Current Issues in Linguistic Theory

Anaphora Processing

Linguistic, cognitive and computational modelling

EDITED BY

António Branco
Tony McEnery
Ruslan Mitkov

ANAPHORA PROCESSING

Volume 263

António Branco, Tony McEnery and Ruslan Mitkov (eds)

Anaphora Processing
Linguistic, cognitive and computational modelling

ANAPHORA PROCESSING

LINGUISTIC, COGNITIVE AND COMPUTATIONAL MODELLING

Edited by

ANTÓNIO BRANCO
Universidade de Lisboa
TONY McENERY
Lancaster University
RUSLAN MITKOV
University of Wolverhampton

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM/PHILADELPHIA

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences — Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

DAARC 2002 (2002 : Lisbon, Portugal)
Anaphora processing : linguistic, cognitive, and computational modelling : selected papers from DAARC 2002 / edited by António Branco, Anthony Mark McEnery, Ruslan Mitkov.
p. cm. -- (Amsterdam studies in the theory and history of linguistic science. Series IV, Current issues in linguistic theory, ISSN 0304-0763 ; v. 263)
Papers presented at the 4th Discourse Anaphora and Anaphor Resolution Colloquium held in Lisbon in Sept. 2002.
Includes bibliographical references and index.
Anaphora (Linguistics)--Data processing--Congresses. Anaphora (Linguistics)--Psychological aspects--Congresses.
P299.A5 D3 2002
415--dc22 2004062375
ISBN 90 272 4777 3 (Eur.) / 1 58811 621 2 (US) (Hb; alk. paper)

John Benjamins Publishing Co. • P.O.Box 36224 • 1020 ME Amsterdam • The Netherlands
John Benjamins North America • P.O.Box 27519 • Philadelphia PA 19118-0519 • USA

CONTENTS

EDITORS' FOREWORD

Anaphora is a central topic in the study of natural language and has long been the object of research in a wide range of disciplines such as theoretical, corpus and computational linguistics, philosophy of language, psycholinguistics and cognitive psychology. On the other hand, the correct interpretation of anaphora has played an increasingly vital role in real-world natural language processing applications including machine translation, automatic abstracting, information extraction and question answering. As a result, the processing of anaphora has become one of the most popular and productive topics of multi- and inter-disciplinary research, and has enjoyed increased interest and attention in recent years.

In this context, the biennial Discourse Anaphora and Anaphor Resolution Colloquia (DAARC) have emerged as the major regular forum for presentation and discussion of the best research results in this area. Initiated in 1996 at Lancaster University and taken over in 2000 by the University of Lisbon, the DAARC series established itself as a specialised and competitive forum for the presentation of the latest results on anaphora processing, ranging from theoretical linguistic approaches through psycholinguistic and cognitive work to corpus studies and computational modelling. The series is unique in that it covers anaphora from such a variety of multidisciplinary perspectives. The fourth Discourse Anaphora and Anaphor Resolution Colloquium (DAARC'2002) took place in Lisbon in September 2002 and featured 44 state-of-the-art presentations (2 invited talks and 42 papers selected from 61 submissions) by 72 researchers from 20 countries.

This volume includes extended versions of the best papers from DAARC'2002. The selection process was highly competitive in that all authors of papers at DAARC'2002 were invited to submit an extended and updated version of their DAARC'2002 paper which was reviewed anonymously by 3 reviewers, members of a Paper Selection Committee of leading international researchers. It is worth mentioning that whilst we were delighted to have so many contributions at DAARC'2002, restrictions on the number of papers and pages which could be included in this volume forced us to be more selective than we would have liked. From the 44 papers presented at the colloquium, we had to select the 20 best papers only.

The book is organised thematically. The papers in the volume have been topically grouped into three sections:

(i) Computational treatment (6 papers)
(ii) Theoretical, psycholinguistic and cognitive issues (7 papers)
(iii) Corpus-based studies (7 papers)

However, this classification should not be regarded as too strict or absolute, as some of the papers touch on issues pertaining to more than one of three above topical groups.

We believe this book provides a unique, up-to-date overview of recent significant work on the processing of anaphora from a multi- and inter-disciplinary angle. It will be of interest and practical use to readers from fields as diverse as theoretical linguistics, corpus linguistics, computational linguistics, computer science, natural language processing, artificial intelligence, human language technology, psycholinguistics, cognitive science and translation studies. The readership will include but will not be limited to university lecturers, researchers, postgraduate and senior undergraduate students.

We would like to thank all authors who submitted papers both to the colloquium and to the call for papers associated with this volume. Their original and revised contributions made this project materialise.

We would like to express our gratitude to all members of the DAARC programme committee as well as the members of this volume's paper selection committee. Without their help, we would not have been able to arrive at such a high quality selection. The following is a list of those who participated in the selection process for the papers in this volume:

Mira Ariel
(Tel Aviv University, Israel)

Amit Bagga
(Avaya Inc., USA)

Branimir Boguraev
(IBM T. J. Watson Research Center, USA)

Peter Bosch
(University of Osnabrück, Institute of Cognitive Science, Germany)

Donna Byron
(The Ohio State University, Computer and Information Science Dept., USA)

Francis Cornish
(CNRS and Université de Toulouse-Le Mirail, Département des Sciences du Langage, France)

Dan Cristea
(University "Al. I. Cuza" of Iasi, Faculty of Computer Science, Romania)

Robert Dale
(Macquarie University, Division of Information and Communication Sciences, Centre for Language Technology, Australia)

Iason Demiros
(Institute for Language and Speech Processing, Greece)

Richard Evans
(University of Wolverhampton, School of Humanities, Languages and Social Sciences, UK)

Martin Everaert
(OTS, The Netherlands)

Claire Gardent
(INRIA-Lorraine, LORIA, France)

Jeanette Gundel
(University of Minnesota, USA and NTNU, Norway)

Sanda Harabagiu
(University of Texas, USA)

Graeme Hirst
(University of Toronto, Department of Computer Science, Canada)

Yan Huang
(University of Reading, Department of Linguistics, UK)

Andrew Kehler
(University of California, San Diego, Department of Linguistics, USA)

Rodger Kibble
(University of London, Department of Computing, UK)

Andrej Kibrik
(Russian Academy of Sciences, Institute of Linguistics, Russia)

Emiel Krahmer
(Tilburg University, The Netherlands)

Shalom Lappin
(King's College, UK)

Naila Pimodri
(University of Cambridge)

Maria Mercedes Piñango
(Yale University, Department of Linguistics, USA)

Massimo Poesio
(University of Essex, Department of Computer Science, UK)

Eric Reuland
(OTS, The Netherlands)

Marco Rocha
(Universidade Federal de Santa Catarina, Brazil)

Antonio Fernandez Rodriguez
(University of Alicante, Spain)

Monique Rolbert
(Université de Marseille, France)

Tony Sanford
(University of Glasgow, Department of Psychology, UK)

Roland Stuckardt
(Johann Wolfgang Goethe University Frankfurt am Main, Germany)

Linda Van Guilder
(MITRE, USA)

We would like also to acknowledge the help received from our series editor, Prof. Konrad Koerner and from Ms Anke de Looper of John Benjamins in Amsterdam. We are also very grateful to João Silva for his patience and considerable help in the nitty gritty of file formatting. Without them this book would not have been a viable project.

September 2004

António Branco
Tony McEnery
Ruslan Mitkov

Section I

Computational Treatment

A Sequenced Model of Anaphora and Ellipsis Resolution

Shalom Lappin
Department of Computer Science, King's College

I compare several types of knowledge-based and knowledge-poor approaches to anaphora and ellipsis resolution. The former are able to capture fine-grained distinctions that depend on lexical meaning and real world knowledge, but they are generally not robust. The latter show considerable promise for yielding wide coverage systems. However, they consistently miss a small but significant subset of cases that are not accessible to rough-grained techniques of interpretation. I propose a sequenced model which first applies the most computationally efficient and inexpensive methods to resolution and then progresses successively to more costly techniques to deal with cases not handled by previous modules. Confidence measures evaluate the judgements of each component in order to determine which instances of anaphora or ellipsis are to be passed on to the next, more fine-grained subsystem.[1]

1 Introduction

Anaphora and ellipsis resolution have been an important focus for work in natural language processing over the past twenty-five years. Providing adequate solutions to these tasks is necessary for the development of genuinely robust systems for (among other applications) text interpretation, dialogue management, query answering, and machine translation. A wide variety of methods have been applied to the treatment of anaphora and ellipsis ranging from knowledge intensive and inference-based techniques to statistical modelling and machine learning. In this paper, I will provide an overview of the main approaches and summarize their comparative strengths and limitations. My concern in this survey is not to offer a detailed account of the numerous computational treatments of anaphora and ellipsis that appear in the literature but to indicate the main advantages and shortcomings of the primary approaches that have been suggested.[2]

[1] Earlier versions of this paper were presented at the 4th Discourse Anaphora and Anaphora Resolution Colloquium in Lisbon in September 2002, the Linguistics Colloquium at the University of Toronto in November 2002, and the Linguistics Colloquium at the University of Reading in January 2003. I am grateful to the audiences of these forums for useful discussion of the ideas presented here. I would also like to thank Ruslan Mitkov and Andy Kehler for their encouragement and their helpful comments on this work.

[2] See (Mitkov, 2002) for a recent study of anaphora resolution that includes a history of the problem within natural language processing. See (Mitkov *et al.*, 2001) for examples of current work on anaphora resolution. (Huang, 2000) offers an extensive cross-linguistic investigation of anaphora and examines alternative linguistic theories of this relation. See (Lappin, 1996) and (Lappin & Benmamoun, 1999) for theoretical and computational discussions of ellipsis resolution.

I will then sketch an integrated model which employs alternative techniques in a sequence of ascending computational cost and domain specificity. This model first invokes relatively inexpensive wide coverage procedures for selecting an antecedent for a pronoun or an elided element. It then moves through successively more expensive, fine-grained measures to handle the cases not resolved by the preceding modules. It applies confidence measures to the decisions of each module to evaluate the reliability of its output. In this way it determines, for each module, which cases have been correctly resolved and which ones are passed on to the following component.

In Section 2, I look at knowledge-based and inference driven approaches to pronominal anaphora resolution. Section 3 considers various knowledge-poor methods for anaphora interpretation. Section 4 extends the comparison to VP ellipsis, and Section 5 takes up fragment interpretation in dialogue viewed as a type of ellipsis. Finally, in Section 6, I describe the proposed sequenced model. Section 7 states conclusions and indicates directions for future work.

2 Knowledge-Based and Inference-Driven Approaches to Anaphora

Knowledge-based approaches to anaphora resolution generally rely on rules of inference that encode semantic and real world information in order to identify the most likely antecedent candidate of a pronoun in discourse. An interesting example of this approach is Kehler's (2000, 2002) use of Hobbs *et al.*'s (1993) model of abductive reasoning to establish coherence relations among the sentences of a text. In Kehler's theory, pronouns are assigned antecedents through the abductive inference chains required for text coherence. Hobbs and Kehler (1997), and Kehler (2002) also invoke abductive inference to interpret elided VP's and resolve pronouns within VP ellipsis.

To illustrate this approach consider (1), to which Kehler (2000) assigns the representation (2).

(1) John hid Bill's keys. He was drunk.

(2) a. $hide(e_1,John,Bill,ck) \wedge car\_keys(ck,Bill)$
b. $drunk(e_2,he)$

He uses axioms like those in (3) to construct the backwards abductive inference chain in (4) from (2) to a conclusion in which *he* is resolved to *Bill* (4g).

(3) a. $\forall e_i,e_j(cause(e_j,e_i) \Rightarrow explanation(e_i,e_j))$
b. $\forall x,y,e_i(drunk(e_i,x) \Rightarrow \exists e_j,e_k(not\_want(e_j,y,e_k) \wedge drive(e_k,x), \wedge cause(e_i,e_j)))$

(4) a. $explanation(e_1,e_2)$
b. $cause(e_2,e_1)$
c. $cause(e_2,e_3) \wedge cause(e_3,e_1)$
d. $cause(e_2,e_4) \wedge cause(e_4,e_3)$

e. not_want(e_3,John,e_5) $\wedge$ have(e_5,Bill,ck)
f. not_want(e_4,John,e_6) $\wedge$ drive(e_6,Bill)
g. drunk(e_2,Bill)

The main strength of knowledge-based systems is their capacity to capture fine-grained semantic and pragmatic distinctions not encoded in syntactic features or frequency of co-occurence patterns. These distinctions are not accessible to knowledge-poor approaches. They are crucial to correctly resolving pronominal anaphora and VP ellipsis in a small but important set of cases that arise in text and dialogue.

The two main difficulties with these systems are that (i) they require a large database of axioms encoding real world knowledge, and (ii) they apply defeasible inference rules which produce combinatorial blow up very quickly. Assigning cost values to inference rules and invoking a cost driven preference system for applying these rules (as in (Hobbs *et al.*, 1993)) may reduce the blow up to some extent, but the problem remains significant.

As a result, knowledge-based models of anaphora resolution are generally not robust. Their rules are often domain-dependent and hard to formulate in a way that will support inference over more than a small number of cases. Moreover, the semantic/discourse representations to which the inference rules apply are not reliably generated for large texts.

3 Knowledge-Poor Approaches

Knowledge-poor systems of anaphora resolution rely on features of the input which can be identified without reference to deep semantic information or detailed real world knowledge. One version of this approach employs syntactic structure and grammatical roles to compute the relative salience of candidate antecedents. Another uses machine-learning strategies to evaluate the probability of alternative pronoun-antecedent pairings by training on large corpora in which antecedent links are marked.

Hobbs (1978) suggests one of the first instances of a syntactic salience procedure for resolving pronouns. He formulates a tree search algorithm that uses syntactic configuration and sequential ordering to select NP antecedents of pronouns through left-right, breadth-first traversal of a tree. Lappin and Leass (1994) propose an algorithm which relies on weighted syntactic measures of salience and recency to rank a filtered set of NP candidates. This algorithm applies to full syntactic parses. Kennedy and Boguraev (1996), Mitkov (1998), and Stuckardt (2001) modify and extend this approach to yield results for partial syntactic representations rather than full and unambiguous parse structures. Grosz *et al.* (1995) employ a grammatical role hierarchy and preference rules for managing informational state change to select the local NP centre (focus) for each element of the sequence of sentences in a discourse.

A recent instance of the application of machine learning to anaphora is (Soon *et al.*, 2001). They describe a procedure for training a classifier on a corpus annotated with coreference chains, where the NP elements of these chains are assigned a set of features. The classifier goes through all pairs of referential NP's in a text to identify a subset of coreferential pairs.

The obvious advantage of knowledge-poor systems relative to knowledge-based models is that the former are computationally inexpensive and potentially robust. However, these claims of resource efficiency and wide coverage must be qualified by recognition of the expense involved in generating accurate syntactic representations for systems that apply to full parses or detailed grammatical role information. Salience-driven systems also require domain specific and, possibly, language specific values for syntactic salience measures. In the case of machine learning techniques, it is necessary to factor in the cost of annotating large corpora and training classification procedures.

An important weakness of these models is that they cannot handle a small but significant core of anaphora resolution cases in which salience cannot be identified solely on the basis of syntactic and morphological properties, and relative recency. These features are also the basis for the candidate rankings that machine learning methods generate.

Dagan *et al.* (1995) attempt to enrich a syntactic salience system by modelling (a certain amount of) semantic and real world information cheaply. They combine the Lappin-Leass algorithm with a statistically trained lexical co-occurrence preference module. Elements of the candidate antecedent list are assigned both salience and lexical preference scores. The latter are based on frequency counts for verb-NP and prep-NP pairs in a corpus, and the substitution of the candidate for the pronoun in the observed head-argument relation of the pronoun. When the difference between the salience scores of the two highest ranked candidates is below a (experimentally) determined threshold and the lexical preference score of another candidate C_i exceeds that of the first by a (experimentally) specified ratio, then C_i is selected.

Consider the pronoun *it* in 5.

(5) The utility (CDVU) shows you a LIST4250, LIST38PP, or LIST3820 file on your terminal for a format similar to that in which *it* will be printed.

The statistical preference module overrides the higher syntactic salience ranking of *utility* to select *file* as the antecedent of *it*. This preference is due to the fact that *print file* has a significantly higher frequency count than *print utility*. The statistical module improved the performance of Lappin and

Leass' (1994) syntactic salience-based algorithm from 86.1% to 88.6% on a blind test of 360 pronoun cases in a set of sentences taken from a corpus of computer manuals.

However, there are cases which still resist resolution even under the finer grain of lexical co-occurrence information that such a statistical preference module provides. The contrast between (6) (= 1) and (7) illustrates the limits of syntactic salience enriched with a statistically trained lexical preference metric.

(6) John hid Bill's keys. He was drunk.
(7) John hid Bill's keys. He was playing a joke on him.

John receives the highest syntactic salience ranking in both (6) and (7). Lexical preference conditions do not select between *John* and *Bill* in these cases. Reliance on real world knowledge and inference are needed to identify *Bill* as the antecedent of *he* in (6), and *John* and *Bill* as the antecedents of *he* and *him*, respectively, in (7).

4 VP Ellipsis

Asher *et al.* (2001) specify a knowledge-based approach to the interpretation of VP ellipsis. They employ a general parallelism constraint based on Asher's (1993) notion of *Maximal Common Theme* (MCT) to resolve ambiguity in VP ellipsis. They define a *Theme* for a Discourse Representation Structure (DRS) K as a DRS K' obtained from K by the application of 0 or more operations of a certain type on K. These operations are (i) deletion of a discourse marker, (ii) deletion of an atomic condition, and (iii) systematic renaming of a bound discourse marker. A *Common Theme* (CT) T of two DRS's J and K is a theme of J and a theme of K which is such that for any other theme T' of J and K, T is a theme of T'. Asher *et al.*'s maximalisation constraint states that in resolving scope ambiguity within a VP ellipsis construction, the preferred DRS for the elided VP and its antecedent is the DRS that provides the MCT for the DRS's representing each clausal constituent. This constraint effectively constitutes a unification principle for the discourse representations of the sentences containing the elided and antecedent VP's.[3]

The MCT condition selects the parallel wide scope reading of the quantified NP *every* student relative to *a test* in (8).

(8) John gave every student a test, and Bill did too.

This is because the DRS's corresponding to this reading of each clause yield a CT that is a theme of the DRS's for the wide scope interpretation of *a test*

[3] Asher *et al.* (2001) also invoke this condition to resolve pronouns in ambiguous elided VP's.

relative to *every student*. The DRS's of the wide scope reading for *a test* do not produce a theme for the DRS's of the wide scope reading of *every student*.

Several other instances of knowledge-based and inference-driven models of VP ellipsis interpretation are as follows. Hobbs and Kehler (1997), and Kehler (2002) use parallelism constraints for text coherence to identify VP antecedents. Dalrymple *et al.* (1991) and Shieber *et al.* (1996) apply higher-order unification to resolve the predicate variable in the semantic representation of an elided VP. Crouch (1999) constructs derivations in linear logic to provide alternative ways of assembling the constituents in the representation of an antecedent in order to obtain possible interpretations of the clause containing the ellipsis site.

This approach to VP ellipsis enjoys the same advantages and suffers from the same weaknesses that we noted with respect to the knowledge intensive view of pronominal anaphora resolution.

Turning to a knowledge-poor model, Hardt (1997) describes a procedure for identifying the antecedent of an elided VP in text that applies to the parse structures of the Penn Treebank.[4]

It constructs a list of candidate VP's to which it applies a syntactic filter. The elements of the filtered candidate list are assigned scores on the basis of syntactic salience factors and recency.

On a blind test of 96 examples from the *Wall Street Journal* the procedure achieved a success rate of 94.8% according to a head verb overlap criterion (the head verb of the system's selected candidate is contained in, or contains the head verb of the coder's choice of antecedent). It achieved 85.4% for exact head verb match and 76% for full antecedent match. A comparison procedure that relies only on recency scored 75% for head verb overlap, 61.5% for exact head verb match, and 14.6% for full antecedent match.

Hardt's syntactic salience-based procedure uses essentially the same strategy and design for identifying the antecedent of an elided VP as Lappin and Leass' (1994) algorithm applies to pronominal anaphora resolution. Its higher success rate may, in part, be due to the fact that recency and syntactic filtering tend to reduce the set of candidates more effectively for elided VP's than for pronouns.

As in the case of pronouns, a small set of elided VP cases are not accessible to resolution by salience ranking or statistically modelled lexical preference. The following examples clearly indicate that inference based on semantic and real world knowledge appears to be inescapable for these cases.[5]

---

[4] Hard's procedure applies to elided VP's that have already been recognized. Nielsen (2003, 2004) presents preliminary results for the application of a variety of machine learning methods to the identification of elided VP's in text.

[5] Dalrymple (1991), Hardt (1993), and Kehler (2002) claim that the fact that inference is required to identify the antecedents of the elided VP's in (9) and (10) shows that ellipsis resolution applies to semantic rather than

(9) Mary and Irv want to go out, but Mary can't, because her father disapproves of Irv. (Webber, 1979)
Mary can't go out with Irv

(10) Harry used to be a great speaker, but he can't anymore, because he lost his voice. (Hardt, 1993)
he can't speak anymore

5 The Interpretation of Fragments in Dialogue

Fernández *et al.* (to appear) present SHARDS, a system for interpreting non-sentential phrasal fragments in dialogue. Examples of such fragments are short answers (11), sluices (short questions, 12), and bare adjuncts (13). The latter are possible even when no wh-phrase adjunct appears in the antecedent to anchor them, as in (14).

(11) A: Who saw Mary?
B: John.
John saw Mary.

(12) A: A student saw John.
B: Who?
Which student saw John?

(13) A: When did Mary arrive?
B: At 2.
Mary arrived at 2.

(14) A: John completed his paper.
B: When?
When did John complete his paper?

SHARDS is a Head Driven Phrase Structure Grammar (HPSG)-based system for the resolution of fragments in dialogue. It treats the task of resolving fragment ellipsis as locating for the (target) ellipsis element a parallel (source) element in the context, and computing from contextual information a property which composes with the target to yield the resolved content. This basic view of ellipsis resolution is similar in spirit to the higher-order unification (HOU) approach of Dalrymple *et al.* (1991) and Pulman (1997).

---

syntactic representations. In fact, it is not obvious that the need for inference in (some cases of) ellipsis resolution in itself determines the nature of the representation to which the inference rules apply. Lappin (1996) argues that inference can apply to syntactic representations of sentences to generate structures corresponding to (i) and (ii).

(i) Mary wants to go out with Irv.

(ii) Harry used to speak.

These structures supply appropriate antecedents for the syntactic reconstruction of the elided VP's in (9) and (10), respectively. The need for inference in ellipsis resolution, on one hand, and the nature of the level of representation to which inference and ellipsis resolution apply, on the other, are independent questions which should be distinguished.

Two new attributes are defined within the CONTEXT feature structure: the *Maximal Question Under Discussion* (MAXQUD) and the *Salient Utterance* (SALUTT). The MAXQUD is the most salient question that needs to be answered in the course of a dialogue. The SALUTT represents a distinguished constituent of the utterance whose content is the current value of MAXQUD. In information structure terms, the SALUTT specifies a potential parallel element correlated with an element in the antecedent question or assertion. The SALUTT is the element of the MAXQUD that corresponds to the fragment phrase. By deleting the SALUTT from the MAXQUD, SHARDS produces the representation of a property from which the propositional core of the CONTENT value for the fragment can be constructed.

(15) is the (simplified) typed feature structure that (Fernández *et al.*, to appear) posit for a bare fragment phrase.

(15)bare-arg-ph $\Rightarrow$

$$\begin{bmatrix} \text{STORE} & \{\ \} \\ \text{CONT}\,|\,\text{SOA}\,|\,\text{NUCL} & \boxed{1} \\ \text{CTXT} & \begin{bmatrix} \text{MAX}-\text{QUD}\,|\,\text{SOA}\,|\,\text{NUCL} & \boxed{1} \\ \text{SAL}-\text{UTT} & \begin{bmatrix} \text{CAT} & \boxed{2} \\ \text{CONT}\,|\,\text{INDEX} & \boxed{3} \end{bmatrix} \end{bmatrix} \\ \text{HD}-\text{DTRS} & \begin{bmatrix} \text{CAT} & \boxed{2} \\ \text{CONT}\,|\,\text{INDEX} & \boxed{3} \end{bmatrix} \end{bmatrix}$$

SHARDS interprets a fragment in dialogue by computing from context (represented as a dialogue record) the values of MAXQUD and SALUTT for the assertion or question clause that the fragment expresses. It uses these feature values to specify the CONTENT feature of the clause for the fragment. The basic propositional content of the fragment clause is recovered from the MAXQUD, whose NUCL feature value is shared with the NUCL feature of the fragment clause's CONT feature.

The value of SALUTT is of type *sign*, enabling the system to encode syntactic categorial parallelism conditions, including case assignment for the fragment. The SALUTT is computed as the (sub)utterance associated with the role bearing widest quantificational scope within the MAXQUD.

SHARDS computes the possible MAXQUD's from each sentence which it processes and adds them to the list of MAXQUD candidates in the dialogue record. When a fragment phrase FP is encountered, SHARDS selects the most recent element of the MAXQUD candidate list which is compatible with FP's clausal semantic type.

(16) is the Attribute Value Matrix (AVM) produced for the CONT of *Who saw Mary*. $\boxed{1}$ is the index value of *who* and $\boxed{2}$ of *Mary*:

(16)

$$\begin{bmatrix} \text{PARAMS} & \begin{bmatrix} \text{INDEX} & \boxed{1} \\ \text{RESTR} & \{person-rel(\boxed{1})\} \end{bmatrix} \\ \text{SOA | NUCL} & \begin{bmatrix} saw-rel & \\ \text{see}-\text{er} & \boxed{1} \\ \text{seen} & \boxed{2} \end{bmatrix} \end{bmatrix}$$

This is the feature structure counterpart of the λ-abstract λπ.(…π…).

The (abbreviated) AVM for the SALUTT *who* is (17).

(17)

$$\begin{bmatrix} \text{CAT} & \text{NP}[+\text{nom}] \\ \text{CONTENT} & \boxed{1} \\ \text{STORE} & \begin{bmatrix} \text{INDEX} & \boxed{1} \\ \text{RESTR} & \{person(\boxed{1})\} \end{bmatrix} \end{bmatrix}$$

(18) is the AVM produced for *John* as a short answer, where $\boxed{1}$ is the index value of *John* and $\boxed{2}$ of *Mary*.

(18)

$$\begin{bmatrix} \text{SOA} & \begin{bmatrix} \text{NUCL} & \begin{bmatrix} see-rel & \\ \text{see}-\text{er} & \boxed{1} \\ \text{seen} & \boxed{2} \end{bmatrix} \\ \text{RESTR} & \{person-rel(\boxed{1})\} \end{bmatrix} \end{bmatrix}$$

(19) is the feature structure generated for *Who* as a sluice in response to *A student saw John*. $\boxed{1}$ is the index value of *Who* and $\boxed{2}$ of *John*.

(19)

$$\begin{bmatrix} \text{PARAMS} & \begin{bmatrix} \text{INDEX} & \boxed{1} \\ \text{RESTR} & \left\{\begin{array}{l} person-rel(\boxed{1}), \\ student-rel(\boxed{1}) \end{array}\right\} \end{bmatrix} \\ \text{SOA | NUCL} & \begin{bmatrix} see-rel & \\ \text{see}-\text{er} & \boxed{1} \\ \text{seen} & \boxed{2} \end{bmatrix} \end{bmatrix}$$

For at least some cases it is necessary to relax the requirement of strict syntactic *category* match between the fragment and the SALUTT to allow correspondence to be specified in terms of an equivalence class of categories.

(20) A: What does Mary want most?
B: a good job/that people should like her/to have her freedom

(21) A: When did John complete his paper?
B: yesterday/on Wednesday/after the teacher spoke to him

There are also instances where the scope criterion for determining the SALUTT must be overridden.

(22)a. A: Each student will consult a supervisor.
B: Which one?

b. Which supervisor will each student consult?

(22b), which selects *a supervisor* as the SALUTT, is the most natural interpretation of *which one* in (22a) B, even when *a supervisor* receives a narrow scope reading relative to *each student* in (22a) A.

Similarly, the recency condition for selecting the MAXQUD from the list of MAXQUD candidates in the dialogue record does not always yield the correct results, as the dialogue sequence in (23) illustrates.

(23) A: Why did Mary arrive early?
B: I can't tell you.
A: Why can't you tell me?
B: Okay, if you must know, to surprise you.

The fragment phrase *to surprise you* is a reply to the first question that A asks, *Why did Mary arrive early?*, rather than the second, *Why can't you tell me?*.

Knowledge-based inference is required to select the more distant question as the preferred MAXQUD in this case.

The following example for the British National Corpus is an example of a naturally occurring dialogue in which the recency criterion for determining the MAXQUD is defeasible.[6]

(24) A1: That new tyre law comes in soon dunnit?
B2: That what?
A3: New tyre law.
C4: First of <pause> first of November it came in.
A5: Oh.
C6: Why?
A7: I'd better check my two back ones then.

The sluice in (24) C6 is a case of clarificatory ellipsis (Ginzburg & Cooper, to appear). It takes as its MAXQUD antecedent the illocutionary statement corresponding to (24) A1 rather than the statement in (24) C4.

[6] (24) is from the dialogue component of the British National Corpus, File KB4, sentences 144-150. I am grateful to Raquel Fernández for providing this example.

6 A Sequenced Model

As we have seen, work on anaphora and ellipsis within the framework of the knowledge-poor approach indicates that syntactic measures of salience combined with recency provide a highly effective procedure for antecedent identification across a wide range of ellipsis and anaphora resolution tasks in text and dialogue. These methods are computationally inexpensive and generally robust. It is possible to deal with a subset of the (significant) minority of cases which are not amenable to syntactic salience-based resolution through statistical modelling of semantic and real world knowledge as lexical preference patterns. For the remaining cases abductive inference appears to be unavoidable. These considerations suggest that a promising approach is to apply the techniques in an ascending sequence of computational cost. (25) gives the outline of a plausible architecture for such an integrated sequenced model of anaphora and ellipsis resolution.

(25) <P,Candidate_Antecedent_List> ⇒

Module 1
Syntactic Salience & Recency Measures +
Syntactic & Morphological Filtering →
Ranked Candidate List →
Confidence Metric 1 →
correctly resolved;
unresolved ⇒

Module 2
Statistically Determined Lexical Preference Measures →
New Ranked Candidate List →
Confidence Metric 2 →
correctly resolved;
unresolved ⇒

Module 3
Abductive Inference ⇒
resolved

The sequenced model of anaphora and ellipsis resolution proposed here moves successively from computationally inexpensive and interpretationally rough-grained procedures to increasingly costly and fine-grained methods. The model encodes a strategy of maximizing the efficiency of an anaphora (ellipsis) resolution system by invoking fine-grained techniques only when necessary.

In order to succeed, this strategy must use reliable confidence metrics to evaluate the candidate rankings which the first two modules produce. Such metrics can be constructed on the model of the criteria that Dagan *et al.* (1995) use to evaluate the reliability of salience scores. When the distance between the salience scores of the two top candidates in a list falls below a certain threshold, the ranking is taken as an unreliable basis for antecedent selection and the statistical lexical preference module is activated. Intensive experimental work using machine learning techniques will be needed to determine optimal values for both the salience factors of Module 1 and the confidence metrics used to assess the outputs of Modules 1 and 2.

A computationally viable abductive inference component will require resource sensitive inference rules to control the size and number of the inference chains that it generates.[7] Resource sensitivity and upper bounds on derivations in abductive inference are essential to rendering the procedures of Module 3 tractable.

7 Conclusions and Future Work

While the knowledge-based and inference driven approach to anaphora and ellipsis resolution can deal with cases that require fine-grained semantic interpretation and detailed real world knowledge, it does not provide the basis for developing computationally efficient, wide coverage systems. By contrast, knowledge-poor methods are inexpensive and potentially robust, but they miss an important minority of recalcitrant cases for which real world knowledge and inference are indispensable. A promising solution to this engineering problem is to construct an integrated system that orders the application of anaphora and ellipsis interpretation techniques in a sequence of modules that apply increasingly fine-grained techniques of interpretation with an attendant rise in computational cost. Confidence metrics filter the output of each module to insure that the more expensive components are invoked only when needed.

In order to implement the proposed model, it is important to achieve optimisation of the selection of salience parameters and their relative values through statistical analysis of experimental results. A considerable amount of work has been done on the application of salience parameters and values to minimal syntactic representations rather than fully specified parse structures. This is a fruitful area of investigation which merits further research, as it holds out the promise of major gains in efficiency and robustness for the salience methods that comprise the first module of an integrated system. Another problem worth pursuing is the generalization of lexical preference patterns to relations between semantic classes. Measuring preference in terms of semantic categories rather than specific lexical head-argument and head-adjunct patterns

[7] Kohlhase and Koller (2003) propose resource sensitive inference rules for model generation in which the salience of a referential NP in discourse is used to compute the relative cost of applying inference rules to entities introduced by this NP. They measure salience in discourse largely in terms of the sorts of syntactic and recency factors that Lappin and Leass (1994) use in their anaphora resolution algorithm.

will increase the power and reliability of Module 2. The viability of the entire system depends upon determining reliable confidence metrics for both salience-based and lexical preference-based antecedent selection. Finally, to implement the third module much work must be done to develop efficiently resource sensitive procedures for abductive inference in different domains.

References

Asher, N. 1993. *Reference to Abstract Objects in English*. Dordrecht: Kluwer.

Asher, N., D. Hardt and J. Busquets 2001. “Discourse Parallelism, Ellipsis, and Ambiguity.” *Journal of Semantics* 18.

Crouch, D. 1999. “Ellipsis and Glue Languages.” S. Lappin and E. Benmamoun (eds.) *Fragments: Studies in Ellipsis and Gapping.* New York: Oxford University Press.32-67.

Dagan, I., J. Justeson, S. Lappin, H. Leass and A. Ribak 1995. “Syntax and Lexical Statistics in Anaphora Resolution.” *Applied Artificial Intelligence* 9:6.633-644.

Dalrymple, M. 1991. “Against Reconstruction in Ellipsis.” Xerox PARC, Palo Alto, CA: unpublished ms.

----------, S. Shieber and F. Pereira 1991. “Ellipsis and Higher-Order Unification.” *Linguistics and Philosophy* 14.399-452.

Fernández, R., J. Ginzburg, H. Gregory and S. Lappin. To appear. “SHARDS: Fragment Resolution in Dialogue.” H. Bunt and R. Muskens (eds.) *Computing Meaning* 3. Dordrecht: Kluwer.

Ginzburg, J. and R. Cooper. To appear. “Clarification, Ellipsis, and the Nature of Contextual Updates.” *Linguistics and Philosophy*.

Grosz, B., A. Joshi and S. Weinstein 1995. “Centering: A Framework for Modeling the Local Coherence of Discourse.” *Computational Linguistics* 21.203-225.

Hardt, D. 1993. “Verb Phrase Ellipsis: Form, Meaning, and Processing.” University of Pennsylvania: unpublished Ph.D dissertation.

---------- 1997. “An Empirical Approach to VP Ellipsis.” *Computational Linguistics* 23.

Hobbs, J. 1978. “Resolving Pronoun References.” *Lingua* 44.339-352.

---------- and A. Kehler 1997. “A Theory of Parallelism and the Case of VP Ellipsis.” *Proc. of the 35th Conference of the ACL.* 394-401. Madrid.

----------, M. Stickel, D. Appelt and P. Martin 1993. “Interpretation as Abduction.” *Artificial Intelligence* 63.69-142.

Huang, Y. 2000. *Anaphora: A Cross-Linguistic Study*. Oxford: Oxford University Press.

Kehler, A. 2000. “Pragmatics, Chapter 18.” D. Jurafsky and J. Martin (eds.) *Speech and Language Processing*. Upper Saddle River, NJ: Prentice Hall.

---------- 2002. *Coherence, Reference, and the Theory of Grammar*. Stanford, CA: CSLI.

Kennedy, C. and B. Boguraev 1996. “Anaphora for Everyone: Pronominal Anaphora Resolution without a Parser.” *Proc. of the 16th Int. Conference on Computational Linguistics* (COLING'96). Copenhagen. 113-118.

Kolhase, M. and A. Koller 2003. “Resource-Adaptive Model Generation as a Performance Model.” *Logic Journal of the IGPL*. 435-456.

Lappin, S. 1996. “The Interpretation of Ellipsis.” S. Lappin (ed.) *Handbook of Contemporary Semantic Theory* Oxford: Blackwell. 145-175.

---------- and E. Benmamoun (eds.) 1999. *Fragments: Studies in Ellipsis and Gapping*. New York: Oxford University Press.

---------- and H. Leass 1994. “A Syntactically Based Algorithm for Pronominal Anaphora Resolution.” *Computational Linguistics* 20.535-561.

Mitkov, R. 1998. “Robust Pronoun Resolution with Limited Knowledge.” *Proc. of ACL'98 and COLING'98*. 869-875. Montreal.

---------- 2002. *Anaphora Resolution*. London: Longman.

----------, B. Boguraev and S. Lappin (eds.) 2001. *Computational Linguistics* 27. Special Issue on Computational Anaphora Resolution.

Nielsen, L. 2003. “Using Machine Learning Techniques for VPE Detection”. *Proc. of RANLP 2004*. 339-346. Borovetz.

Nielsen, L. 2004. “Verb Phrase Ellipsis Detection Using Automatically Parsed Text”. *Proc. of COLING 2004*. Geneva.

Pulman, S. 1997. “Focus and Higher-Order Unification.” *Linguistics and Philosophy* 20.

Shieber, S., F. Pereira and M. Dalrymple 1996. “Interactions of Scope and Ellipsis.” *Linguistics and Philosophy* 19.527-552.

Soon, W., H. Ng and D. Lim 2001. “A Machine Learning Approach to Coreference Resolution of Noun Phrases.” *Computational Linguistics* 27.521-544.

Stuckardt, R. 2001. “Design and Enhanced Evaluation of a Robust Anaphor Resolution Algorithm.” *Computational Linguistics* 27.479-506.

How to Deal with Wicked Anaphora?

Dan Cristea[1,2] and Oana-Diana Postolache[1]
[1] *Al.I.Cuza University, Faculty of Computer Science*
[2] *Romanian Academy, Institute for Theoretical Computer Science*

This paper revises a framework (called AR-engine) capable of easily defining and operating models of anaphora resolution. The proposed engine envisages the linguistic and semantic entities involved in the cognitive process of anaphora resolution as represented in three layers: the referential expressions layer, the projected layer of referential expression's features and the semantic layer of discourse entities. Within this framework, cases of anaphora resolution usually considered difficult to be tackled are investigated and solutions are proposed. Among them, one finds relations triggered by syntactic constraints, lemma and number disagreement, and bridging anaphora. The investigation uses a contiguous text from the belletrist register. The research is motivated by the view that interpretation of free language in modern applications, especially those related to the semantic web, requires more and more sophisticated tools.

1 Introduction

Although it is generally accepted that semantic features are essential for anaphora resolution, due to the difficulty and complexity of achieving a correct semantic approach, authors of automatic systems mainly preferred to avoid the extensive use of semantic information (Lappin & Leass, 1994; Mitkov, 1997; Kameyama, 1997). It is well known that anaphora studies reveal a psychological threshold around the value of 80% precision and recall that seems to resist any attempt to be surmounted by present systems (Mitkov, 2002). It is our belief that one of the causes for the current impasse of devising an anaphora resolution (AR) system with a very high degree of confidence should be searched also in the choice for a sub-semantic limitation. Drawn mainly on strict matching criteria, in which morphological and syntactic features are of great value, these systems disregard resolution decisions based on more subtle strategies that would allow lemma and number mismatch, gender variation, split antecedents, bridging anaphora or cataphora resolution. Moreover, types of anaphora different than strict coreference, like type/token, subset/superset, is-element-of/has-as-element, is-part-of/has-as-part, etc. often impose more complex types of decision-making, which could get down to the semantic level as well.

Our study makes use of the AR framework defined by Cristea and Dima (2001), and Cristea *et al.* (2002a) (called **AR-engine**) with the aim of applying it to the treatment of cases of anaphora resolution usually considered to be difficult. The AR-engine approach is settled on a view that sees anaphoric relations as having a semantic nature (Halliday & Hassan, 1976), as opposed to a textual nature.

This paper discusses the tractability of implementing AR-models capable of tackling cases of anaphora usually considered difficult. The validation of the approach is currently being done on a contiguous free text by informally appreciating the computational feasibility of the proposed solutions within the AR-engine framework.

The research is motivated by the belief that interpretation of free language in modern applications, especially those related to semantic web, justifies more and more sophisticated tools. We think that our investigation is a step forward towards dealing with really hard anaphora resolution problems as those occurring in free texts. The study intends to determine a psychological boundary beyond which is really hard to process anaphora. It is our belief that the usual lack of interest for considering hard cases of anaphora in practical settings is not always motivated by high modelling and computational costs and their notoriety of "untouchables", tacitly accepted, is exaggerated. The real hard life in dealing with AR happens only when world knowledge is to be put on the table. In this paper, we try to prove that until then, there is still a lot to do.

The presentation proceeds as follows: Section 2 describes AR-engine: its basic principles, the constituent parts in the definition of a model within the framework and the basic functionality of the engine put to analyse a free text. Sections 3 to 7 discuss cases of AR, from more simple to more complex. Finally, Section 8 presents preliminary evaluation data and conclusions.

2 The framework

2.1 The AR-engine[1] basic principles

In (Cristea & Dima, 2001; Cristea *et al.*, 2002a) a framework having the functionality of a general AR engine and able to accommodate different AR models is proposed. This approach recognizes the intrinsic incrementality of the cognitive process of anaphora interpretation during reading a text or listening a discourse. It sees the linguistic and semantic entities involved in the process of AR as settled on two fundamental layers: a **text layer** – populated with referential expressions (REs),[2] and a deep **semantic layer** – where discourse entities (DEs), representations of entities the discourse is about, are placed.

---

1 AR-engine and the related documentation are freely available for research purposes at http://consilr.info.uaic.ro.

2 We will restrict this study only to nominal referential expressions.

Within such a view, two basic types of anaphoric references can be expressed: coreferences, inducing equivalence classes of all REs in a text which participate in a coreference chain, and functional references (Markert *et al.*, 1996), also called indirect anaphora or associative anaphora (Mitkov, 2002), which express semantic relations between different discourse entities, including type/token, is-part-of/has-as-part, is-element-of/has-as-element, etc. As sketched in Figure 1, chains of coreferential REs are represented as corresponding to a unique DE on the semantic layer, whereas functional references are represented as relational links between the DEs of the corresponding REs.

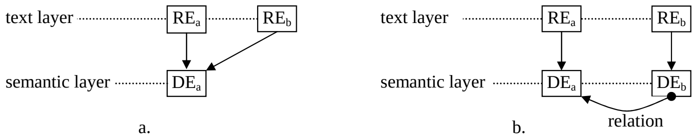


Figure 1*: Representation of anaphoric relations revealing their semantic nature: a. coreferences; b. functional references.*

Representations involving only REs and DEs are the result of an interpretation process applied to a text. Even if the semantic level is kept hidden, these types of representations are implicitly assumed by the majority of anaphora resolution annotation tasks. Indeed, DEs of the semantic layer could be short-circuited by appropriate tags associated to coreferential REs, where each RE points either to the first RE of the chain or to the most recent antecedent RE. Analogously, in the case of functional references, the annotation tags associated to the surface REs name the nature of the referential function. However, if we are interested to model the interpretation process itself, in a way that simulates the cognitive processes developed in a human mind during text reading, the need for another intermediate layer can immediately be argued for. On this layer, that we will call the **projection layer**, feature structures (in the following, projected structures – PSs) are filled-in with information fetched from the text layer and all the resolution decisions are to be negotiated between PSs of the projection layer and DEs of the semantic layer. We will say that a PS is **projected** from an RE and that a DE is **proposed** (if it appears for the first time in the discourse) or **evoked** (if it exists already) by a PS (Figure 2).

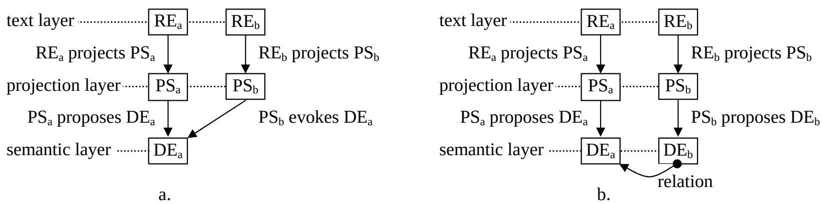


Figure 2*: The three-layer representation of:*
a. two coreferring expressions; b. two functional referential expressions

We term referential expression (RE) any noun phrase having a referential function, including the first mention of an entity. The coreference relation (two REs are coreferent if they refer to the same entity (Hirschman *et al.*, 1997)) is, in most of the cases, anaphoric,[3] while not all anaphoric relations are coreferential (e.g. bridging anaphora). Then, according to the usual acceptance (see for instance (Mitkov, 2002)), if RE_b corefers with RE_a, with RE_b following RE_a in text, we say that RE_b is the **anaphor** and RE_a the **antecedent**. In order to stress the semantic nature of anaphora as a referential relation (Halliday & Hassan, 1976), if anaphors and antecedents remain intrinsically connected to the text, discourse entities belong to the semantic layer and are said to be the **referents** of REs. The unique DE that is referred to by a set of REs disposed in sequence reveals thus the equivalence class of these REs as a chain of coreferencing expressions.

Figure 3 presents a sequence of phases during the functioning of the AR-engine in which two referential expressions are found to corefer. First, the referential expression RE_a is identified on the text layer. It projects down to the projection layer a feature structure composed of a set of attribute-value pairs – PS_a (Figure 3a). Supposing the model decides in favour of considering RE_a as introducing a new discourse entity during interpretation, the feature structure PS_a proposes an adequate semantic representation on the semantic layer – DE_a, mainly a copy of PS_a (Figure 3b). Because the aim of the projected structure is

[3] For the definition of anaphoric relations we adopt a somehow different position than Deemter and Kibble (2000), for instance. They argue that, following the definition of anaphora: *an NP* α_1 *is said to take an NP* α_2 *as its anaphoric antecedent if and only if* α_1 *depends on* α_2 *for its interpretation* (e.g. (Kamp & Reyle, 1993)), *W.J.Clinton* and *Hillary Rodham's husband*, are not anaphoric since *Hillary Rodham's husband* can be understood as *W.J.Clinton* by itself, therefore without the help of the former RE. Our meaning for α_1 *depends on* α_2 *for its interpretation* is α_1 *and* α_2 *are related in the given setting*. In this sense, the two REs above are anaphoric if the intent of the writer is to let the reader establish a link between the two mentions, in this particular case, as the same person. In (Cristea, 2000), co-referential non-anaphoric references are called *pseudo references*. These are REs which, although referring to the same entity, can be understood independently without making the text interpretation to suffer if a relation between them is not established (for instance, two mentions of *the sun*: *I waked up this morning when the sun rose;* and later on: *I read a book about Amenomphis the IV*th*, the Egyptian pharaoh, son of the sun*).

to help the proposal/identification of a discourse entity, once this task has been fulfilled, the projected structure can be discarded. The result is a bidirectional link that will be kept between RE_a and the corresponding DE_a. Some moments later, when a referential expression RE_b is identified on the text layer, it projects a features structure PS_b on the projection layer (Figure 3c). Finally, if the model takes the decision that PS_b evokes DE_a, a bidirectional link between RE_b and DE_a is established and PS_b is discarded (Figure 3d). A similar sequence takes place when other types of anaphoric relations than strict coreference are established.

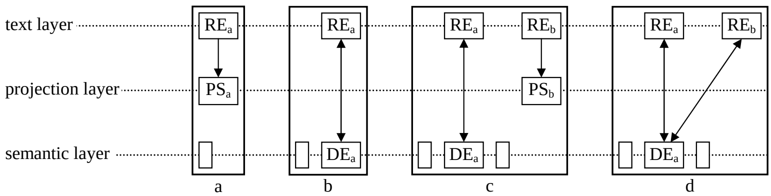


Figure 3: *a. Projection of PS_a from RE_a; b. Proposing of DE_a from PS_a; c. Projection of PS_b from RE_b; d. Evocation of DE_a by PS_b*

2.2 Definition of an AR model

The AR-engine framework can accommodate different **AR models**. Such a model is defined in terms of four components. The first component specifies **the set of attributes of the objects populating the projection and semantic layers and their corresponding types**. Different approaches in AR may lead to consider specific options for what features of the anaphor and the referent are to be considered important in the resolution process. An analysis of the state of the art of the existing approaches suggests a classification of the possible features (attributes) on the following lines:

a. morphological features:

- grammatical number;
- grammatical gender;
- case.

All known approaches use morphological criteria to filter out antecedents. However, there are frequent cases when elimination of possible referential links based on mismatches of morphological features may lead to erroneous conclusions. Barlow (1998), for instance, presents examples when gender concord between a pronominal anaphor and a common noun antecedent seems to be unobserved (*Su Majestad suprema… él,*[4] in which the antecedent is a feminine NP and the anaphor – a masculine pronoun; in English *his supreme*

[4] In Spanish: <genderless possessive pronoun> *supreme Majesty* (feminine noun) … *he*.

Majesty… he, displays no such problem because English nouns do not have genders). Also, most languages acknowledging gender distinction have a number of nouns or phrases that can be referred to by both masculine and feminine pronouns, according to the natural gender of the person designated (*le docteur… elle*; in English *the doctor… she*). Though we do not share Barlow's view in this respect, namely that morphology should be ignored in AR, a less categorical approach with respect to a filtering rule based on morphology is preferable.

b. syntactical features:

- full syntactic description of REs as constituents of a syntactic tree (Lappin & Leass, 1994; Hobbs, 1978);
- marking of the syntactic role for subject position or obliqueness (the subcategorisation function with respect to the verb) of the REs, as in all centering based approaches (Grosz *et al.*, 1995; Brennan *et al.*, 1987), syntactic domain based approaches (Chomsky, 1981; Reinhart, 1981; Gordon & Hendricks, 1998; Kennedy & Boguraev, 1996);
- quality of being adjunct, embedded or complement of a preposition (Kennedy & Boguraev, 1996);
- inclusion or not in an existential construction (Kennedy & Boguraev, 1996);
- syntactic patterns in which the RE is involved, that can lead to the determination of syntactic parallelism (Kennedy & Boguraev, 1996; Mitkov, 1997);
- the quality of being in an apposition or a predicative noun position.

c. lexico-semantic features:

- lemma;
- person;[5]
- name (for proper nouns);
- natural gender;
- the part-of-speech of the head word of the RE. The domain of this feature contains: zero-pronoun (also called zero-anaphora or non-text string), clitic pronoun, full-flagged pronoun, reflexive pronoun, possessive pronoun, demonstrative pronoun, reciprocal pronoun, expletive "it", bare noun (undetermined), indefinite determined noun, definite determined noun, proper noun (name);[6]
- the sense of the head word of the RE, as for instance, given by a wordnet;[7]
- position of the head of the RE in a conceptual hierarchy (hypo/hypernymy) as in all models using wordnets (Poesio *et al.*, 1997; Cristea *et al.*, 2002a).

---

[5] Since, among the nominal REs, only pronouns can distinguish the person, for our purposes person is a lexical feature.

[6] As mentioned already, this classification takes into account only nominal anaphors, therefore ignoring verbal, adverbial, adjectival, etc. (Mitkov, 2002).

[7] We prefer to use wordnet as a common noun when we refer to any language variant (Vossen, 1998; Tufiş & Cristea, 2002a) of the original American English WordNet (Miller *et al.*, 1993).

Features as animacy, sex (or natural gender) and concreteness could be considered simplified semantic tags derived from a conceptual hierarchy;
- inclusion in a wordnet synonymy class;
- semantic roles, out of which selectional restrictions, inferential links, pragmatic limitations, semantic parallelism and object preference can be verified.

d. positional features:
- offset of the first token of the RE (an NP) in the text (Kennedy & Boguraev, 1996);
- inclusion in an utterance, sentence or clause, considered as a discourse unit (Azzam *et al.*, 1998; Cristea *et al.*, 1998). This feature allows, for instance, calculation of the proximity between the anaphor and the antecedent in terms of the number of intervening discourse units.

e. other features:
- inclusion or not of the RE in a specific lexical field, dominant in the text (this is called “domain concept” in (Mitkov, 1997));
- frequency of the term in the text (Mitkov, 1997);
- occurrence of the term in a heading (Mitkov, 1997).

The second component of a model is a **set of knowledge sources** intended to fetch values from the text to the attributes of the PS. A knowledge source is a virtual processor able to fill in values for one single attribute on the projection layer. Depending on the application the AR-engine is coupled to, as well as on the format of the input, sometimes more than just one such virtual processor could be served by one NLP processor. Thus, a morpho-syntactic tagger usually serves several knowledge sources as it can provide at least lemma, grammatical number and gender, case, person and part of speech of the head word of the RE (Brill, 1992; Tufiş, 1999). An FDG (functional dependency grammar) parser (Järvinen & Tapanainen, 1997) fetches the syntactic role of the RE, while wordnet access functions can bring all the headword senses (or synsets), and their position in a conceptual hierarchy. If word sense disambiguation (WSD) is available as a knowledge source, then the exact word sense of the head-word in the corresponding context can be determined. The membership of an RE to a certain segment can be the contribution of a discourse segmenter or a syntactic parser.

The third component is **a set of matching rules and heuristics** responsible to decide whether the PS corresponding to an RE introduces a new DE or, if not, which of the existing DEs it evokes. This set includes rules of the following four types:

- **certifying rules**, which if evaluated to 'true' on a pair (PS, DE), certify without ambiguity the DE as a referent of the PS. For instance, coreference based on proper name identity could be implemented, in most application settings, by a certifying rule;

- **demolishing rules**, which rule out a possible DE as referent candidate of a PS (and, therefore, of its corresponding RE). These rules lead to a filtering phase that eliminates from among the candidates those discourse entities that cannot possibly be referred to by the RE under investigation. The order of application of certifying and demolishing rules is specified in the model through priority declarations;
- **promoting/demoting rules** (applied after the certifying and demolishing rules), which increase/decrease a resolution score associated with a pair (PS, DE). The evaluation of these rules allows the run of a proposing/evoking phase, in which either the best DE candidate of a PS is chosen from the ones remained after the demolishing rules have been applied, or a new entity is introduced. The use of promoting/demoting rules can be assimilated with the preferences paradigm, employed by many classical approaches;
- a special section of the third component is dedicated to **attribute filling rules**, which are activated each time a new DE is proposed. These rules, behaving similar to the certifying ones, are responsible for the setting of anaphoric relations of a functional type. Each such rule receives as parameters: the name of an attribute (a functional relation), and a pair (DE1, DE2), in which DE1 is the current DE and DE2 is a DE previously introduced. If a matching is verified, that attribute of DE1 mentioned as the rule's first parameter, receives as value the identifier of DE2.

Finally, the fourth component is **a set of heuristics that configure the domain of referential accessibility**, establishing the order in which DEs have to be checked, or certain proximity restrictions. For instance, if we want to narrow the search for an antecedent to a vicinity of five sentences (or discourse units) with the intent to reduce the resolution effort on the basis that the great majority of the anaphors can find an antecedent within this range, e.g. (McEnery *et al.*, 1997), then the fourth component of the model will record that only those DEs linked with REs belonging to the last five discourse units are considered. Not the least, the domain of referential accessibility can model a linear search back order (Mitkov, 2000), or a hierarchical search back order on the discourse tree structure. Figures 4 and 5 display an example of a domain of referential accessibility for the linear case, respectively the hierarchical case. Figure 4a shows a case when RE_a evokes DE_a and RE_b evokes DE_b. Then the order to search the candidate referents for PS_c (projected from RE_c) is DE_b first, then DE_a. If a match between PS_c and DE_a is found (Figure 4b) then, for a subsequent RE_d, the order to search the candidate referent matching the correspondent PS_d is DE_a first, then DE_b (Figure 4c). If, instead, hierarchical order is preferred, considering that RE_a, RE_b and RE_c belong to three adjacent discourse units whose vein structure (Cristea *et al.*, 1998, 2000) is the one depicted in Figure 5 in bold line,[8] then the order to consider the candidate referents for PS_c (projected from RE_c) is DE_a first and DE_b after, since, hierarchically, RE_a (and therefore its corresponding DE_a) is closer to RE_c than RE_b (and its corresponding DE_b).

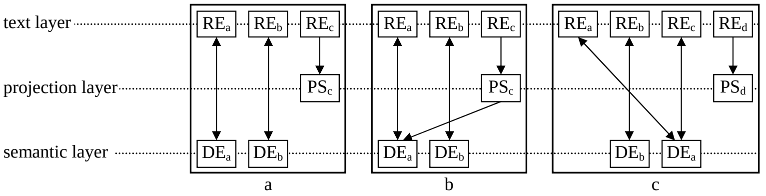


Figure 4: *Linear search order*

In certain cases, it could be of help to see the domain of referential accessibility as dynamically scaled on the type of the anaphor. A synthesis done by Mitkov (2002: 24) evidences that demonstrative anaphors find their antecedents more distantly than pronouns, while this distance could be even greater in the case of definite nouns and proper nouns. Rules of this kind could be included in the fourth component of the AR-engine.

The framework is language independent, in the sense that the adjustment to one language or another consists in defining a specific set of attributes, establishing the language specific knowledge sources capable to fill them and devising evoking heuristics/rules specific to each language. The domain of referential accessibility is thought to be stable to language change.

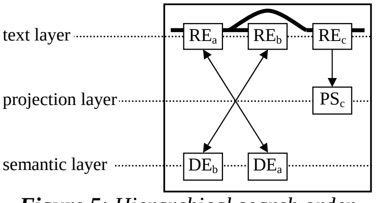


Figure 5: *Hierarchical search order*

2.3 Processing anaphors with AR-engine

Figure 3 depicts the main processing stream of AR-engine. The fundamental assumption is that anaphors should be resolved in a left-to-right order (in left-

[8] The vein expression of an elementary discourse unit (*edu*) *u*, following Veins Theory, is a sequence of *edu*s, proceeding, including and following *u*, which account to the minimal coherent sub-discourse focused on *u*. The gray lines in Figure 5 exemplify a situation in which RE_b, the linearly most recent RE from RE_c, is short-circuited by the vein expression of the *edu* RE_c belongs to, which means that RE_a is more pregnant in the reader's memory than RE_b when RE_c is read.

to-right reading languages) and vice versa in right-to-left reading languages. This way, the linear processing done by humans while reading, from the beginning of the text to its end is mimicked. At any moment during processing, just one RE is under investigation, which we will call – **the current RE**. As the current RE is momentarily the last one on the input stream, all resulting activity is performed against DEs already existent and, therefore, all found relations will point towards the beginning of the text. One processing cycle of the engine deals with the resolution of one RE and develops along three compulsory phases and an optional one.

The first (mandatory) phase is the **projection phase** when a PS (called the current PS) is build on the projection layer, using the information centred on the current RE obtained from the text layer with the contribution of the available knowledge sources.

The second (mandatory) phase, **proposing/evoking**, is responsible for matching the current PS towards one DE, either by proposing a new discourse entity or by deciding on the best candidate from the existent ones. This process involves first running the certifying and demolishing rules (if available), followed by the promoting/demoting rules. In the end, either an existent DE is firmly identified by a certifying rule, or matching scores between the current PS and a class of referent DEs are computed. Based on these scores, three possibilities can be judged:

1. all candidate DEs range under $\mathtt{threshold}_{\mathtt{min}}$, a parameter of the engine in the range 0 to 1: the interpretation is that none of the preceding DEs is sufficiently convincing as a referent for the current RE, and therefore a new DE is build. Each time a DE is created, a relation (`type-of`, `is-part-of`, etc.) is searched for between the new DE and previous DEs in a certain length window. Responsible for this activity are the **attribute-filling rules**;
2. the best rated scores are above $\mathtt{threshold}_{\mathtt{min}}$, but in the $\mathtt{threshold}_{\mathtt{diff}}$ range (a parameter usually less than 0.1) more than one candidate is placed: this situation should be interpreted as a lack of enough evidence to firmly consider one referent (the one scored the best) as the selected candidate. Consequently, the decision to choose a referent is postponed in order to allow following resolutions to bring supplementary clues to the resolution of the current RE, and the postponed corresponding PS is left on the projection layer;
3. the best score rated above $\mathtt{threshold}_{\mathtt{min}}$ and there is no other score under it in the $\mathtt{threshold}_{\mathtt{diff}}$ range: the interpretation is that the corresponding candidate individualises itself strongly among the rest of DE candidates. It will be confirmed as the referent and any of the preceding REs of the current RE, which correspond to the identified DE, should be considered antecedents of the current RE.

In the third compulsory phase, the **completion phase**, the data contained in the resolved PS is combined with the data configuring the found referent, if such a DE has been identified or, simply, the PS content is copied onto the newly build DE if none of the already existing DEs has been recognised. The resolved PS is afterwards deleted from the projection layer since any information that it used to capture can now be recuperated from the DE. So, to give an extreme example, if for some reason a model chooses to look for previous syntactic patterns of chained REs, they can be found on the semantic level. Although apparently contradictory to the "semantic" significance of the layer, this behaviour can mimic the short-term memory that records information of value for immediate anaphoric resolution.

Finally, the optional **re-evaluation phase** is triggered if postponed PSs remained on the projection layer at a former step. The intent is to apply the matching rules again on all of them. Humans usually resolve anaphors at the time of reading, but sometimes decisions should be postponed until the acquisition of complementary information adds enough data to allow a disambiguation process. Cases of postponed resolution will be discussed in Section 7.2. At the end of processing, each RE should record a link towards its corresponding DE and each DE should record a list of links towards its surface REs.

As we shall see in Sections 3 to 6, when referential relations different than strict coreference are to be revealed, DE attributes, which are not directly triggered from the corresponding PSs, appear as necessary. As mentioned at item 2 of the proposing/evoking phase, a section dedicated to actions to be performed for the filling-in of specific attributes following a proposing action is opened in the third component of the framework – the one dedicated to rules and heuristics.

In the following examples, we will mark REs by italic letters (as *a car*) and their corresponding DEs by a paraphrasing text in bold fonts and within square brackets (as [**the car**]). The following sections will analyse, within the AR-engine framework, a set of AR cases, usually considered difficult to interpret.

The discussion intends to evidence specific difficulties inherent to a large range of anaphoric phenomena, to imagine solutions in terms of an AR model, by indicating knowledge sources and rules/heuristics capable to deal with the identified tasks and to informally appreciate the tractability of these solutions. The discussion remains under the universal panacea for all the failures in AR, world knowledge (WK).

3 Relations triggered by positional and/or syntactic constraints

3.1 Nested referential expressions

(1) *the University building*
(2) *Amenomphis the IVth's wife*
(3) *the face of the beautiful queen*

In constructions of these types, two included (nested) REs are involved. They refer to two distinct DEs, which are linked by a certain relation. In (1), the two DEs are [**the University building**] and [**University**], where [**the University building**] `belongs-to` [**University**]. In (2), between [**Amenomphis the IVth's wife**] and [**Amenomphis the IVth**] a variant of the `belongs-to` relation holds, perhaps a `commitment`. In (3), between [**the face of the beautiful queen**] and [**the beautiful queen**] a still different type of `belongs-to` relation holds, perhaps a `is-part-of` relation. In all cases, the possessed object (or the part) corresponds to the outer RE while the possessing entity (or the whole) corresponds to the inner RE on the surface string. The incremental type of processing, including surface string parsing, and the included pattern of the REs allow that processing of the possessing entity (corresponding to the inner RE) be performed before the possessed entity (corresponding to the outer RE). If RE1 is nested on RE2 on the text layer, a knowledge source should fetch the value `RE1` to a `nesting` slot of the PS corresponding to RE2. On DE2 of the semantic layer, this slot will later on be transformed, by an attribute-filling rule, into a `belongs-to` (or some variation of it) attribute indicating the DE corresponding to RE1. Other constructions where a `belongs-to` or variations of it are correctly included are:[9] (*the center of* (*the hall opposite the big telescreen*)), (*emblem of* (*the Junior Anti-Sex League*)), (*one of* (*the middle rows*)), (*one of* (*them*)), (*one of* (*the novel-writing machines*)). In some cases the rule should be applied recursively: (*the waist of* ((*her*) *overalls*)), (*the shapeliness of* ((*her*) *hips*)). However, in expressions like: (*the hall opposite* (*the big telescreen*)), (*preparation for* (*the Two Minutes Hate*)), (*some mechanical job on* (*one of the novel-writing machines*)), (*a bold-looking girl, of* (*about twenty-seven*)), (*the girl with* (*dark hair*)), the relation between the two constituents are different than `belongs-to` or its variations. Our refinement of the types of relations to consider did not go so far. Moreover, a demolishing rule should always prevent a coreference relation between the DEs corresponding to the two REs.

[9] From G. Orwell's "1984".

3.2 Apposition

(4) *Mrs. Parsons, the wife of a neighbour on the same floor*
(5) *Nefertiti, Amenomphis the IVth's wife*
(6) *Jane, beautiful girl, come to me!*
(7) *a sort of gaping solemnity, a sort of edified boredom*

An apposition usually brings supplementary knowledge on a discourse entity. Also according to other approaches (Mitkov, 2002), but in disagreement with the annotation convention of MUC-7 which sees the apposition as one RE and the pair of the two elements as another RE, we consider the two elements of the apposition as different REs. In the model that we have built, the type of relation linking the two REs obeys the following heuristic: definite determined NP, genitival appositions and undetermined NP, as in (4), (5) and (6) yield coreferences, whereas indefinite noun appositions as in (7) yield `type-of` relations between the DE corresponding to the second RE towards the DE corresponding to the first RE. Let RE2 be an apposition of RE1 on the text level. We will suppose a knowledge source capable to apply syntactic criteria in order to fetch a `apposition-of=RE1` slot attached to PS2. As PS1 should have matched a DE1 the moment PS2 is being processed, a certifying rule must unify PS2 with DE1, in case RE2 is a definite determined NP, undetermined NP or a genitival construction. As a result, DE1 will accumulate all the attributes of PS2. Examples of cases correctly interpreted following this strategy are:[10] (*Emmanuel Goldstein*), (*the Enemy of the People*); (*the primal traitor*), (*the earliest defiler of the Party's purity*). If the apposition is an indefinite determined NP, a demolishing rule will rule out as a possible antecedent the argument of the `apposition-of` attribute in the current PS. As a consequence, the usual proposing/evoking mechanism will work, finalized in finding a target DE. Then, only if the found DE is new, a rule in the attribute-filling section of the set of rules/heuristics will exploit the `apposition-of=RE1` slot attached to PS2 in order to transform it into a `type-of=DE1` value. This strategy will correctly interpret an apposition like (*a narrow scarlet sash*), (*emblem of the Junior Anti-Sex League*). Unfortunately, the knowledge source responsible to detect appositions can easily go into errors. This is the case when apposition is iterated over more than just two adjacent constituents: (*the most bigoted adherents of the Party*), (*the swallowers of slogans*), (*the amateur spies*) and (*nosers-out of unorthodoxy*); (*a man named O'Brien*), (*a member of the Inner Party*) and (*holder of some post so important and remote*), where clear criteria to disambiguate from enumerations or from

[10] From G. Orwell: “1984”.

indications of locations (as in (*the same row as Winston*), (*a couple of places away*)), the only two types of exceptions found so far matching the patterns of our apposition-finding knowledge source, are difficult to devise.

3.3 The subject – predicative noun relation

(8) *Maria is the best student of the whole class.*
(9) *John is a high school teacher.*
(10) *Your rival is a photo.*
(11) *The young lady became a wife.*

Supposing RE1 is the subject and RE2 is the predicative noun, a knowledge source of a syntactic nature should be able to fetch a `predicative-noun-of=RE1` attribute into the PS2 corresponding to the predicative noun RE2. Definite determined predicative nouns as *the best student of the whole class* in (8) are, in our model, considered coreferential with the subject. The resolution should aim at injecting into the DE [**Maria**] the information brought by the predicative noun RE2, and temporarily stored on PS2. Suppose the DE [**Maria**] is something of the kind: [`name="Maria", sem=person1, Ngen=fem, num=sg`], where `person1` is the first sense of the word *person* according to WordNet. Then, the fact that she is seen now also as a student must not affect any of the attributes `name`, `Ngen` (natural gender) or `num` (grammatical number) but instead add into the description an attribute `lemma=student` (if only the head of the RE is considered in the representation, or a more sophisticated description if the constituents are also kept: *the best of the whole class*), and replace the `person1` value of the `sem` attribute with a more specific one: `student1`.[11] When the predicative noun is an indefinite NP, as in (9), our model interprets it as the semantic type of the subject. The more general concept is replaced with a more specific one both when a concept is predicated as a more specific one (*the animal is an elephant*) as well as when the reverse predication holds (*the elephant is a heavy animal with a trump*). Other examples of the same kind are:[12] (*one of them*) *was* (*a girl*); (*she*) *was* (*a bold-looking girl, of about twenty-seven*); (*who*) *were* (*the most bigoted adherents of the Party*); (*the other person*) *was* (*a man named O'Brien*); (*O'Brien*) *was* (*a large, burly man*); (*she*) *might be* (*an agent of the Thought Police*).[13]

Conceptual hierarchies like WordNet can help to identify, in examples like (10), that *a photo* (an object) cannot be a type for [**the rival**] (hyponym of a person, according to WordNet). On the contrary, to find out that *a photo* is a

[11] The implicit assumption here was that WSD capabilities were used as a knowledge source.
[12] From G. Orwell's "1984".
[13] The present model does not implement specific criteria to deal with modalities.

substitute for *the person faced in the photo* necessitates deep WK. To offer a substitute of a solution in cases like that, a generic relation like `metaphoric-type-of` can be adopted.

The solution we adopted for representing discourse entities subject to time changes, *different* than the one proposed in MUC-7 (Hirschman & Chinchor, 1997), is described in (Cristea & Dima, 2001): we have linked entities as the ones in example (11) with the `same-as` relation, triggered by the occurrence of the interposed predicate *become*.

In all cases (8) to (11), a complication arises when the resolution of RE1 (the subject) was *postponed* to the moment RE2 (the predicative noun) is processed.[14] If this happens, either the unification makes PS2 coreferential with the postponed PS1, or the semantic relation is established between the current proposed DE and the postponed PS1. Later on, when the postponed PS is lowered at the semantic level, these relations are maintained.

4 Lemma disagreement of common nouns

4.1 Common NPs displaying identical grammatical number but different lemmas

(12) *Amenomphis the IVth's wife … the beautiful queen*

The discovering of the coreference relation in this case should mainly be similarity-based. In principle, a queen should be found more similar to a wife then to a pharaoh, supposing Amenomphis is known to be as such. If, instead, this elaborate *knowledge* is not available, and all that is known about Amenomphis, as contributed by a name-entity recogniser knowledge source, is his quality of being a man, the moment *the beautiful queen* is processed, a queen should again be found more similar to a wife than to a man. Many approaches to measure similarity in NLP are already known and some use wordnets (e.g. (Resnik, 1999)). When a sense disambiguation procedure is lacking, then a wordnet-driven similarity that counts the common hypernyms of all senses of the two lemmas could be a useful substitute in some cases.[15] Still, criteria to decide similarity are not elementary and a simple intersection of the wordnet hypernymic paths of the anaphor lemma and the candidate antecedent lemma often does not work. The following is an example of a chain of erroneous coreferences found on the basis of this simplistic criteria: *the centre of the hall opposite the big telescreen | his place | some post so important and*

[14] The same is true for apposition.

[15] There is good reason to believe that such an approach is successful when lexical ontologies, as fine graded in word senses as WordNet, are used. This criterion is based on the assumption that senses displaying common ancestors must be more similar than the ones whose hierarchical paths do not intersect.

remote | the back of one's neck | a chair | places away | the end of the room | the protection of his foreign paymasters.[16]

Sometimes, a useful criterion for the identification of coreferential common noun REs with *different* lemmas could be the natural gender (*queen* and *wife* are both feminine in natural gender). In other cases, the antecedent could be recuperated by looking at the modifiers of the head nouns. Consider example (13):

(13) *the most beautiful women… those beauties*

A *promoting* rule should be able to confront the lemma *beauty* with modifiers of the head *women* in the DE for [**the most beautiful women**].

4.2 Common NPs with different grammatical number and different lemmas

(14) *a patrol … the soldiers*
(15) *the government… the ministers*

According to WordNet, in two out of three senses, a patrol is a group and, in one sense out of four, government is also a group. This suggests to fill-in a `sem=group` feature if the `group, grouping -- (any number of entities (members) considered as a unit)` synset is found on a hypernymic path of the lemma of a candidate antecedent of the plural NP (see examples (14) and (15)). However, this criterion could prove to be weak because many words have senses that correspond to groups (*a garden*, for instance, has a sense that means a group of flowers, and in a text like *A patrol stopped by the garden. The soldiers…* there is high chance to find *the soldiers* coreferring to [**the garden**] rather than to [**the patrol**]). Different criteria should be combined to maximize the degree of confidence, among which a similarity criteria, for instance based on wordnet glosses (as in `forest – the trees and other plants in a large densely woodened area`) or on meronymy, (as in `flock – a group of sheep or goats – HAS MEMBER: sheep – woolly usu. horned ruminant mammal related to the goat`), or even the simple identification of antecedents within a fixed collection of collective nouns, as suggested in (Barbu *et al.*, 2002). In principle, this case is similar to the preceding one if an attribute of being a group is included in the representation of the DE referent.

4.3 Common nouns referring proper nouns

(16) *Bucharest… the capital*

[16] From G. Orwell's "1984".

There are no other means to solve this reference than enforcing the labelling of *Bucharest*, in its corresponding DE, the very moment when it is processed, with, for instance, a `city1` value of a `sem` attribute. If this labelling information is available, fetched by a name-entity recogniser, then the framework processes the reference the same way it does with common nouns with different lemmas.

5 Number disagreement

5.1 Plural pronouns identifying split antecedents

(17) *John waited for Maria. They went for a pizza.*

Despite the opinion of other scholars on the matter (see, for instance, (Eschenbach *et al.*, 1998)) we do not think that, during the interpretation of (17) above, a discourse entity for the group [**John, Maria**] must have been proposed, as soon as the referential expression *Maria* is parsed. Or else, we have to face a very uncomfortable indecision regarding what groups to consider and when. The mentioned group is seen as a DE only because at a certain moment, as the text unfolds, an anaphor coreferring to it appears: *they*. In (18) below, there is no need for such a group representation, as the reader is perhaps not conscious of its existence:

(18) *John waited for Maria. He invited her for a pizza.*

Neither vicinity in the location space of the story, nor textual vicinity or framing in a wording pattern are a sufficient constraining criteria for proposing groups on the semantic layer, see examples (19) and (20):

(19) *John was in New York when Maria wrote him that she finally made up her mind. They got married the next month.*

(20) *John finished his classes. He went to a football match. As it was a rainy day, no more than 10 people were on the stadium. Maria happened to be there too. They went for a pizza and one month later got married.*

To make life even harder, note that in (20) 12 people are candidates for different groups of persons ([**John, 10pers**], [**10pers, Maria**], [**John, 10pers, Maria**], [**John, Maria**] or only [**10pers**]). Nevertheless, the reader has no difficulty to identify *they* with the group [**John, Maria**]. But why not to attach to the group also [**John's classes**], [**the football match**], [**the rainy day**] or [**the stadium**]? The obvious WK-based answer is: because none of the others can go for a pizza! And also because getting married is an occupation for exactly two people! But this is deep WK and, as agreed, we would not want to rely on it.

From the discussion above, we know that group formation is triggered by a first reference to it. A group, unless it is verbalised to as such in the text, does not exist until it is referred to. Still, two questions remain: how much we can do in the absence of WK for the group content identification, and what are the criteria to trigger the creation of group DEs, therefore by what means a plural pronoun is considered as referring to a group. The answer to the first question stays again in the use of similarity measures (common association basis in (Eschenbach *et al.*, 1998)) to identify members of groups in the text preceding the plural pronoun. As for the second question, the framework policy is to propose new DEs when no match between the current PS and the preceding DEs rises above $\mathtt{threshold}_{\mathtt{min}}$. This policy is good enough for our purpose as long as no plural DEs, toward which the plural anaphor could match, are in the recent proximity. If an ambiguity arises, then the second framework policy to postpone resolution until sufficient discrimination criteria leaves a unique candidate within a $\mathtt{threshold}_{\mathtt{diff}}$ range is well suited again. The combination of these two policies in example (21) below, for instance, would maintain the indecision whether *they* should corefer to [**John, Maria**] or to [**the classes**] as long as no WK is available to state that only people can go for a pizza, and this should be a correct behaviour.

(21) *John waited for Maria when the classes were over. They went for a pizza.*

5.2 Plural nouns identifying split antecedents

Supplementary to the problems identified above, when the anaphor is a noun, the similarity criteria found to characterize the group should extend to the anaphor as well. Consider the following example:

(22) *Athos, Porthos and Aramis … the musketeers*

The similarity criteria sketched above yields `person, individual, someone, somebody, mortal, human, soul – (a human being)` as the WordNet concept characteristic to the discovered group, while the word *musketeer* means also a `person`. As such, there is enough evidence to conclude that a DE [**the musketeers**] should be proposed that points to each of the DEs [**Athos**], [**Porthos**] and [**Aramis**] as members. As already discussed in Section 2.3, the decoration of existing DEs with attributes different than those inherited from the PS it evolves from, in our case the completion of the DE [**the musketeers**] with an attribute `has-as-element=<x,y,z>`, with `x`, `y`, `z` being identifiers of the DEs [**Athos**], [**Porthos**] and [**Aramis**], is an action characteristic to the attribute-filling rules.

6 Bridging anaphora

6.1 Elements-to-set references

(23) *all the weapons for the underwater hunting… the masque… the rifle… the ribbon paws*

In this example, to each of the REs *the masque*, *the rifle*, and *the ribbon paws* must correspond a proper DE. Moreover, in a proper representation, each of them must contain an attribute `is-element-of` pointing to the DE [**the weapons for the underwater hunting**]. *The rifle* against [**the weapons…**] is the only relation of this kind that can be easily inferred based on a similarity computation. A masque and a paw are not in themselves weapons, although the context helps to acquire this interpretation. Only reasoning on deep WK would allow for such assignments. If, however, this kind of WK is available, assigning the `is-element-of` links from all component DEs towards the DE [**the weapons…**] is also an action characteristic to the attribute-filling rules. Suppose now a case in which, between two coreferring anaphors, a set to which the corresponding entity belongs is mentioned, like in *<u>John</u> and Mary decided <u>they</u> should go to the party, in this order: Mary first, <u>John</u> after* (only *John* and the group mentioning pronoun are underlined, although the same is also true for *Maria*). The `is-element-of` relation between the element DE and the group DE cannot be established because the element DE is build before the group DE. **[John]** is build before the group identified by **they=[John and Mary]**. However, this relation can be inferred as the inverse of an already acquired `has-as-element` relation, supposed to have been filled between the group DE and the element DE the moment the group was mentioned, and on the basis of a genuine coreference relation established between the second mention of the element and its corresponding DE representation.

6.2 Hidden discourse entities

(24) *When I got into the room I saw a strange screen saver on the big monitor. The other computer was off.*

Interesting debates could arise around this example. Any human person reading this text is aware of the existence of two computers in the mentioned room: one with a big screen, on which a strange screen saver was running, and another one which was off. One question is whether both computers should be represented on the semantic layer or only [**the other computer**]. Since the mentioning of *the other computer* doesn't make sense, but if [**some (first) computer**] exists, this can be taken as an implicit mentioning of the first computer. However there is no RE in the text explicitly referring to this DE, excepting from *the big monitor*, which is interpreted as part of this computer.

But a representation for a [**some (first) computer**] entity cannot appear the moment *the strange screen*, a part of it, is mentioned, because otherwise we see no reason why to consider only the `is-part-of` relation and to neglect others like `made-of`, spatial relations like `laying-upon`, etc. There is no end to describe all objects to which a certain mentioned object could consciously interact. For instance, in some reader's mind at least the image of a table on which one or both computers lay is present. A saver solution (at this level of automatic reasoning which is insinuated by our framework) is to consider as candidates for being represented on the semantic layer strictly those objects that are explicitly mentioned in the text. If a more elaborated resolution model is to be attached on top of the work performed by the AR-engine, then those hidden DEs should be put into evidence through an inference mechanism, which is not supported by the current level of processing.

What the engine would have to do in the case of example (24) is to build a DE corresponding to the RE *the big monitor* and another DE for the RE *the other computer*. No relations link these representations.[17] On the contrary, in a sequence like the one in example (25) the DE [**the computer**] should display a `has-as-part` relation towards the DE [**the big monitor**].

(25) *When I got into the room I saw a strange screen saver on the big monitor. The computer was left open by my colleague.*

In this example, an attribute-filling rule must be responsible for filling-in a value of a `has-as-part` attribute. The difference between examples (24) and (25) is that in (24) the method should prevent from retaining, as the value of the attribute `has-as-part` of the DE [**the other computer**], the identifier of the DE [**the big monitor**], while in (25) it should mainly go for it.

7 The resolution moment

7.1 Resolution in the case of cataphora

A rather controversial anaphoric phenomenon is cataphora, which is said to arise “when the reference is made to an entity that is mentioned subsequently in the text” (Mitkov, 2002). In our terms, a cataphoric relation is given by a pair of coreferring mentions in which the first one introduces the referent and is information-poorer than the subsequent one. The only cases that merit a special attention are those defined as ‘first-mention’ cataphora (Mitkov, 2002) or ‘backwards anaphora’ (Carden, 1982), like the one in the following text placed at the beginning of O. Wilde’s “The Picture of Dorian Gray”:

[17] Behavior as sophisticated as simultaneously projecting two PSs from an RE as *the other x*, which would allow for the identification of two objects of the type [**x**], is not currently implemented in the AR-engine framework.

(26) *"From the corner of the divan of Persian saddle/bags on which <u>he</u> was lying, smoking, as was <u>his</u> custom, innumerable cigarettes, <u>Lord Henry Wotton</u> could just catch the gleam of the honey-sweet blossoms of a laburnum…"*

In cases where a pronoun precedes a noun but the text contains an earlier more informative mention of the same entity, also in accordance with other scholars (see, for instance, an analysis done by Tanaka (2000)), the pronoun should be resolved against the preceding text as in ordinary anaphora.

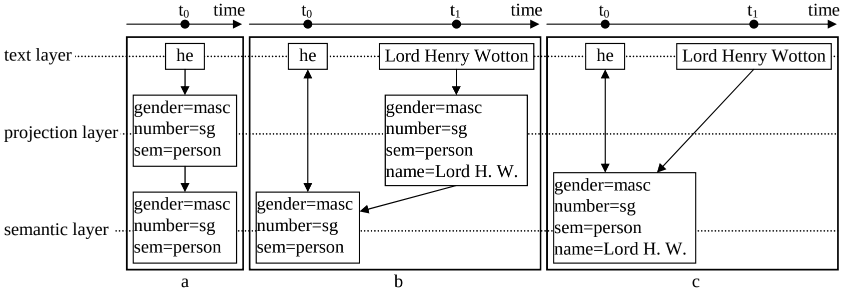


Figure 6: *A cataphora resolution example*

The view we have on this topic is that once a linear processing model, from the beginning of the text to its end, is adopted when reading the cataphoric referential expressions, there is no way in which one would look towards the end of the text in order to recuperate a referred entity. Consequently, the moment the pronoun is read/processed, a poorly decorated discourse entity must have been introduced into the state of mind of the reader, and subsequent coreferring expressions evoke this entity, eventually adding new features to it. As remarked on Section 2.3, the linear (incremental) processing hypothesis also implies that the anaphoric relation should always be projected on the text axis towards the beginning of the text. At the moment of reading/processing the pronoun *he* in the example above, first a PS is projected. Then this is immediately lowered to the semantic layer as a proposed DE. This moment is marked $\mathbf{t_0}$ on Figure 6a, and the corresponding semantic representation could not contain more features than those contributed by genuine morphology (gender and number) and a semantic feature of being a person. As the text unfolds, at a later moment $\mathbf{t_1}$, Lord Henry Wotton is processed and a PS containing morpho-semantic features, as suggested by Figure 6b, is proposed. As this feature structure strongly matches (in `gender`, `number` and `sem`) the previously created DE, the evoking phase will most probably indicate it as the referent. Then, during the completion phase, the `name` feature will enrich the original DE, introduced by the pronoun (Figure 6c).

7.2 Postponed resolution

The mechanism of postponed resolution that AR-engine incorporates allows solving of otherwise intractable cases. Consider example (24):

(27) *No one knew who was the driver who drove up the actor home that night. Later on, everybody found out that this was the best driver Hollywood ever had.*

The moment *this* is to be resolved, there is no sufficient knowledge in order to decide whether it refers [**the person**], or [**the actor**] or even [**that night**]. However, immediately after the predicate noun *the best driver* is read, two things happen: a) *the best driver* is found to refer [**the driver**], a DE already introduced, and b) the predicative noun should corefer to the subject (see Section 3.3). So, from the fact that the predicative noun *the best driver* is coreferential with the subject *this*, and the same *the best driver* is resolved against the DE [**the driver**], it can be inferred the recognition of *this* as the same DE [**the driver**]. This is a postponed resolution and its completion is realised during the re-evaluation phase of the RE following it on the text level, as discussed in Section 2.3. In example (28), application of the same mechanism produces the recuperation of *this* as [**the actor**]:

(28) *No one knew who was the driver who drove up the actor home that night. And when you think that this was the actor that used to be in vogue not long time ago…*

8 Final considerations

8.1 Evaluation

To evaluate the proposed solutions, we have used four chapters, summing approximately 17,500 words, from the original English version of novel “1984” by George Orwell. The choice of a text belonging to the belletrist register, instead of the scientific or technical register, was justified by the intend to appreciate how frequently the mentioned cases occur in a free text and also how well are fitted the proposed solutions for the wide variety of types of referential expressions and anaphoric phenomena encountered there.

The text was first POS-tagged, then FDG-tagged and then manually annotated by a group of master students (by using the Palinka annotator (Orăsan, 2002)), for coreference. The annotation task did not contain a phase dedicated to markables, as they were extracted automatically from the FDG structure (all structures dominated by a head noun, from which clauses were removed). NP heads were also automatically marked. Our markables generally are conformant with the MUC-7 criteria (Hirschman & Chinchor, 1997), although ours do not include relative clauses, each term of an apposition is taken separately, and we have marked also *wh*- noun phrases. Some errors that the FDG parser makes and which inflict on the NP annotation were manually corrected. Four approximately equal parts were assigned to teams of two master students in Computational Linguistics. The students had to annotate their assigned parts individually. To simplify the annotation task, the annotators were instructed to mark only coreference relations. Agreement between pairs of annotators are in the range of 60% to 90%. After seeing the mismatches reported by a program, they had to negotiate common decisions. The document obtained after merging the final negotiated versions was considered the gold standard. To perform the evaluation, all cases of `belongs-to`, `type-of`, `is-part-of`, `has-as-part`, `is-element-of`, `has-as-element` and `same-as` relations were collected manually.

The model implemented at this stage of research was rather a simple one, since our focus was not so much on refining the coreference performance towards attaining or surmounting the 80% psychological limit, as to see whether feasible solutions for the investigated cases of wicked anaphora can be imagined. As such, the incorporated AR model contained only the following attributes: `lemma`, `number`, `pos`, `femaleName` (`YES`, if lemma is a female name), `maleName` (`YES`, is lemma is a male name), `familyName` (`YES`, if lemma is a family name), `HeSheItThey` (the probability of a noun phrase to be referred to by *he*, *she*, *it* or *they* pronouns), `includes` (containing a vector of REs Ids nested in the current RE, possible empty), `indefinite` (`YES` if the RE is an indefinite determined NP and `NO` if the RE is definite determined or undetermined), `predicateNameBE` (contains the Id of the subject when the current RE is a predicative noun of a form of the predicate *to be*), `predicateNameBECOME` (contains the Id of the subject when the current RE is a predicative noun of a form of the predicate *to become*), `apposition` (contains the Id of the RE towards whom the current RE is in an apposition relation), `SYNOMYMS` (the list of the WordNet synonyms of the lemma, no matter the sense), `HYPERNYMS` (the list of the hypernymic synset Ids in WordNet, no matter the sense), `MERONYMS` (the list of the has-parts synset Ids in WordNet, no matter the sense), `HOLONYMS` (idem, for the part-of relations). No syntactic attributes, other than predicative noun and apposition, were retained. The knowledge sources were implemented based on the following processors: a POS-tagger, an FDG parser, a very simple name-entity recognizer, and a WordNet navigator. The model includes 4 certifying rules, 2 demolishing rules, 5 promoting rules and 5 attribute-filling rules. The domain of referential accessibility considered is linear and the anaphors were searched within a distance of 10 sentences for coreference and 3 sentences for functional relations. Table 1 shows the dimension of the experiment.

	total	%
nested REs	1097	29.37
coreferential appositions	19	0.51
`type-of` appositions	9	0.24
coreferential subject-predicative noun relations	40	1.07
`type-of` subject-predicative noun relations	45	1.20
`same-as` relations	1	0.03
different lemmas	1115	29.85
group noun to split antecedents	1	0.03
pl. noun to split antecedents	4	0.11
pl. pron. to split antecedents	20	0.54
`is-element-of`	34	0.91
`is-part-of` and `has-as-part`	110	2.95
cataphorae	8	0.21
total REs	5522	
total DEs	3107	
total relations	3735	

Table 1: Dimension of the experiment

The total number of relations was computed by adding the number of coreferential relations (number of REs minus number of DEs) with the number of functional relations. The investigated phenomena amounted to 2/3 of the total number of anaphoric relations in the corpus (approx. 67%). The rest are genuine coreference relations.

Nested REs raised no problem, because the simple identification of this surface pattern yields a `belongs-to` relation or a variation of it. At this stage of research, no effort was devoted to improve different subtypes.

By far, the best results (precision and recall between 0.8 and 0.92) are obtained for predicative noun to subject relations, relatively easy to identify and catalogue as either coreference or `type-of` relations.

Recognition of `type-of` relations in case of appositions had also a good degree of success (0.8 precision and 0.88 recall). Bad precision was obtained for appositional coreferences, explained by the tendency of our external sources to classify also enumerations as appositions, rather by inappropriate decisions made in the resolution process itself. Still, a good recall of 0.94 was obtained in these cases. The difference in precision is explained by the scarcity of cases were terms of enumerations are expressed as indefinites.

We obtained very good precision but bad recall in cases of coreferences involving cataphora. Examples of failed resolutions of this type are: *'Do you think you could come across and have a look at <u>our</u> kitchen sink?'… The <u>Parsons</u>' flat was bigger than Winston's...* (for reasons of a too large distance in between: 218 interposed words in more than 12 sentences), *<u>someone</u>* … *<u>the children</u>* (intended disagreement in number), "*We didn't ought to 'ave trusted* *'em...That's what comes of trusting 'em...We didn't ought to 'ave trusted the buggers."* (the parser does not recognize *'em* as a pronoun), *"Take your places, please." Winston sprang to attention in front of the telescreen..."Take your time by me... Come on, comrades..."* (where, because of number confusion, the referent of the first occurrence of *your* is found to be the already existent [**Winston**] DE, which will furthermore prevent *comrades* to corefer it).

The only singular group noun to split antecedents example found in the corpus was correctly processed, but an optimistic conclusion here would be premature. Of the examined cases, the most frequent are found to be different lemmas coreferences. Our implemented model is still too weak to handle properly these anaphoric phenomena: a better similarities-valuing model is needed. The results of plural noun and pronoun referring split antecedents as well as the recognition of the `is-part-of` relations are in approximately the same range of precision and recall (32% - 58%). The following are examples of failed plural-noun-to-split-antecedents references: *"Of course it's only because Tom isn't home" said Mrs Parsons vaguely. The Parsons' flat...* (failure to discover the second occurrence of *Parsons* as a plural noun); *Oceania was at war with Eurasia and in alliance with Eastasia… the three powers* (no WordNet or other name entity help in identifying the names as state names); *Winston was dreaming of his mother. ... His father he remembered more vaguely … (Winston remembered …) .... The two of them must evidently have been swallowed up in one of the first great purges of the fifties* (only deep understanding of the context in which the reference is used could disambiguate *the two of them* as being the group of [**mother**] and [**father**] and not of [**Winston**] and [**father**], for example; also cardinality of groups as a restrictions feature is not yet in the model); *At this moment his mother was sitting in some place deep down beneath him, with his young sister in her arms... Both of them were looking up at him* (*both of them* could not be linked to the group [**mother**] and [**sister**] for the same reasons as above; this is also a good example of postponed evaluation, since only later, at the moment of reading *him*, one could decide that the group does not include also the person referred by *him* on the ground that a group cannot look at a member of it – WK); *an old man and an old woman … "We didn't ought to 'ave trusted 'em"* (153 in between words in 8 sentences, and two more persons mentioned). To note that the current model does not implement group nouns referring split antecedents when the split antecedents are nouns different in number, as in *Mary and her friends went to the cinema. They saw a good movie.*

Finally, the worst results were obtained for `is-element-of` relations. Here are some commented failures: *Victory Mansions were old flats… The Parsons'* *flat was bigger than Winston's* (lack of knowledge sources to recognize elliptical heads of genitival constructions, like *Winstons'*); *The sacred principles of Ingsoc*. *Newspeak*, *doublethink*, *the mutability of the past* (invented terms, not in English, impossibility to consult WN in order to detect `is-element-of` relations); *All their ferocity was turned outwards, against the enemies of the State, against foreigners, traitors, saboteurs, thought-criminals* (the still weak capacity of the model to recognize similarity: only WordNet hypernymic chains contribute).

An interesting example of failure to recognize the `has-as-part` relations is the following: *Both of them were dressed in the blue shorts, grey shirts, and red neckerchiefs which were the uniform of the Spies*. Here, *which* is a pronoun referring the DEs of the group of elements {[**the blue shorts**], [**grey shirts**], [**red neckerchiefs**]}. Although number neuter, when seen in isolation, this pronoun was found to be in plural, as the subject of the plural verb *were*. As a result, a new DE was proposed to represent the set of the three elements, and a relation `has-as-element` linking this DE with each of its members. Further on, there is a subject-predicative noun construction with a definite predicative noun: *which were the uniform of the Spies*, implying therefore a coreference relation. This will finally yield `has-as-element` relations between [**the uniform**] and each of its mentioned elements, instead of `has-as-part` relations (a short, a shirt and a neckerchief can be parts of a uniform, not elements of it). Perhaps WK is needed to correct this error.

8.2 Conclusions

Modern applications, especially those related to the semantic web, compel to apply combined and complex methods in NLP. These application environments require more and more sophisticated tools to be put to work and, where necessary, AR methods should be prepared to tackle also hard problems raised by the interpretation of free text.

The paper investigated cases of difficult AR problems and proposed a set of solutions within the framework of a general incremental AR solver, called AR-engine, previously introduced by Cristea and Dima (2001). The basic principles and architecture of the engine were presented. Our investigation went on cases of AR resolution that were not in focus in previous evaluation attempts (Cristea *et al.*, 2002a, 2002b), and where the evaluation was conducted on examples chosen by hand or reported by other authors to be difficult to tackle (Mitkov, 2001; Barbu *et al.*, 2002). This time, a corpus of continuous text taken from the belletrist register was used. We investigated four categories of anaphoric relations that, we believed, display an ascending degree of difficulty: coreference relations whose resolution could be triggered by positional (syntactic) constrains, coreference relations in which the anaphor and the antecedent are common nouns with disagreement in lemma, noun and pronoun anaphors displaying number disagreement with the antecedents, and bridging anaphora. For the first time, anaphoric references other than genuine coreferences were experimented with AR-engine. Using the framework, we discussed also two less studied situations of recuperation of referential links: the case of cataphoric references and situations when resolution cannot be accomplished synchronously with the reading moment of the anaphor.

The examples discussed in the paper revealed different degrees of difficulties. Consequently, the knowledge sources put on stage were also spread on a very large scale, from cheap, as a POS-tagger, capable to tag words with morphological features, to extremely expensive, like WSD, capable to infer word senses in context (however, our implemented model did not make use of a WSD knowledge source).

Due to the difficulty of organizing a large corpus annotated for such a large diversity of referential links, the dimension of the experiment was limited. The language under investigation was English. However, the framework is not restricted to one language. Language dependent expertise is incorporated in a model, which is a configurable component that should be plugged-into the engine. Also, any application specific behaviour, as for instance the type of references to identify, can be described into the model.

If infrequent cases require costly implementation solutions and costly computations, the effort is not justified. Instead, if a model can be easily updated to take into account also these cases with little difference in computation time, then the effort is worth doing. It is also worth questioning whether there exist an algorithmic optimisation solution, that call for expensive methods only when other cheaper methods proved to be inefficient. To take the coreference task as an example, expensive methods would have to be put to work only when cheap methods would have failed to point firmly an antecedent among more closely rated candidates. This behaviour can be easily added to the functionality of the AR-engine by adequately exploiting thresholds. A disambiguation decision between two candidates is usually taken when their computed scores are different on a certain threshold. Then, one could make this threshold larger if it was computed based on rules using cheap knowledge sources and narrower if it was computed on the basis of rules using expensive knowledge sources.

In this stage of the research, the interest was focused, on one hand, in enhancing the AR engine and, on the other hand, in devising rules and heuristics that integrated into a model, to foreshadow the feasibility of the expected solutions for the specific types of anaphora enumerated. Another goal was to neatly define the benefit that certain knowledge sources can bring for certain types of problems. Knowledge sources, as well as resources, should always constitute a configurable component in an AR task. A designer should be able to add or to remove to/from an AR engine any such knowledge source depending on their availability, the complexity of the task and the running constraints. In such a configurable setting, it should then be clear what behaviour to expect any time a "surgery" of this genre is operated.

Although it is perhaps too soon to draw conclusions related to the feasibility of the approach, we consider our results to be promising. The engine has reached a certain stability vis-à-vis the updates encumbered by the specific type of processing imposed by a large diversity of anaphoric phenomena. The most spectacular part of the research is only now ahead of us when the focus will be on the refinement of the incorporated model.

Acknowledgments: The authors are grateful to our master students in Computational Linguistics, Petronela Dumitraşcu, Corina Forăscu-Nicu, Maria Husarciuc, Monica Lupu, Delia Mihu, Daniel Pintilie, and Diana Trandabăț-Balaşa, who have pursued the annotation task and to the two anonymous reviewers for their comments that helped a lot to improve the final shape of the paper.

References

Azzam, S., K. Humphreys and R. Gaizauskas 1998. "Evaluating a Focus-Based Approach to Anaphora Resolution." *Proc. of the 17th Coling and the 36th Annual Meeting of the ACL (COLING-ACL'98),* Montreal, Canada, 74-78.

Barbu, C., R. Evans and R. Mitkov 2002. "A Corpus Based Investigation of Morphological Disagreement in Anaphoric Relations." *Proc. of Language Resources and Evaluation Conference (LREC 2002),* Las Palmas, VI:1995-1999.

Barlow, M. 1998. "Features Mismatches and Anaphora Resolution." S. Botley and T. McEnery (eds.) *New Approaches to Discourse Anaphora, Technical Papers*. vol. 11.

Brennan, S., M. Friedman and C. Pollard 1987. "A Centering Approach to Pronouns." *Proc. of the 25th Annual Meeting of the ACL (ACL'87).* Stanford, CA. 155-162.

Brill, E. 1992. "A Simple Rule-based Part of Speech Tagger." *Proc. of the 3rd Conference of Applied Natural Language Processing,* Trento, 152-155.

Carden, G. 1982. "Backwards Anaphora in Discourse Context." *Journal of Linguistics* 18.361-387.

Chomsky, N. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Cristea, D., N. Ide and L. Romary 1998. "Veins Theory: A Model of Global Discourse Cohesion and Coherence." *Proc. of the 17th Coling and the 36th Annual Meeting of the ACL (COLING-ACL'98),* Montreal, 281-285.

---------- and G.-E. Dima 2001. "An Integrating Framework for Anaphora Resolution." *Information Science and Technology,* Bucharest 4:3-4, 273-291.

----------, O.-D. Postolache, G.-E. Dima and C. Barbu 2002a. “AR-Engine – A Framework for Unrestricted Coreference Resolution.” *Proc. of Language Resources and Evaluation Conference (LREC 2002),* Las Palmas, VI:2000-2007.

----------, G.-E. Dima, O.-D. Postolache and R. Mitkov 2002b. “Handling Complex Anaphora Resolution Cases.” *Proc. of the Discourse Anaphora and Anaphor Resolution Colloquium,* Lisbon.

Deemter van, K. and R. Kibble 2000. “On Coreferring: Coreference Annotation in MUC and Related Schemes.” *Computational Linguistics* 26:4.615-623.

Eschenbach, C., C. Habel, M. Herweg and K. Rehkamper 1989. *Proc. of the 27th Annual Meeting of the ACL (ACL ’89),* 161-167.

Gordon, P. and R. Hendrick 1998. “The Representation and Processing of Coreference in Discourse.” *Cognitive Science* 22.389-424.

Grosz, B., A. Joshi and S. Weinstein 1995. “Centering: a Framework for Modeling the Local Coherence of Discourse.” *Computational Linguistics* 21:2.

Halliday, M. and R. Hassan 1976. *Cohesion in English.* London: Longman.

Hirschman, L. and N. Chinchor 1997. “MUC-7 Coreference Task Definition, version 3.0.” *MUC-7 Proc*. See also: http://www.muc.saic.com.

----------, P. Robinson, J. Burger and M. Villain 1997. “Automatic Coreference: The Role of Annotated Training Data.” *Proc. of AAAI String Symposium on Applying Machine Learning to Discourse Processing.*

Hobbs, J. 1978. “Resolving Pronoun References.” Barbara Grosz, Karen Sparck-Jones and B. Webber (eds.) *Lingua.* 44. Also in *Readings in Natural Language Processing, Morgan Kaufmann,* Los Altos, *1986,* 339-352.

Järvinen, T. and P. Tapanainen 1997. “A Dependency Parser for English.” *The Technical Reports of the Department of General Linguistics, University of Helsinki.*

Kameyama, M. 1997. “Recognizing Referential Links: An Information Extraction Perspective.” *Proc. of a Workshop “Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts”,* Madrid, 46-53.

Kamp, H. and U. Reyle 1993. *From Discourse to Logic*. Dordrecht: Kluwer.

Kennedy, C. and B. Boguraev 1996. “Anaphora for Everyone: Pronominal Anaphora Resolution without a Parser.” *Proc. of the 16th Int. Conference on Computational Linguistics,* vol I, 113-118.

Lappin, S. and H. Leass 1994. “An Algorithm for Pronominal Anaphora Resolution.” *Computational Linguistics* 20:4.535-561.

Markert, K., M. Strube and U. Hahn 1996. “Inferential Realization Constraints on Functional Anaphora in the Centering Model.” *Proc. of the CogSci 1996,* 609-614.

McEnery, A., I. Tanaka and S. Botley 1997. “Corpus Annotation and Reference Resolution.” *Proc. of the ACL’97/EACL’97 Workshop on Operational Factors in Practical, Robust Anaphora Resolution,* Madrid, 67-74.

Miller, G., R. Beckwith, C. Fellbaum, D. Gross and K. Miller 1990. “Introduction to Wordnet: An On-line Lexical Database.” *Int. Journal of Lexicography* 3:4.235-244.

Mitkov, R. 1997. “Factors in Anaphora Resolution: They are not the Only Things that Matter. A Case Study Based on Two Different Approaches.” *Proc. of the Workshop “Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts”, Universidad Nacional de Educación a Distancia,* Madrid ed. Ruslan Mitkov and Branimir Boguraev, 14-21.

---------- 2001. “Outstanding Issues in Anaphora Resolution.” Al. Gelbukh (ed.) *Computational Linguistics and Intelligent Text Processing.* Berlin: Springer. 110-125.

---------- 2002. *Anaphora Resolution*. London: Longman.

Orăsan, C. 2000. “CLinkA a Coreferential Links Annotator.” *Proc. of Language Resources and Evaluation Conference (LREC 2000),* Athens, 491-496. See also: http://clg.wlv.ac.uk/projects/PALinkA/.

Poesio, M., R. Vieira and S. Teufel 1997. “Resolving Bridging References in Unrestricted Texts.” R. Mitkov and B. Boguraev (eds.) *Proc. of the Workshop “Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts”, Universidad Nacional de Educación a Distancia,* Madrid.

Reinhart, T. 1981. “Definite NP Anaphora and C-command Domains.” *Linguistic Inquiry* 6:12.605-635.

Resnik, P. 1999. “Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language.” *Journal of Artificial Intelligence Research* 11.95-130.

Tanaka, I. 1999. “The Value of an Annotated Corpus in the Investigation of Anaphoric Pronouns, with Particular Reference to Backwards Anaphora in English.” Phd. thesis, University of Lancaster.

A Machine Learning Approach to Preference Strategies for Anaphor Resolution

Roland Stuckardt
Johann Wolfgang Goethe University Frankfurt am Main[1]

In the last few years, much effort went into the design of robust anaphor resolution algorithms. Many algorithms are based on antecedent filtering and preference strategies that are manually designed. Along a different line of research, corpus-based approaches have been investigated that employ machine learning or statistical techniques for deriving strategies automatically, thus considerably facilitating knowledge engineering. Since, however, manually designing the robust antecedent filtering strategies constitutes a once-for-all effort, the question arises whether at all they should be derived automatically. In this article, it is investigated what may be gained by combining the best of two worlds: designing the universally valid antecedent filtering strategies manually, and deriving the potentially genre-specific antecedent preference strategies automatically by applying machine learning techniques. Following this paradigm, an anaphor resolution system ROSANA-ML is designed. Through a thorough formal evaluation, it is shown that, while exhibiting additional advantages, ROSANA-ML performs similar to its manually designed ancestor ROSANA.

1 Introduction

The interpretation of textual anaphoric expressions is a subtask which is crucial to a wide range of natural language processing problems. In the last few years, much effort went into the design of robust, knowledge-poor algorithms that are capable of processing potentially noisy data. Many approaches take as a starting point the landmark work of Lappin and Leass (1994), in which an algorithm for interpreting third person pronouns is developed that relies upon the idealistic assumption that, for the sentences to be interpreted, complete syntactic parses are available. For achieving robustness, various solutions have been suggested, e.g. to employ a robust part-of-speech tagger instead of full syntactic parsing (Kennedy & Boguraev, 1996), or to generalize the strategies to work on possibly fragmentary syntactic descriptions (Stuckardt, 1997, 2001).

Along a different line of research, corpus-based approaches have been investigated that employ machine learning or statistical techniques for deriving anaphor resolution strategies automatically (Soon *et al.*, 2001; Paul *et al.*, 1999; Ge *et al.*, 1998; Aone & Bennett, 1995, 1996; Dagan *et al.*, 1995; McCarthy & Lehnert, 1995; Connolly *et al.*, 1994). These approaches are considered

[1] Im Mellsig 25, D-60433 Frankfurt am Main, Germany
E-mail: roland@stuckardt.de. Phone: +49 (0)69 517797

particularly attractive because the effort for designing and implementing the strategies is reduced. In general, the automatic derivation of anaphor resolution strategies relies upon the availability of sufficiently large text corpora that are tagged, in particular, with referential information.[2]

For instance, Aone and Bennett (1995) employ supervised decision tree learning for deriving an anaphor resolution master strategy that covers antecedent filtering as well as antecedent preference criteria. They primarily aim at providing an elegant solution to the robustness issue per se; as an important advantage, they point out that their approach automatically generalizes to additional types of anaphoric expressions. However, the inventory of relevant types of anaphoric expressions is limited. Moreover, recent research has revealed that some classical approaches to robust anaphor resolution which descend from the work of Lappin and Leass (1994) are, with respect to the robust operationalization of the antecedent filtering strategies of syntactic disjoint reference and agreement in person/number/gender, nearly optimal (Stuckardt, 2001; Kennedy & Boguraev, 1996). Since the robust implementation of these successful anaphor resolution strategies constitutes a once-for-all effort, the question arises whether at all they should be derived automatically through the application of machine learning techniques.

In the present article, it is investigated what may be gained by employing machine-learned *preference* strategies[3] as part of a robust anaphor resolution approach according to the Lappin and Leass (1994) paradigm in which the antecedent filtering strategies are manually designed. The algorithm ROSANA described in (Stuckardt, 2001) is taken as the starting point. Empirical studies in this article have shown that, for achieving optimal interpretation results, the antecedent preference strategies, which come as sets of *weighted salience factors*, should be designed genre-specifically, since text genres seem to differ with respect to the characteristic properties of their typical coherence structures. Hence, there is no once-for-all optimal design of preference heuristics. Consequently, antecedent preference strategies are ideal targets for applying machine learning techniques.

Thus, it is explored what may be gained by combining the best of two worlds: designing the universally valid antecedent filtering strategies manually – once and for all –, and deriving the genre-specific antecedent preference

[2] While some approaches employing unsupervised learning have been explored, the most promising ones make use of supervised techniques. Some referentially annotated corpora have been developed during the last few years (particularly for the DARPA Message Understanding Conferences (MUCs)). However, the total amount of available tagged texts is still quite restricted.

[3] In emphasizing the application case of these criteria during anaphor resolution (viz., the antecedent selection phase, see Section 3.2), one could equally well speak of antecedent *selection* strategies.

strategies automatically by applying machine learning techniques. An anaphor resolution system ROSANA-ML, which follows this paradigm, is designed and implemented. Through a thorough formal evaluation, it is shown that, with respect to two important evaluation measures, ROSANA-ML reaches a level of performance that compares with the interpretation quality of its manually designed ancestor ROSANA. More specifically, the evaluation reveals that, whereas regarding third person possessive pronouns, a gain is achieved, the results regarding third person non-possessives slightly lag behind the performance of the manually designed system. In particular, the evaluation results regarding non-possessives indicate that the set of features over which the classifiers are learned should be suitably supplemented; it is expected that this enhances the need for still larger corpora of referentially annotated training texts, thus confirming similar findings of other researchers (e.g. (Mitkov, 2001)). Moreover, the results of a series of further experiments indicate that, regarding third-person pronominal anaphora in English, by biasing ROSANA-ML towards precision, better (precision, recall) tradeoffs (henceforth referred to as *(P,R)* tradeoffs) can be obtained than those determined by Aone and Bennett (1995) for the case of Japanese zero pronouns.

The article is organized as follows. In Section 2, the fundamental methodology is described. In particular, the machine learning approach, which employs the C4.5 decision tree algorithm of Quinlan (1993), is outlined; moreover, it is sketched how the training data are obtained and how the learned decision trees are applied for selecting antecedents. In Section 3, formal specifications of the algorithms are given, and the underlying paradigm of learning preference strategies is further illustrated; an implementation, the ROSANA-ML system, is briefly described. In Section 4, a series of experiments regarding, in particular, the choice of features, the employed training strategies, and the learning performance is designed. In Section 5, the respective empirical evaluation results are interpreted. Finally, in Sections 6 and 7, the findings are compared with the results of other approaches to anaphor resolution, and promising directions of further research are identified.

2 Methodology

In Figure 1, the machine learning approach to anaphor resolution followed by ROSANA-ML is outlined. It is distinguished between the training phase, which is shown in the upper part of the figure, and the application (anaphor resolution) phase sketched in the lower part of the figure.

During the *training phase*, based on a training text corpus, a set of feature vectors is generated which consists of feature tuples derived from the *(anaphor, antecedent candidate)* pairs that are considered during the antecedent

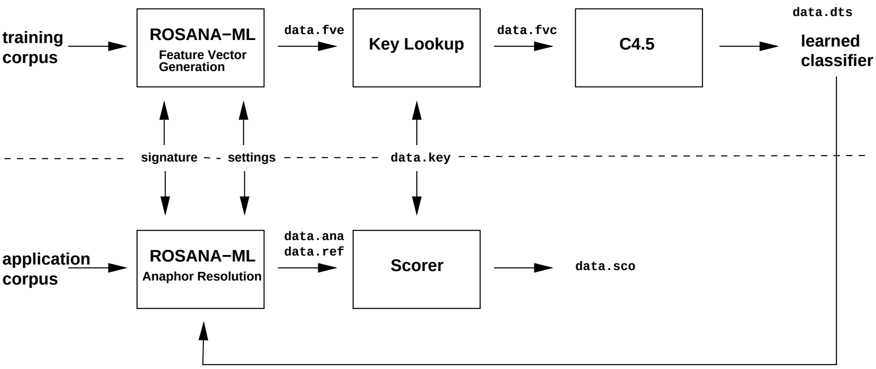


Figure 1: ROSANA-ML: training phase vs. application (anaphor resolution) phase

selection phase of the anaphor resolution algorithm ROSANA. This output is written to a file *data.fve*, which, during the next step, is classified by employing intellectually gathered key data (file *data.key*). The result consists of a set of training vectors (file *data.fvc*) which are classified as either COSPEC or NON_COSPEC, depending on whether, according to the key, the respective occurrences of anaphor and antecedent candidate are *cospecifying* or *not cospecifying*. Finally, these training cases are submitted to the C4.5 machine learning algorithm: C4.5 derives a decision-tree-shaped classifier (file *data.dts*) suitable for categorizing arbitrary feature vectors that are of the same signature as the training vectors.

In the *application (anaphor resolution) phase*, the learned classifiers are employed for antecedent selection: to discern between more and less plausible candidates, instead of applying a set of salience factors (as done by the manually designed algorithm ROSANA), a decision tree lookup is performed, which yields a (heuristic) prediction COSPEC or NON_COSPEC. In combination with a secondary preference criterion (such as surface distance), this prediction renders possible an ordering of the antecedent candidates of an anaphor according to decreasing plausibility. The anaphor resolution output is written to the files *data.ana* (coreference classes) and *data.ref* (basically, anaphoric resumption chains). During *formal evaluation*, the interpretation quality of ROSANA-ML will be measured with respect to various evaluation disciplines, among which are immediate antecedency (*ia*) and non-pronominal anchors (*na*) (see Section 4.6).

For being compatible with the classifiers learned during the training phase, the application version of ROSANA-ML has to employ the identical feature vector signature, i.e. the same Cartesian product of attribute sets to which the individual instances of anaphors and antecedent candidates are mapped. There are further settings, such as the exact way how the antecedent filtering criteria are to be applied, which should be identical during training set generation and application phase (see below).

3 Algorithms and Implementation

The algorithms employed by ROSANA-ML for training data generation and anaphor resolution are immediate descendants of the robust anaphor resolution algorithm underlying the manually designed system ROSANA. Regarding the full details of how robustness is achieved by ROSANA, the reader is referred to the article (Stuckardt, 2001). Like its ancestor, ROSANA-ML handles a broad range of entity-specifying expressions, in particular ordinary, possessive, reflexive/ reciprocal, and relative pronouns, definite NPs, and names. The machine learning experiments described in this article will focus on the key cases of third person non-possessive and possessive *pronominal* anaphora.[4]

In aiming at determining the coreference classes of non-zero linguistic expressions which specify entities[5], ROSANA-ML covers the coreference task of the Message Understanding Conferences (Hirschman, 1998; Vilain *et al.*, 1996; Grishman & Sundheim, 1996).

3.1 Training data generation

Figure 2 gives the specification of the training data generation algorithm. The antecedent filtering step 1, in which different kinds of restrictions for eliminating impossible antecedents (in particular, agreement in person/number/gender and syntactic disjoint reference) are applied, is immediately taken over from the original ROSANA algorithm. In step 2, however, no salience ranking of the remaining antecedent candidates is performed. Rather, each remaining anaphor-candidate pair (α_i,γ_j) is mapped to a feature vector $fv(\alpha_i,\gamma_j)$, the attributes $f_1,...,f_{k\sigma}$ of which comprise individual and relational features derived from the descriptions of the occurrences α_i and γ_j.

[4] As modeled, e.g., by Binding Principle A of the Government and Binding (GB) theory by Chomsky (1981), there are tight syntactic bounds that confine the antecedent options for reflexives and reciprocals. Since these restrictions can be robustly implemented, pronouns of these types can be resolved with very high precision and recall anyway. A similar observation holds with respect to relative pronouns, which, too, can be interpreted with high accuracy by mere surface positional and syntactic means. Hence, these types of anaphoric expressions are not considered during the machine learning experiments; anyway, they are dealt with by appropriate manually designed interpretation strategies as specified in (Stuckardt, 2001). Regarding definite NPs and names, machine learning experiments should be based on additional lexico-semantic and ontological information not taken into account in the purely syntactic framework of the ROSANA approach.

[5] This contrasts with coreference resolution of expressions that, e.g., specify events.

1. *Candidate filtering*: for each anaphoric NP α, determine the set of admissible antecedents γ:
   a. verify morphosyntactic or lexical agreement with γ;
   b. if the antecedent candidate γ is intrasentential: apply the robust syntactic disjoint reference filter as specified in (Stuckardt, 2001), Figure 4;
   …
2. *Feature vector generation*: for each remaining anaphor-candidate pair (α_i,γ_j):
   a. generate, according to the feature signature σ under consideration, the feature vector $fv(\alpha_i,\gamma_j) := (n_{\alpha i}, n_{\gamma j}, f_1, \ldots, f_{k\sigma})$ where $n_{\alpha i}$ and $n_{\gamma j}$ are the number (unique identifiers referred to in the key) of the occurrences α_i and γ_j, and $f_1, \ldots, f_{k\sigma}$ are (individual and relational) features derived from α_i and γ_j with respect to the signature σ;
   b. write $fv(\alpha_i,\gamma_j)$ to the training data file *data.fve*.

Figure 2: ROSANA-ML: training data generation

The *signature* of the feature vectors, i.e. the inventory of features to be taken into account[6] has to be chosen carefully in order to fulfill the conditions of robust processing: instead of requiring complete and unambiguous descriptions, they should be computable from potentially partial representations such as *fragmentary* syntactic representations.[7]

As initially motivated, by restricting the consideration to (α_i,γ_j) instances in which γ_j denotes an antecedent candidate that, relatively to α_i, doesn't violate any tight condition, the learning approach focuses on a subset of cases that, from a knowledge engineering point of view, are difficult to decide upon algorithmically (see Figure 3).[8] In other words, machine learning techniques are applied only for handling the (presumably) difficult cases, i.e. to discern between cospecifying and non-cospecifying candidates that, at current, cannot be distinguished by applying one of the robustly computable restrictions.[9]

---

[6] Formally, the feature vectors are instances of an underlying signature, which is defined as the Cartesian product of the sets of attributes taken into account: $fv(\alpha_i,\gamma_j) \in A_1 \times A_2 \times \ldots \times A_{k\sigma}$

[7] The two additional "technical" features $n_{\alpha i}$ and $n_{\gamma j}$, which correspond to the head token surface number of anaphor and candidate, are required for relating the output *data.fve* to the key data; they are removed during the generation of the file *data.fvc* of classified vectors.

[8] Of course, as pointed out by Mitkov (1997), the distinction between tight constraints and fuzzy preference criteria is all but uncontroversial. Suffice it to say that what is perceived as a tight constraint is determined through our current state of knowledge: what, today, may be taken as a weak preference criterion, could in future, based on a deeper insight into the problem, be stepwisely refined such that, eventually, a tight criterion emerges.

[9] A closer analysis reveals that the initially mentioned statistical approaches, too, do at most partially match this clear-cut paradigm. Dagan and Itai (1990) explore a related approach, in which selectional preferences are automatically derived through a statistical analysis of large corpora. Contrary to the methodology followed in the article at hand, they don't make use of coreference information. Importantly, the acquired selectional preference criteria are intended to supplement, rather than substitute, other preference strategies. This has been further explored by Dagan *et al.* (1995) and by Lappin and Leass (1994), who showed that, by supplementing a syntactic salience-based anaphor resolver with statistical preferences, an improvement of 2.5 percent can be obtained.

Ge *et al.* (1998) investigate a statistical approach that employs corpora annotated with syntactic and coreference information to derive antecedent preference criteria. The tree search algorithm developed by Hobbs (1978) is used as the base strategy for anaphor resolution. Thus, in requiring complete syntactic

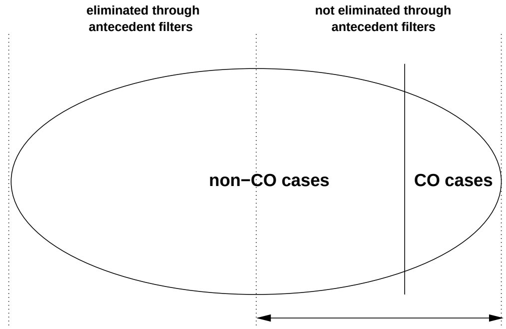


Figure 3: focusing on the relevant cases

3.2 Anaphor resolution

The specification of the ROSANA-ML anaphor resolution algorithm proper is given in Figure 4. Again, step 1 is identical with the antecedent filtering phase of the manually designed ROSANA algorithm. Step 2, however, is modified. For a specific instance (α_i,γ_j) of anaphor and antecedent candidate, after the computation of the feature vector $fv(\alpha_i,\gamma_j)$, the decision tree lookup takes place; basically, its result $\Psi_\sigma^{type(\alpha i)}(fv(\alpha_i,\gamma_j))$ consists in a prediction $\in$ {COSPEC,NON_COSPEC}.[10] In the subsequent step, these predictions are employed for computing a ranking over the candidate sets of each anaphor. In its base version, candidates which are classified to COSPECify with the anaphor rank higher than candidates that are predicted to NON_COSPECify; surface nearness (i.e. word distance) serves as the secondary criterion. Among the possible refinements are: further ranking the candidates according to the classification error probability yielded by the decision tree lookup, and eliminating candidates which are (fuzzily) classified as NON_COSPECifying (see Section 5.5 below, in which results of a series of respective experiments

---

descriptions, this approach cannot be considered to be truely robust, i.e. operational on potentially noisy data. Moreover, while the syntactic disjoint reference conditions can be regarded to be implicitly covered by the tree search algorithm, the filtering criterion of agreement in person/number/gender is left to the responsibility of the statistically derived word-oriented preference criteria. To that extent, their approach differs from the paradigm outlined in Figure 3.

Notably, Paul *et al.* (1999) investigate the opposite scenario in which decision tree classifiers are employed as candidate *filters* only; since, however, they are studying the case of Japanese restricted domain dialogues, an immediate comparison of their evaluation results with the figures determined below is problematic. Moreover, as initially mentioned, various approaches employ decision tree classifiers as the solitary master strategy, thus implicitly covering antecedent preference and filtering, e.g. (Aone & Bennett, 1995; McCarthy & Lehnert, 1995; Connolly *et al.*, 1994).

[10] To put it formally: a *classifier function* $\Psi_\sigma^{type(\alpha i)}$: $A_1 \times A_2 \times ... \times A_{k\sigma} \mapsto$ {COSPEC, NON_COSPEC} is applied that maps instances of the underlying signature σ to cospecification / non-cospecification predictions.

1. *Candidate filtering*: for each anaphoric NP α, determine the set of admissible antecedents γ:
   a. verify morphosyntactic or lexical agreement with γ;
   b. if the antecedent candidate γ is intrasentential: apply the robust syntactic disjoint reference filter as specified in (Stuckardt, 2001), Figure 4;
      …
2. *Candidate scoring and sorting*:
   a. for each remaining anaphor-candidate pair (α_i,γ_j):
      i. determine, according to the feature signature σ underlying the learned classifier (decision tree) $\Psi_\sigma^{type(\alpha i)}$ to be applied, the internal representation of the feature vector $fv(\alpha_i,\gamma_j) := (f_1,\ldots,f_{k\sigma})$ where $f_1,\ldots,f_{k\sigma}$ are (individual and relational) features derived from the occurrences α_i and γ_j with respect to the applicable signature σ;
      ii. *decision tree lookup*: determine the prediction $\Psi_\sigma^{type(\alpha i)}(fv(\alpha_i,\gamma_j))$ of the learned classifier with respect to the instance $fv(\alpha_i,\gamma_j)$.
   b. for each anaphor α: sort candidates γ_j according the following criteria:
      i. *primary*: candidates γ_j for which $\Psi_\sigma^{type(\alpha)}(fv(\alpha,\gamma_j))$ = COSPEC are preferred over candidates $\gamma_{j'}$ for which $\Psi_\sigma^{type(\alpha)}(fv(\alpha,\gamma_{j'}))$ = NON_COSPEC;
      ii. *secondary*: surface nearness.
   c. sort the anaphors α according to the above criteria applied to their respective best antecedent candidates.
3. Antecedent selection: consider anaphors α in the order determined in step 2c. Suggest antecedent candidates $\gamma_j(\alpha)$ in the order determined in step 2b. Select $\gamma_j(\alpha)$ as candidate if there is no interdependency, i.e. if
   a. the morphosyntactic features of α and $\gamma_j(\alpha)$ are still compatible,
   b. for all occurrences $\delta_{\gamma j(\alpha)}$ and δ_α the coindexing of which with $\gamma_j(\alpha)$ and (respectively) α has been determined in the *current* invocation of the algorithm: the coindexing of $\delta_{\gamma j(\alpha)}$ and δ_α, which results transitively when choosing $\gamma_j(\alpha)$ as antecedent for α, does neither violate the binding principles nor the i-within-i condition. (see the full specification in (Stuckardt 2001), Figure 4)

Figure 4: ROSANA-ML: anaphor resolution through learned classifiers

will be given). There is a final step 3 in which the actual antecedent selection takes place. The remaining candidates are considered in the order determined by the ranking procedure; additional means are taken to avoid combinations of antecedent decisions that are mutually incompatible (Stuckardt, 2001).

3.3 Implementation

Based on the algorithms described in Figure 2 and Figure 4, the machine-learning-based anaphor resolution system ROSANA-ML has been implemented. Regarding its peripheral modules (definition of basic data structures, preprocessing of the externally provided parsing results, occurrence identification, restriction verification, result scoring), ROSANA-ML is code-identical with its manually designed ancestor. Further components have been added to provide the functionality for training data generation, feature vector classification, decision tree lookup, and modified candidate ranking. The ROSANA-ML System has been implemented in Common Lisp.[11]

For the task of learning decision tree classifiers from the training data, the C4.5 implementation for Unix of the University of Regina[12] is employed.

4 *Layout of Experiments*

A series of experiments at different levels of consideration will be carried out:

4.1 Variation of feature vector signatures

The first and most fundamental question regards the set of attributes, i.e. the signature of the feature vectors from which the classifiers will be learned. As pointed out above, the choice is confined to attributes that are *robustly computable* over the morphological, syntactical, and semantic information available under application conditions. Actual signatures are then defined by selecting subsets of the above attributes.

In Table 1, the set of attributes currently taken into account is shown. *type(o)* denotes the type of the respective occurrence *o*, in particular PER3/POS3 (third person non-possessive/possessive pronouns), VNOM (ordinary noun phrases), and NAME (proper names); regarding the anaphor (o = α), the choice is restricted to PER3 and POS3 in the current experiments. The feature *synfun(o)* describes the syntactic function of *o*. *synlevel(o)* captures a coarse notion of (non-relational) syntactic prominence,[13] which is measured by counting the number of principal categories[14] occurring on the path between *o* and the root of the respective parse fragment. Features *number(o)* and *gender(o)* capture the respective morphological characteristics of anaphor α and candidate γ. Furthermore, surface context information about the three neighbours to the left and to the right of α and γ is taken into account, comprising the syntactic category (*syncateg(o)*) and, again, the syntactic function (*synfun(o)*) of the respective token(s). Finally, four relational features are considered: *dist(α,γ)* (sentence distance, only distinguishing between three cases: same sentence, previous sentence, two or more sentences away), *dir(α,γ)* (whether γ topologically precedes α or vice versa), *synpar(α,γ)* (identity of syntactic

[11] The FDG parser for English of Järvinen and Tapanainen (1997) has been chosen as the syntactic preprocessor. In giving robustness and processing speed priority over normativeness and syntactic coverage of the underlying grammar, the parser meets the requirements on a preprocessor for robust anaphor resolution on unrestricted texts.

[12] Release 8 for Unix, available (February 1, 2002) at http://www.cs.uregina.ca/~dbd/cs831/notes/ml/dtrees/c4.5/tutorial.html

[13] This contrasts with *relational* notions of syntactic prominence, in which the relative position to the other occurrence is taken into account (e.g. *c-command*).

[14] Formally, principal categories can be defined as surface structure nodes that, in the sense of the Government and Binding (GB) theory, constitute *binding categories* (Chomsky, 1981).

Feature	Examples of Instances	Description
type(α)	PER3, POS3	type of anaphor α
synfun(α)	subje, trans	syntactic function of α
synlevel(α)	TOP, SUB, SUBSUB	syntactic position of α
number(α)	SG	morphological number of α
gender(α)	MASK	gender of α
syncateg($ln_i(\alpha)$)	N, DET	category of left neighbour i, $1 \leq i \leq 3$
synfun($ln_i(\alpha)$)	subje, trans	synt.function of left neighbour i, $1 \leq i \leq 3$
syncateg($rn_i(\alpha)$)	N, DET	category of right neighbour i, $1 \leq i \leq 3$
synfun($rn_i(\alpha)$)	subje, trans	synt. function of right neighbour i, $1 \leq i \leq 3$
type(γ)	VNOM, NAME, PER3	type of candidate γ
synfun(γ)	subje, trans	syntactic function of γ
synlevel(γ)	TOP, SUB, SUBSUB	syntactic position of γ
number(γ)	SG	morphological number of γ
gender(γ)	MASK	gender of γ
syncateg($ln_i(\gamma)$)	N, DET	category of left neighbour i, $1 \leq i \leq 3$
synfun($ln_i(\gamma)$)	subje, trans	synt. function of left neighbour i, $1 \leq i \leq 3$
syncateg($rn_i(\gamma)$)	N, DET	category of right neighbour i, $1 \leq i \leq 3$
synfun($rn_i(\gamma)$)	subje, trans	synt. function of right neighbour i, $1 \leq i \leq 3$
dist(α ,γ)	INTRA, PREV, PPREV	sentence distance between γ and α
dir(α ,γ)	ANA, KATA	resumption: anaphoric or cataphoric?
synpar(α ,γ)	YES, NO	syntactic role identity (parallelism)?
syndom(α ,γ)	$[\alpha \rightarrow \gamma]$, $[\gamma \rightarrow \alpha]$, none	synt. dominance relations betw. clauses?

Table 1: complete set of features over which the signatures are defined

function),[15] and *syndom(α,γ)* (relative syntactic position of the clauses of anaphor α and candidate γ if they occur in the same sentence).[16]

At the first experimental level, different subsets of attributes will be considered. In particular, it is experimented with signature σ_{full}, which comprises the complete attribute set (38 features), signature σ_{n1}, in which only the syntactic categories and functions of the *immediately* preceding and following neighbours are taken into account (22 features), and signature σ_{n0}, where the *syncateg* and *synfun* attributes of the neighbours are completely ignored (14 features).

4.2 Variation of training data generation settings

At the second experimental level, different settings regarding the extension of the sets of training vectors to be generated are taken into consideration.

Extending the training set by switching off recency limits: in the manually designed algorithm ROSANA, a recency filter is applied that eliminates candidates for pronominal anaphors that are, in terms of sentence distance, too

[15] Thus, the role inertia information that has been found to be useful in the classical, manually designed approaches (Lappin & Leass, 1994; Stuckardt, 2001) is also captured.
[16] E.g., $[\alpha \rightarrow \gamma]$ describes the case in which the clause of γ is syntactically subordinated to the clause of α

far away. This strategy, which is justified by the observation that, in most cases, antecedent occurrences are available in the immediately preceding sentences, drastically reduces the amount of generated training data. It is experimentally evaluated what may be gained by switching off this filter during the training data generation phase.

General or specialized classifiers: according to an important result of the formal evaluation of the manually designed system ROSANA, *different* antecedent preference criteria should be employed for third-person non-possessive and possessive pronouns. To reflect this observation, it is experimented with the strategy of generating training data sets for learning two different, i.e. *specialized* classifiers, one of which dedicated for dealing with non-possessives, the other of which designed for handling possessives.

From a learning-theoretical point of view, the so-far sketched variations of the experimental settings are considered to be redundant as long as enough training data are available: the decision-tree learning algorithm should be able to automatically discern between important and unimportant features, and, moreover, should determine for itself whether the classifiers for non-possessive and possessive third-person persons should be kept apart, reflected by the occurrence of the type(α) feature at or near the root of the derived decision tree classifier; a similar argument holds with respect to the extension of the training set by switching off the recency limits, which results in a (supposedly slight) adulterating of the training data since cases are taken into account that may, according to their general characteristics, differ from the general characteristics of the application-relevant cases inside the recency bounds. However, it will turn out that these experimental variations are successful techniques for heuristically coping with the problem of training data sparsity.

4.3 Variation of C4.5 decision tree learning settings

The C4.5 algorithm provides for different settings that determine how the decision trees are learned. One key parameter, the so-called *pruning confidence factor CF*, characterizes the amount of *pruning* (given in percent) performed for avoiding overfitting the training data (Mitchell, 1997). The optimal setting of this parameter depends upon the amount and the reliability of the available training data. Using the empirically best signature and training data generation settings as the point of departure, it will be experimented with different values for this factor.

4.4 Intrinsic (decision tree) cross-validation and learning curve analysis

The final output of the learning phase consists of decision tree classifiers. At the level of *intrinsic cross-validation*, it will be evaluated how these classifiers perform with respect to their basic predictions ∈ {COSPEC,NON_COSPEC}, averaged over ten experiments with varying sets of training and evaluation data. For this 10-fold cross-validation, the set of feature vectors of the training/evaluation corpus is randomly split into 10 parts of equal size. Ten different experiments are run in which decision trees are determined over nine of the ten subsets and evaluated on the remaining subset. The cumulated (average) results are given as *confusion matrices* that describe how the learned classifiers perform with respect to the two classes COSPEC and NON_COSPEC.

Furthermore, for the training set / evaluation set split on which the classifiers yield median results, the *learning curves* of the classifiers will be analysed, thus, in particular, giving evidence regarding the amount of training data typically needed for obtaining an empirically optimal (classifier-intrinsic) performance.

4.5 Extrinsic (application-level) cross-validation

A similar experiment of (here: 6-fold) cross-validation will be carried out at the application (i.e. anaphor resolution) level. In contrast to the above experiment of classifier-oriented cross-validation, the random split of the training data is performed at the document level, i.e. the overall corpus, which comprises 66 documents, is split into six subsets of eleven documents each. In this case, cross-validation results with respect to the main evaluation disciplines of the anaphor resolution task will be determined.

4.6 Text corpus and disciplines of formal evaluation

The training and evaluation of the ROSANA-ML system will be performed on a corpus of 66 news agency press releases, comprising 24,712 words, 406 third-person non-possessives[17], and 246 third-person possessive pronouns. For the first three experimental stages, the corpus is firmly partitioned into a training subset (31 documents, 11,808 words, 202 non-possessives, 115 possessives) and an evaluation subset (35 documents, 12,904 words, 204 non-possessives, 131 possessives); for the cross-validation stages, further partitions are generated randomly (see above). In all experiments, the training data generation and the application of the trained system take place under conditions of potentially noisy data, i.e. without a priori intellectual correction of orthographic or syntactic errors.

The anaphor resolution performance will be evaluated with respect to two evaluation disciplines: *immediate antecedency* (*ia*) and *non-pronominal*

[17] Relative pronouns are excluded from consideration since they are effectively resolvable with high accuracy by surface-topological means.

anchors (*na*). In the first-mentioned discipline, an elementary accuracy measure is employed that determines the precision of correct immediate antecedent choices; by further taking into account cases of unresolved anaphors, the respective recall measure is obtained.[18] In the last-mentioned discipline, the performance with respect to the (application-relevant) determination of *non-pronominal* antecedents is evaluated: precision and recall measures are defined in the same way; however, only non-pronominal antecedent candidates are considered. Thus, the anaphor resolution performance is measured according to the tradeoffs (P_{ia},R_{ia}) and (P_{na},R_{na}). For formal definitions and an in-depth discussion of the evaluation measures, the reader is referred to (Stuckardt, 2001).

5 Experiments and Empirical Results

5.1 Optimising the signature and the training data generation settings

In Table 2, the results of the formal, corpus-based evaluation on the *News Agency Press Releases* corpus are summarized. In the upper line, the scores of the manually designed ROSANA system are given. The next three groups of rows display the evaluation results for the signatures σ_{n0}, σ_{n1}, and, respectively, σ_{full} (see Section 4.1, first level of experimental variation). Inside these groups, the training data generation settings are varied (second level). In these stages of experimentation, the partition of the corpus into training data and evaluation data remains fixed ($[d_1^{31},d_{32}^{66}]$).

In the base level experiment of signature variation (rows labeled (1), (2), (3)), non-possessive and possessive pronouns behave nonuniformly: whereas, with growing number of considered neighbours, non-possessives score marginally better, the performance on possessive pronouns slightly deteriorates.

More importantly, an in-depth qualitative analysis of the typical failure cases concerning the determination of immediate antecedents revealed that a substantial amount of incorrect decisions could have been avoided by dispreferring cataphoric resumptions.[19] In the system ROSANA, this negative preference criterion, which is known to promote a good overall antecedent selection performance, is manually encoded as the so-called cataphora malus factor, which is applied during the antecedent scoring phase. ROSANA-ML, however, failed to learn a respective criterion from the training data, which may

[18] Under the assumption that *all* pronouns are resolved, the precision measure yields results that are immediately comparable with the accuracy figures given in the evaluations of the classical approaches of, e.g., Lappin and Leass (1994), and Kennedy and Boguraev (1996). By further allowing for unresolved pronouns, *(P,R)* tradeoffs are obtained that seem to be comparable with the evaluation results that are given by Aone and Bennett (1995).

[19] That is, cases of anaphora with antecedents surface-topologically *following* the anaphor.

be attributed to the fact that the cospecification information employed at the learning-relevant level of *individual* antecedent decisions is *inherently symmetrical*.[20] This observation gave rise to a further variation at the level of feature vector generation settings: *eliminating instances of cataphoric resumption* in the training as well as in the application phase.

	antecedents *(P_{ia},R_{ia})*		anchors *(P_{na},R_{na})*	
	PER3	POS3	PER3	POS3
ROSANA (manually)	(0.71,0.71)	(0.76,0.76)	(0.68,0.67)	(0.66,0.66)
(1) σ_{n0}, [d_1^{31},d_{32}^{66}]	(0.61,0.60)	(0.71,0.71)	(0.54,0.53)	(0.67,0.66)
(1_{nc}) = (1) $\wedge$ no cataphors	(0.62,0.62)	(0.77,0.77)	(0.57,0.56)	(0.70,0.70)
(1^{tc}) = (1) $\wedge$ type(α)-spec. class.	(0.61,0.60)	(0.69,0.69)	(0.56,0.55)	(0.66,0.65)
(1_{nc}^{tc}) = (1^{tc}) $\wedge$ no cataphors	**(0.63,0.63)**	**(0.76,0.76)**	**(0.60,0.59)**	**(0.73,0.73)**
(1_{nc}^{tc}+) = (1_{nc}^{tc}) $\wedge$ no recency filt.	(0.63,0.63)	(0.73,0.73)	(0.58,0.58)	(0.63,0.63)
(2) σ_{n1}, [d_1^{31},d_{32}^{66}]	(0.62,0.61)	(0.70,0.70)	(0.54,0.54)	(0.65,0.65)
(2+) = (2) $\wedge$ no recency filter	(0.60,0.60)	(0.70,0.70)	(0.52,0.50)	(0.61,0.60)
(2_{nc}) = (2) $\wedge$ no cataphors	(0.63,0.62)	(0.74,0.74)	(0.57,0.57)	(0.66,0.66)
(2^{tc}) = (2) $\wedge$ type(α)-spec. class.	(0.60,0.60)	(0.70,0.70)	(0.56,0.55)	(0.68,0.67)
(2_{nc}^{tc}) = (2^{tc}) $\wedge$ no cataphors	(0.63,0.63)	(0.73,0.73)	(0.60,0.59)	(0.65,0.65)
(3) σ_{full}, [d_1^{31},d_{32}^{66}]	(0.62,0.62)	(0.69,0.69)	(0.55,0.55)	(0.62,0.62)
(3^{tc}) = (3) $\wedge$ type(α)-spec. class.	(0.61,0.61)	(0.69,0.69)	(0.57,0.56)	(0.63,0.62)
(3_{nc}^{tc}) = (3^{tc}) $\wedge$ no cataphors	(0.62,0.62)	(0.75,0.75)	(0.57,0.56)	(0.64,0.64)
(3+) = (3) $\wedge$ no recency filter	(0.60,0.59)	(0.69,0.69)	(0.49,0.49)	(0.57,0.57)
(3^{tc}+) = (3+) $\wedge$ type(α)-spec. cl..	(0.62,0.61)	(0.68,0.68)	(0.54,0.53)	(0.64,0.63)
(3_{nc}^{tc}+) = (3^{tc}+) $\wedge$ no cataphors	(0.62,0.62)	(0.76,0.76)	(0.58,0.57)	(0.68,0.68)

Table 2: evaluation results: signature and settings variation

The evaluation results illustrate that, under the *no cataphor* setting, with only one minor exception, results improve considerably. In particular, this holds for possessive pronouns: in experiment (1_{nc}), e.g., the gain in the *immediate antecedency* discipline amounts to 6 points of percentage for P_{ia} and R_{ia} each; in the *nonpronominal anchor* discipline, the improvement is reflected too, amounting to 3% for P_{na}, and 4% for R_{na}.

Extending the training set by switching off the recency limits seems to induce, at first sight, a deterioration: compare, e.g., experiments (1_{nc}^{tc}) and (1_{nc}^{tc}+), or (2) and (2+). However, the comparison of the case series [(3), (3^{tc}), (3_{nc}^{tc})] vs. [(3+), (3^{tc}+), (3_{nc}^{tc}+)] shows that this observation doesn't generalize.

[20] Antecedent selection interdependency comes into play here. If a cataphoric antecedent candidate, which itself embodies an anaphor, is selected prior to being resolved, some antecedent options are ruled out for this occurrence since selecting a candidate that is known to be cospecifying (in particular: the cataphor that already resumes this occurrence) would not yield any new information. It turned out that in a considerable number of such cases, this occurrence then gets wrongly resolved. This is an immediate consequence of the greedy strategy employed during the antecedent selection phase (step 3 of the ROSANA-ML algorithm, see Figure 4), which, in order to avoid exponential time complexity, doesn't optimize the *combined* plausibility of the antecedent decisions.

Rather, it seems to depend on the further settings: in the last-mentioned case, in which the *no cataphor* as well as the *type-specific classifier* settings are activated, there is a slight gain with respect to immediate antecedency of possessive pronouns, and a slight to considerable gain concerning the *nonpronominal anchors* scores for non-possessives and possessives. This may be explained by referring to the respective training set sizes, which are displayed in Table 3. In the base ("standard") case (3), one general classifier is constructed over 7,696 vectors. In the *type-specific classifier* setting, *two* specialized classifiers have to be learned, the one for non-possessives over 4,804 samples, the one for possessives over 2,892 samples. Under the *no cataphor* setting, the respective training set sizes are further reduced to 4,446 and 2,670, respectively. The observation may thus be explained by referring to the argument of Section 4.2: if the amount of available data is sufficiently large, the adulterating effect of artificially enlarging the training set prevails; if, however, training data are sparse, the overall effect may be positive.

	training set sizes		
training set generation settings	general	PER3	POS3
standard	7,696	4,804	2,892
no cataphors	7,116	4,446	2,670
no recency filter	17,416	11,115	6,301
no cataphors, no recency filter	16,836	10,757	6,079

Table 3: sizes of the training sets

	antecedents (P_{ia},R_{ia})		anchors (P_{na},R_{na})	
	PER3	POS3	PER3	POS3
$(1_{nc}{}^{tc})$ (CF = 25%)	(0.63,0.63)	**(0.76,0.76)**	(0.60,0.59)	**(0.73,0.73)**
$(1_{nc}{}^{tc},15) = (1_{nc}{}^{tc}) \wedge$ CF = 15%	(0.63,0.62)	(0.76,0.76)	(0.61,0.60)	(0.69,0.69)
$(1_{nc}{}^{tc},37) = (1_{nc}{}^{tc}) \wedge$ CF = 37%	**(0.65,0.64)**	(0.72,0.72)	**(0.61,0.61)**	(0.64,0.64)
$(1_{nc}{}^{tc},50) = (1_{nc}{}^{tc}) \wedge$ CF = 50%	(0.63,0.62)	(0.72,0.72)	(0.56,0.56)	(0.62,0.62)
$(1_{nc}{}^{tc},62) = (1_{nc}{}^{tc}) \wedge$ CF = 62%	(0.62,0.61)	(0.72,0.72)	(0.55,0.55)	(0.61,0.61)
$(1_{nc}{}^{tc},75) = (1_{nc}{}^{tc}) \wedge$ CF = 75%	(0.62,0.62)	(0.72,0.72)	(0.56,0.56)	(0.61,0.61)
$(1_{nc}{}^{tc},h) = (1_{nc}{}^{tc}) \wedge$ CF = (37|25)%	**(0.65,0.64)**	**(0.76,0.76)**	**(0.62,0.61)**	**(0.73,0.73)**

Table 4: evaluation results: pruning confidence factor variation

The *type-specific classifiers* setting yields nonuniform effects. In some cases, there are gains as well as losses ((1) vs. (1^{tc}), (2) vs. (2^{tc})). As identified above, however, specialized classifiers seem to pay off in combination with the *extended training set* mode. A particular behaviour is exhibited by the $(1_{nc}{}^{tc})$ experiment, which, in terms of overall (averaged) performance, can be considered to comprise the empirically optimal settings: whereas, concerning signature σ_{n0}, the *type-specific classifier* setting alone doesn't yield an overall positive contribution ((1^{tc}) vs. (1)), together with the *no cataphor* setting, the positive effects prevail. In this specific case, the advantage of employing specialized classifiers may outweigh the disadvantage of the small number of training cases since the number of attributes of signature σ_{n0} is considerably lower than in the case of σ_{full} (14 vs. 38).

Through further experiments, the results of which are not displayed in Table 2, the positive contributions of various subclasses of features have been validated. For example, the evaluation of a signature $\sigma_{full}^{-synpro}$, which consists of the features of σ_{full} minus the *synlevel* and *syndom* attributes (35 features), confirmed the positive contribution of the non-relational and relational attributes of syntactic prominence.

5.2 Optimizing the C4.5 decision tree learning settings

The settings of the experiment (1_{nc}^{tc}) have been taken as the starting point of further variations at the level of the C4.5 decision tree learning proper (see Section 4.3), viz. different settings of the *pruning confidence factor CF*. The base value of CF in all above-discussed experiments was 25%. Hence, it has been experimented with further CF values of 15, 37, 50, 62, and 75%. For possessive pronouns, according to the respective results, which are given in Table 4, the original setting of CF=25% yields the best scores; classifiers for non-possessives, however, should be determined with a slightly higher CF of 37%. Again, this may be explained by the different sizes of the training sets: for non-possessives, more training cases are available, resulting in a decision tree that better generalizes, thus necessitating a lower amount of pruning, i.e. allowing for a higher pruning confidence factor.

The row (1_{nc}^{tc},h) displays the evaluation results of a *hybrid setting* in which specialized classifiers for non-possessives and possessives are computed with the respective empirically optimal choices of CF values.

cases	CO	¬ CO
CO	62.7%	37.3%
n = 1,518	952	566
¬ CO	3.5%	96.5%
n = 8,187	284	7,903

cases	CO	¬ CO
CO	59.4%	40.6%
n = 1,066	633	433
¬ CO	4.0%	96.0%
n = 4,777	190	4,587

Table 5: 10-fold cross-validation, confusion matrices
(signature σ_{n0}, experiment (1_{nc}^{tc},h)): PER3 and POS3 classifiers

5.3 Intrinsic cross-validation and learning curves

In Table 5, the results of a *10-fold intrinsic cross-validation* according to the method outlined in Section 4.4 are given. The two confusion matrices display the overall (cumulated) scores for the PER3 and POS3 classifiers that have been derived by employing the settings of experiment (1_{nc}^{tc},h).

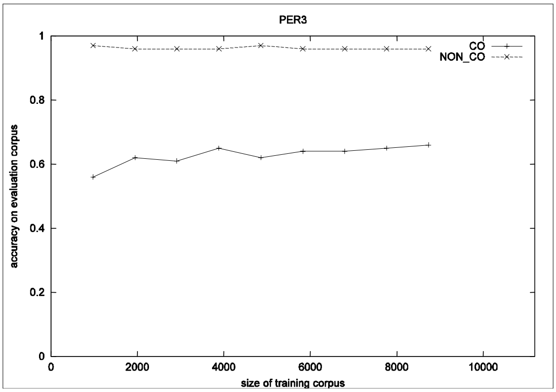


Figure 5: learning curve, signature σ_{n0}, experiment $(1_{nc}^{tc},h)$: PER3 classifier

Regarding the PER3 classifier, the overall number of training/evaluation vectors amounts to 9,705, of which 1,518 are COSPEC instances and 8,187 are NON_COSPEC instances. According to the upper row of the PER3 table, 62.7% of the COSPEC cases are correctly classified, and 37.3% are erroneously classified as belonging to the NON_COSPEC class. The scores for the NON_COSPEC vectors, which are shown in the lower row, are considerably higher: 96.5% correct, 3.5% incorrect. The results obtained for the POS3 classifier are similar. Here, the overall number of training/evaluation vectors is lower (5,843). Of the 1,066 COSPEC cases, 59.4% are correctly classified; instances of the 4,777 vectors belonging to the NON_COSPEC class are identified with an accuracy of 96%.

At first sight, the comparatively low accuracy obtained for the COSPEC cases seems to impose a problem. The situation is not as worse as it looks like since, with respect to step 2b of the ROSANA-ML algorithm (see Figure 4), it is of primary importance not to misclassify the NON_COSPEC cases; wrongly classifying a COSPEC instance is unproblematic as long as there are further cospecifying antecedent candidates that are correctly recognized. A closer analysis shows that, for most anaphors, indeed, *several* cospecifying candidates are available. However, one also has to take in account that the relative amount of NON_COSPEC instances is quite high.[21] Hence, although the NON_COSPEC instance classification error rate lies clearly below 5%, the probability that, for a certain anaphor to be resolved, one of the (typically numerous) non-cospecifying antecedent candidates gets wrongly classified is still not neglectable.

[21] See Table 5: about 84% of all instances for PER3, and about 82% of all instances for POS3.

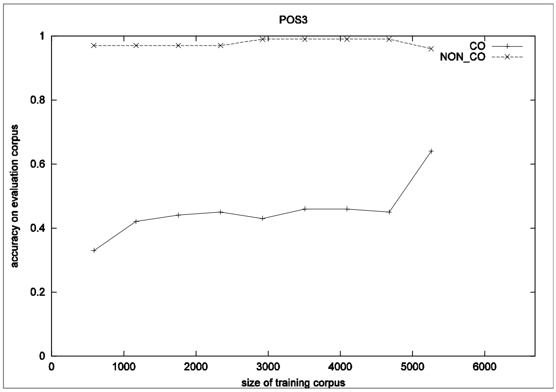


Figure 6: learning curve, signature σ_{n0}, experiment (1_{nc}^{tc},h): POS3 classifier

Out of the random 10-fold partition of the training data, the respective subsets for which median scores were obtained during cross-validation have been employed as the target data for a *learning curve analysis* of the two classifiers. In Figure 5, the learning curve of the PER3 classifier is shown (1,378 CO + 7,356 NON_CO = 8,734 training cases, 140 CO + 831 NON_CO = 971 evaluation cases). It turns out that, for obtaining a classifier that achieves a high performance regarding NON_COSPEC instances, only a small amount of training cases is needed, whereas, regarding the COSPEC cases, a corpus of at least 4,000 sample vectors is necessary for obtaining a performance near the level that was empirically observed during cross-validation.[22] Regarding the POS3 classifier (957 CO + 4,302 NON_CO = 5,259 training cases, 109 CO + 475 NON_CO = 584 evaluation cases), the situation is similar (see Figure 6). Hence, in particular, the COSPEC case recognition curve illustrates that training data sparsity is an issue: at least a training corpus of size 5,000 (1,000 COSPEC instances) seems to be needed for approaching the 60% level that was empirically observed to be achievable.

5.4 Extrinsic (application-level) cross-validation

The results of a *6-fold cross-validation at the application (anaphor resolution) level* are displayed in Table 6. According to the method outlined in Section 4.5, the data have been randomly split into six subsets d_{si}, $1 \leq i \leq 6$, of eleven

[22] Partly, this may be regarded as a consequence of the fact that the training corpus contains about five times more NON_COSPEC instances than COSPEC cases (7,356 vs. 1,378). However, even upon a respective rescaling of the x-axis, still the NON_COSPEC recognition accuracy curve reaches its empirical optimum faster.

documents each. Hence, there are six base experiments with differing training set / evaluation set assignments, viz. $[d_1^{66} \setminus d_{si}, d_{si}]$, $1 \leq i \leq 6$.

	antecedents (P_{ia}, R_{ia})		anchors (P_{na}, R_{na})	
experiment	PER3	POS3	PER3	POS3
$(1_{nc}^{tc},h)$ $[d_1^{31},d_{32}^{66}]$, cf. Table 4	(0.65,0.64)	(0.76,0.76)	(0.62,0.61)	(0.73,0.73)
(ds1) $[d_1^{66} \setminus d_{s1},d_{s1}]$	(0.71,0.70)	(0.90,0.90)	(0.67,0.65)	(0.79,0.79)
(ds2) $[d_1^{66} \setminus d_{s2},d_{s2}]$	(0.59,0.59)	(0.70,0.70)	(0.51,0.51)	(0.59,0.59)
(ds3) $[d_1^{66} \setminus d_{s3},d_{s3}]$	(0.72,0.72)	(0.72,0.72)	(0.73,0.72)	(0.70,0.70)
(ds4) $[d_1^{66} \setminus d_{s4},d_{s4}]$	(0.82,0.82)	(0.80,0.80)	(0.82,0.82)	(0.74,0.74)
(ds5) $[d_1^{66} \setminus d_{s5},d_{s5}]$	(0.59,0.59)	(0.76,0.76)	(0.53,0.53)	(0.70,0.70)
(ds6) $[d_1^{66} \setminus d_{s6},d_{s6}]$	(0.52,0.52)	(0.69,0.69)	(0.45,0.45)	(0.56,0.56)
(ds1-6) cumulated / averaged	**(0.66,0.66)**	**(0.75,0.75)**	**(0.62,0.62)**	**(0.68,0.68)**

Table 6: 6-fold cross-validation of anaphor resolution results

Regarding the results of the six base experiments, the variance is considerable. Similar observations have been made during the evaluation of the manually designed ROSANA system. Thus, rather than indicating a specific problem of the machine learning approach, the variance seems to be determined by the individual empirical difficulty of the document sets with respect to the anaphor resolution task. With the exception of the nonpronominal anchors result for possessives, which is lower (-5%), the cumulated score (ds1-6) lies close to the figures determined in the $(1_{nc}^{tc},h)$ experiment. One might expect that, since the training sets are considerably larger than in the original $[d_1^{31},d_{32}^{66}]$ experiment (on average, (8,734;5,259) vs. (4,446;2,670)), results should be better, particularly for possessive pronouns. One should, however, keep in mind that the learning characteristics of the classifiers are only indirectly mirrored in the anaphor resolution performance (see the discussion in Section 5.3); in particular, this holds for the secondary discipline of nonpronominal anchor determination. Hence, though it should certainly be instructive to re-run the experiments on larger data sets, the results of the extrinsic (application-level) cross-validation can be interpreted as confirming the order of magnitude of the figures obtained in the original experiment $(1_{nc}^{tc},h)$.

5.5 Trading off recall for precision

Based on the empirically optimal configuration $(1_{nc}^{tc},h)$, a series of further experiments has been carried out to address the question whether, by looking at the additional quantitative information given at the leaves of the C4.5 decision trees, it is possible to gradually bias ROSANA-ML towards *high precision anaphor resolution*. Each decision tree leaf provides the total number μ of *training cases* that match the respective decision path, and the number $\varepsilon \leq \mu$ of these cases that are, through the category prediction of the leaf, *misclassified*. By computing the quotient ε/μ, it should thus be possible to derive an estimate of the classification error probability of the specific leaf.

This information can now be used to gradually bias ROSANA-ML towards high precision anaphor resolution. The base version of the algorithm specified in Figure 4 prefers candidates predicted to COSPECify over candidates predicted to NON_COSPECify, and employs surface-topological distance as the secondary criterion. By looking at the quotient ε/μ, this preference criterion may be refined as follows: prefer COSPEC candidates over NON_COSPEC candidates; at the secondary level, prefer COSPEC candidates with smaller classification error estimate ε/μ over COSPEC candidates with higher ε/μ, and prefer NON_COSPEC candidates with higher classification error estimate ε/μ over NON_COSPEC candidates with lower ε/μ. Finally, by setting a *threshold* $\theta := (\theta_{co},\theta_{nonco})$ as illustrated in Figure 7, i.e. by eliminating all COSPEC candidates the classification error estimate of which falls *above* (>) θ_{co}, and by eliminating all NON_COSPEC candidates the classification error estimate of which falls *below* ($\leq$) θ_{nonco}, a bias may be imposed that gradually trades off recall for precision.

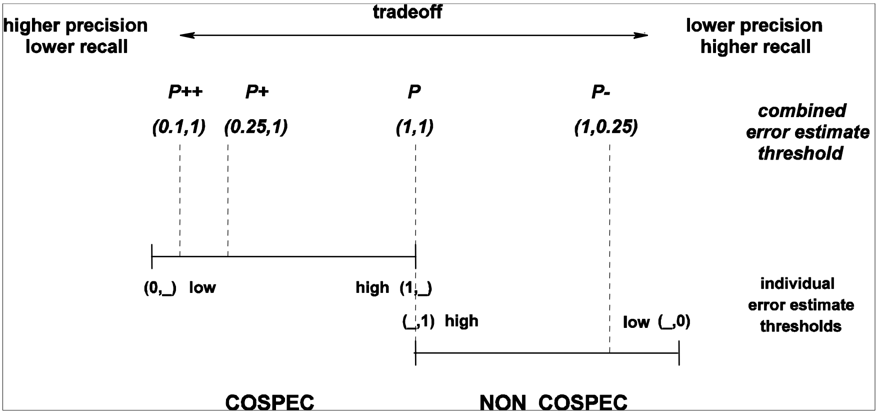


Figure 7: combined error estimate threshold for trading off recall for precision

In Table 7, the results of four respective experiments with different threshold settings are displayed. The basic precision bias setting of experiment $(1_{nc}^{tc},h,p)$ allows ε/μ values of $\leq$ 1 for COSPEC-predicted instances, and > 1 for NON_COSPEC-predicted instances; in other words, all and only the candidates that are predicted not to cospecify are eliminated. The precision bias can be weakened by eliminating only those candidates the NON_COSPEC prediction of which is incorrect with estimated probability falling below a threshold $\theta < 1$, e.g. $\theta = 0.25$ (experiment (1_{nc}^{tc},h,p−)). Similarly, the bias can be strengthened by imposing lower error ratio thresholds for the candidates predicted to COSPECify (experiments (1_{nc}^{tc},h,p+) and (1_{nc}^{tc},h,p++)). Regarding the primary evaluation discipline of immediate antecedency, the scores for the different settings indicate that, as expected, by employing the quantitative information given at the decision tree leaves in the above-described way, one obtains a suitable (albeit heuristic) means for gradually biasing ROSANA-ML towards high precision. Regarding the nonpronominal anchor discipline, which is more indirectly related to the classifier predictions, this conjecture can be regarded to be confirmed.

	antecedents *(P_{ia},R_{ia})*		anchors *(P_{na},R_{na})*	
experiment	PER3	POS3	PER3	POS3
(1_{nc}^{tc},h) [d_1^{31},d_{32}^{66}], cf. Table 4	(0.65,0.64)	(0.76,0.76)	(0.62,0.61)	(0.73,0.73)
(1_{nc}^{tc},h,p) = **(**1_{nc}^{tc},h) $\wedge\ \theta$ = (1.0,1.0)	(0.79,0.51)	(0.86,0.60)	(0.75,0.45)	(0.83,0.54)
(1_{nc}^{tc},h,p−) = **(**1_{nc}^{tc},h) $\wedge\ \theta$ = (1.0,0.25)	(0.74,0.56)	(0.78,0.63)	(0.71,0.52)	(0.76,0.59)
(1_{nc}^{tc},h,p+) = **(**1_{nc}^{tc},h) $\wedge\ \theta$ = (0.25,1.0)	(0.81,0.45)	(0.89,0.50)	(0.74,0.36)	(0.67,0.30)
(1_{nc}^{tc},h,p++) = **(**1_{nc}^{tc},h) $\wedge\ \theta$ = (0.1,1.0)	(0.83,0.31)	(1.00,0.17)	(0.80,0.08)	(1.00,0.12)

Table 7: evaluation results: trading off recall for precision

6 Comparison

6.1 ROSANA-ML vs. ROSANA

The comparison of the evaluation results for experiment (1_{nc}^{tc},h) (see Table 4) with the scores of the manually designed ROSANA system on [d_1^{31},d_{32}^{66}] (see Table 2) leads to a nonuniform assessment. Whereas ROSANA-ML performed better with respect to nonpronominal anchors for possessive pronouns, the results for non-possessives deteriorated. At first sight, this is surprising since, as observed in Section 5.3, regarding the classifier for possessives, the training set size of 2,670 is too small to arrive at the possible accuracy level of around 60% with respect to the recognition of COSPEC cases. However, one has to take into account that, according to the results that have been determined for the manually designed ROSANA system, possessives are generally easier to resolve than non-possessives.

In view of the efforts that went into the refinement of the preference factors employed in manually designed systems, the results can be regarded to be encouraging. With a comparatively low amount of training data, a performance regarding possessives has been achieved that at least reaches, if not outperforms the results of the hand-tuned ROSANA approach.[23] The inferior results on non-possessives can be interpreted as an indicator that the inventory of feature sets over which the signatures are defined should be enlarged. According to the learning curve analysis in Section 5.3, at least in the extrinsic (application-level) cross-validation experiments, the training set size should have been sufficiently large (> 8,000) to arrive at the possible accuracy level of around 60% with respect to the recognition of COSPEC cases. This gives evidence that, for arriving at an anaphor interpretation performance on non-possessives similar to the performance of manually designed systems, a COSPEC accuracy of 60% does not suffice, and, moreover, that yet not a sufficient inventory of features (Table 1) is available.

6.2 ROSANA-ML vs. (Aone & Bennett, 1995)

In their machine learning approach to anaphor resolution of Japanese texts, Aone and Bennett (1995) determine *(P,R)* figures regarding four types of anaphoric expressions: names, definite NPs, quasi-zero pronouns, and zero pronouns. The investigation is restricted to anaphoric expressions that specify organizations. Hence, their findings do not immediately compare with the evaluation results given above, which have a more general scope. A first, coarse impression, however, may be obtained by comparing the results regarding possessive and non-possessive pronouns with the cases of Japanese quasi-zero and zero pronouns, for which Aone and Bennett give immediate antecedency figures of (0.85,0.64) and (0.76,0.38). Under the assumption that similar definitions of the precision and recall measures are employed,[24] these results can be compared to the scores of the high precision anaphor resolution experiments that are summarized in Table 7. Whereas the quasi-zero pronoun figures (0.85,0.64) seem to indicate (at least when compared to the immediate antecedency scores for non-possessives) that the Aone and Bennett (1995) approach outperforms ROSANA-ML, evidence is to the contrary if one takes the zero pronoun figures (0.76,0.38) as the base of comparison. As pointed out by Aone and Bennett, quasi-zero pronouns are easier to resolve since, by definition, they always cospecify with a local subject, and, hence, may be interpreted by purely syntactical means. Consequently, the zero pronoun scores

[23] According to the results of the 6-fold extrinsic (application-level) cross-validation given in Table 6, the performance regarding possessives is expected to be, on average, lower than observed for the $(1_{nc}^{tc},h)$ experiment on $[d_1^{31},d_{32}^{66}]$. At least the level of the ROSANA scores on $[d_1^{31},d_{32}^{66}]$ has been reached; however, a more instructive comparison should be based on data obtained from a comparable in-depth cross-validation of ROSANA.

[24] Aone and Bennett (1995) do not give formal definitions of the employed measures; however, there is clear evidence that the measures are equivalent, or nearly equivalent, to the measures used in the article at hand.

can be regarded as the more suitable reference for comparison, thus urging upon the conclusion that the methodology of ROSANA-ML, according to which machine learning is applied to derive anaphor resolution *preference* strategies, is superior to the unfocused learning approach employed by Aone and Bennett (1995), in which preferences as well as restrictions are learned.

6.3 ROSANA-ML vs. CogNIAC

Baldwin (1997) describes the CogNIAC approach that achieves high precision coreference resolution by restricting antecedent decisions to cases in which no world knowledge or sophisticated linguistic processing seems to be needed for successful resolution. The recognition of such cases is performed by a set of six manually designed rules. The resolution of only those pronouns is tried the interpretation context of which matches one of these rules; all other pronouns remain unresolved. While it remains unclear whether the employed formal *(P,R)* measures neatly match up with the evaluation criteria used above, the evaluation figures of (0.92,0.64), which were obtained on a corpus with 298 cases of English third person pronouns, seem to give evidence that the manual design of a high precision rule set outperforms the machine-learning-based approach which has been obtained above as a side-product by referring to quantitative information available at the decision tree leaves. However, it has to be taken into account that Baldwin (1997) manually corrected the preprocessing results in order to allow for a fair comparison of his approach with the non-robust algorithm of Hobbs (1978), which employs complete and unambiguous parses. In fact, results of recent experiments indicate that, on potentially noisy data and without intellectual intervention, the ROSANA-ML approach to high precision anaphor resolution at least performs on a par with a *robust* reimplementation of Baldwin's algorithm (Stuckardt, 2003).

Moreover, it remains to be investigated whether even better *(P,R)* tradeoffs are obtained if decision trees that have been learned over larger amounts of training data are employed. Classifiers of higher quality are expected to yield better estimates of the classification error probability as defined in Section 5.5.

7 Conclusion and Further Research

Overall, the evaluation results of ROSANA-ML are promising. According to the above experiments, it can be concluded that, by employing a machine learning approach to preference strategies for anaphor resolution, results that at least compare with those of the best manually tuned systems can be reached. With respect to the current best settings and regarding possessive third person pronouns, the resolution quality is slightly higher than for the ancestor system ROSANA, whereas regarding non-possessives, the quality slightly lags behind.

The cross-validation scores range from 75% (possessives, immediate antecedency) to 62% (non-possessives, nonpronominal anchors). Moreover, the investigation has given evidence that, by biasing ROSANA-ML towards precision, better *(P,R)* tradeoffs can be obtained than those achieved by the approach of Aone and Bennett (1995). While this can be interpreted as an indicator that the approach employed by ROSANA-ML, which focuses on machine learning *preferences*, may lead to a better overall performance, it has to be kept in mind that the cases of English third-person pronouns and Japanese zero-pronouns do not immediately compare.

Future efforts should focus on the goal of enhancing the interpretation quality regarding non-possessives. According to the results of the above evaluation, most certainly this will require that the set of features over which the classifiers are learned is appropriately supplemented. Finding suitable candidate features, however, can be considered to be a hard and time-consuming intellectual task, thus illustrating that machine learning approaches do not generally free the knowledge engineer from intellectual fine-tuning. In this specific case, the task may be immediately compared with the intellectual determination of suitable robustly computable salience factors; however, the application of decision-tree learning saves part of the time necessary for optimising the playing together of the overall set of factors in the classical approaches to anaphor resolution.

Larger sets of features over which classifiers are learned will enhance the need for bigger training corpora. This key issue should be considered for further reasons. First, as outlined in Section 4, parts of the first two levels of experimental consideration will become obsolete: if enough training data are available, C4.5 will be able to discover for itself which features are key and which are not, thus freeing the knowledge engineer from experimenting with subset signatures, or from artificially enlarging the set of training cases. Moreover, of paramount importance is the availability of sufficiently large corpora of *different text genres*, which is the enabling condition for empirically addressing the issue of genre-specific preference strategy assignment, a goal that has been put forward as a consequence of the evaluation results of the manually designed ROSANA system.

Based on these further experiments on larger and heterogeneous corpora, the learned classifiers should undergo a thorough *qualitative* analysis. Which features do typically occur at or near the root of the learned decision trees? Which features are typically eliminated during pruning? Are there certain characteristics that are specific to the different training corpus genres? Regarding the qualitative exploration of classifiers, it should be worthwhile to look at C4.5 classifiers in the *lists of rules* format, which are generated by the classifier learning tool *C4.5rules* of the employed C4.5 implementation (see Section 3.3). From the point of view of knowledge engineering, besides having available enhanced pruning options, an important advantage of employing rules instead of trees lies in their better intellectual accessibility. The qualitative analysis of classifiers might ultimately shed new light on the empirical foundation of classical strategies for determining salience, including theories of attentional focusing such as centering.

References

Aone, C. and S. Bennett 1995. "Evaluating Automated and Manual Acquisition of Anaphora Resolution Strategies." *Proc. of the 33rd Annual Meeting of the ACL.* 122-129.

---------- 1996. "Applying Machine Learning to Anaphora Resolution." *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing.* S. Wermter, E. Riloff and G. Scheler (eds.), 302-314. Berlin: Springer.

Baldwin, B. 1997. "CogNIAC: High Precision Coreference with Limited Knowledge and Linguistic Resources." *Proc. of the ACL'97/EACL'97 Workshop on Operational Factors in Practical, Robust Anaphor Resolution for Unrestricted Texts.* R. Mitkov and B. Boguraev (eds.), 38-45.

Chomsky, N. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Connolly, D., J. Burger and D. Day 1994. "A Machine Learning Approach to Anaphoric Reference." *Proc. of the Int. Conference on New Methods in Language Processing (NEMLAP).*

Dagan, I. and A. Itai 1990. "Automatic Processing of Large Corpora for the Resolution of Anaphora References." *Proc. of the 13th Int. Conference on Computational Linguistics (COLING).* III, 330-332.

----------, J. Justeson, S. Lappin, H. Leass and A. Ribak 1995. "Syntax and Lexical Statistics in Anaphora Resolution." *Applied Artificial Intelligence* 9:6.633-644.

Ge, N., J. Hale and E. Charniak 1998. "A Statistical Approach to Anaphora Resolution." *Proc. of the 6th Workshop on Very Large Corpora*, 161-170.

Grishman, R. and B. Sundheim 1996. "Design of the MUC-6 Evaluation." *Proc. of the 6th Message Understanding Conference (MUC-6), 1-11.* San Mateo, CA: Morgan Kaufmann.

Hirschman, L. 1998. "MUC-7 Coreference Task Definition, Version 3.0." *Proc. of the 7th Message Understanding Conference (MUC-7).* Published online, available (December 9, 1999) at http://www.itl.nist.gov/iaui/894.02/related_projects/muc/proceedings/co_task.html

Hobbs, J. 1978. "Resolving Pronoun References." *Lingua* 44.311-338.

Järvinen, T. and P. Tapanainen 1997. *A Dependency Parser for English*. Technical Report TR-1. University of Helsinki: Department of General Linguistics.

Kennedy, C. and B. Boguraev 1996. "Anaphora for Everyone: Pronominal Anaphora Resolution without a Parser." *Proc. of the 16th Int. Conference on Computational Linguistics (COLING), Copenhagen*, I, 113-118.

Lappin, S. and H. Leass 1994. "An Algorithm for Pronominal Anaphora Resolution." *Computational Linguistics* 20(4).535-561.

McCarthy, J. and W. Lehnert 1995. "Using Decision Trees for Coreference Resolution." *Proc. of the 14th Int. Joint Conference on Artificial Intelligence (IJCAI'95), Montreal.*

Mitchell, T. 1997. *Machine Learning.* New York: McGraw-Hill.

Mitkov, R. 1997. “Factors in Anaphora Resolution: They are not the Only Things that Matter. A Case Study Based on Two Different Approaches.” *Proc. of the ACL’97/EACL’97 Workshop on Operational Factors in Practical, Robust Anaphor Resolution for Unrestricted Texts*, R. Mitkov and B. Boguraev (eds.), 14-21.

---------- 2001. “Outstanding Issues in Anaphora Resolution.” *Proc. of the 2nd Int. Conference on Computational Linguistics and Intelligent Text Processing (CICLing), Mexico City,* A. Gelbukh (ed.), 110-125.

Paul, M., K. Yamamoto and E. Sumita 1999. “Corpus-Based Anaphora Resolution Towards Antecedent Preference.” *Proc. of the ACL’99 Workshop on Coreference and its Applications,* 47-52.

Quinlan, J. 1993. *C4.5: Programs for Machine Learning.* San Mateo, CA: Morgan Kaufmann.

Soon, W., H. Ng and D. Lim 2001. “A Machine Learning Approach to Coreference Resolution of Noun Phrases.” *Computational Linguistics* 27(4).521-544.

Stuckardt, R. 1997. “Resolving Anaphoric References on Deficient Syntactic Descriptions.” *Proc. of the ACL’97/EACL’97 Workshop on Operational Factors in Practical, Robust Anaphor Resolution for Unrestricted Texts.* R. Mitkov and B. Boguraev (eds.), 30-37.

---------- 2001. “Design and Enhanced Evaluation of a Robust Anaphor Resolution Algorithm.” *Computational Linguistics* 27(4). 479-506.

---------- 2003. “Coreference-Based Summarization and Question Answering: a Case for High Precision Anaphor Resolution.” *Proc. of the 2003 Int. Symposium on Reference Resolution and Its Application to Question Answering and Summarization (ARQAS), Università Ca’ Foscari, Venice,* 33-41.

Vilain, M., J. Burger, J. Aberdeen, D. Connolly and L. Hirschman 1996. “A Model-Theoretic Coreference Scoring Scheme.” *Proc. of the 6th Message Understanding Conference (MUC-6),* 45-52. San Mateo, CA: Morgan Kaufmann.

Decomposing Discourse

Joel Tetreault
University of Rochester

This paper presents an automated empirical evaluation of the relationship between clausal structure and pronominal reference. Past work has theorized that incorporating discourse structure can aid in the resolution of pronouns since discourse segments can be made inaccessible as the discourse progresses and the focus changes. As a result, competing antecedents for pronouns from closed segments could be eliminated. In this study, we develop an automated system and use a corpus annotated for rhetorical relations and coreference to test whether basic formulations of these claims hold. In particular, we look at naive versions of Grosz and Sidner's theory and Kameyama's intrasentential centering theories. Our results show that incorporating basic clausal structure into a leading pronoun resolution does not improve performance.

1 Introduction

In this paper, we present a corpus-based analysis using Rhetorical Structure Theory (RST) to aid in pronoun resolution. Most implemented pronoun resolution methods in the past have used a combination of focusing metrics, syntax, and light semantics (see (Mitkov, 2000) for a leading method) but very few have incorporated discourse information or clausal segmentation. It has been suggested that discourse structure can improve the accuracy of reference resolution by closing off unrelated segments of discourse from consideration. However, until now, it has been extremely difficult to test this theory because of the difficulty in annotating discourse structure and relations reliably and for a large enough corpus. What limited empirical work that has been done in this area (Poesio & Di Eugenio, 2000; Ide & Cristea, 2000) has shown that structuring discourse can successfully constrain the search space for antecedents. In this paper, we use a different metric, simply, how many pronouns one can resolve correctly with a constrained search space.

The RST-tagged Treebank (Carlson et al., 2001) corpus of Wall Street Journal articles merged with coreference information is used to test this theory. In addition, an existing pronoun resolution system (Byron & Tetreault, 1999) is augmented with modules for incorporating the information from the corpus: discourse structure and relations between clauses. The experiments involve breaking up an utterance into clausal units (as suggested in (Kameyama, 1998)) and basing the accessibility of entities and the salience of entities on the hierarchical structure imposed by RST. We also compare a leading empirical method, the Veins Theory (Ide & Cristea, 2000), with our approaches. Our results show that basic methods of decomposing discourse do not improve performance of pronoun resolution methods.

In the following Section, we discuss theories that relate discourse and anaphora. Next, we discuss two experiments: the first determines the baseline algorithm to be compared against and the second tests different metrics using RST and its relations. Finally, we close with results and discussion.

2 Background

2.1 Discourse Structure

We follow Grosz and Sidner's (1986) work in discourse structure in implementing some of our clausal-based algorithms. They claim that discourse structure is composed of three interrelated units: a linguistic structure, an intentional structure, and an attentional structure. The linguistic structure consists of the structure of the discourse segments and an embedding relationship that holds between them.

The intentional component determines the structure of the discourse. When people communicate, they have certain intentions in mind and thus each utterance has a certain purpose to convey an intention or support an intention. Grosz and Sidner call these purposes “Discourse Segment Purposes” or DSP's. DSP's are related to each other by either dominance relations, in which one DSP is embedded or dominated by another DSP such that the intention of the embedded DSP contributes to the intention of the subsuming DSP, or satisfaction-precedent relations in which satisfying the intentions of a DSP is necessary to satisfy the intentions of the next DSP. Given the nesting of DSP's, the intentional structure forms a tree, with the top node being the main intention of the discourse. The intentional structure is more difficult to compute since it requires recognizing the discourse purpose and the relation between intentions.

The final structure is the attentional state which is responsible for tracking the participant's mental model of what entities are salient or not in the discourse. It is modelled by a stack of focus spaces, which is modified by changes in attentional state. This modification process is called focusing and the set of focus spaces available at any time is the focusing structure. Each discourse segment has a focus space that keeps track of its salient entities, relations, etc. Focus spaces are removed (popped) and added (pushed) from the stack depending on their respective discourse segment purpose and whether or not their segment is opened or closed. The key points about attentional state are that it maintains a list of the salient entities, prevents illegal access to blocked entities, is dynamic, and is dependent on intentional state.

To our knowledge, there has been no large-scale annotation of corpora for intentional structure. In our study, we use Rhetorical Structure Theory, or RST, (Mann & Thompson, 1988) to approximate the intentional structure in Grosz and Sidner's model. RST is intended to describe the coherence texts by labelling relations between clauses. The relations are binary. Thus, after a text has been completely labelled, it is represented by a binary tree in which the interior nodes are relations. With some sort of segmentation and a notion of clauses, one can test pushing and popping, using the depth of the clause in relation to the surrounding clauses.

Using RST to model discourse structure is not without precedent. Moser and Moore (1996) first claimed that the two were quite similar in that both had hierarchal tree structures and that while RST had explicit nucleus and satellite labels for relation pairs, DSP's also had the implicit salience labels, calling the primary sentence in a DSP a "core," and subordinating constituents "contributors." However, Poesio and DiEugenio (2001) point out that an exact mapping is not an easy task as RST relations are a collection of intentional but also informational relations. Thus, it is not clear how to handle subordinating DSP's of differing relations and therefore, it is unclear how to model pushes and pops in the attentional stack.

2.2 Centering Theory

Centering (Grosz *et al.*, 1995) is a theory that models the local component of the attentional state, namely how the speaker's choice of linguistic entities affects the inference load placed upon the hearer in discourse processing. For instance, referring to an entity with a pronoun signals that the entity is more prominently in focus.

In Centering, entities called centres link an utterance with other utterances in the discourse segment. Each utterance within a discourse has a backward looking center (Cb) and forward looking centres (Cf). The backward-looking centre represents the most highly ranked element of the previous utterance that is found in the current utterance. Basically, the Cb serves as a link between utterances. The set of forward-looking centres for an utterance U_0 is the set of discourse entities evoked by that utterance. The Cf set is ranked according to discourse salience; the most accepted ranking is grammatical role (by subject, direct object, indirect object). The highest ranked element of this list is called the preferred centre or Cp. Abrupt changes in discourse topic are reflected by a change of Cb between utterances. In discourses where the change of Cb is minimal, the preferred centre of the utterance represents a prediction of what the backward-looking centre will be in the next utterance. In short, the interaction of the topic, and current and past salient entities can be used to predict coherence as well as constrain the interpretation of pronouns.

Given the above, Grosz *et al.* proposed the following constraints and rules of centering theory:

Constraints:

For each utterance U_i in a discourse segment D, consisting of utterances of $U_1 \ldots U_m$:

1.There is precisely one backward looking centre.

2. Every element of the Cf list for U_i must be realized in U_i

3. The centre: $Cb(U_i, D)$, is the highest ranked element of $Cf(U_{i-1}, D)$ that is realized in U_i.

Rules:

For each utterance U_i, in a discourse segment D, consisting of utterances of $U_1 \ldots U_m$:

1. If some element of $Cf(U_{i-1},D)$ is realized as a pronoun in U_i, then so is $Cb(U_i,D)$.

2. The relationship between the Cb and Cp of two utterances determines the coherence between the utterances. The Centering Model ranks the coherence of adjacent utterances with transitions, which are determined by whether or not the backward looking centre is the same from U_{i-1} to $U_{i,}$ and whether or not this entity coincides with the preferred centre of U_i. Transition states are ordered such that a sequence of “continues” (where the Cb's and Cp are the same entity) is preferred over a “retain,” which is preferred to a “smooth shift” and then to a “rough shift.”

2.3 Long-Distance Pronominalization

Following Centering theory, pronouns are typically used when referring to salient items in the current discourse segment, that is, their antecedents are generally very focused and found in the local text area. This tendency is supported by corpus statistics, which show that an overwhelming majority of the antecedents of pronouns are found in the current or previous utterance (Kameyama, 1998; Hitzeman & Poesio, 1998; Tetreault, 2001). However, there are cases in which a pronoun is used to refer to an entity not in the current discourse segment. Consider the dialogue from Allen (1994, p. 503) between two people E and A discussing engine assembly.

1. E: So you have the engine assembly finished.
2. E: Now attach the *rope* to the top of the engine.
3. E: By the way, did you buy gasoline today?

4.	A:	Yes. I got some when I bought the new lawn mower wheel.
5.	A:	I forgot to take my gas can with me, so I bought a new one.
6.	E:	Did it cost much?
7.	A:	No, and I could use another anyway to keep with the tractor.
8.	E:	OK.
9.	E:	Have you got *it* attached yet?

Figure 1: Long Distance Pronominalization example

The *it* in (9) refers to the *rope* which is seven sentences away in (2) and also has several possible antecedents in the immediate context. One can account for this long-distance pronominalization using the Grosz and Sidner approach by considering sentences (3)-(8) a discourse segment embedded in the segment (1) through (9). The phrase “By the way” can be viewed as a cue phrase that a new discourse state is being started (a push on the attentional stack) and that (8) completes the segment and the state is popped from the top of the stack. With these intervening sentences “removed” it is easy to resolve *it* to the *rope* since the *rope* is the most salient object on the top of the attentional stack.

Although cases of long-distance pronominalization are rare, the phenomenon is important because it can be used as a metric to determine whether or not an application of a pronoun resolution strategy with discourse segmentation is successful or not. Typically, one would not expect recency-based algorithms to be successful in these cases, but algorithms equipped with knowledge of the discourse would be.

In related work, Hitzeman and Poesio (1998) developed an algorithm for addressing long-distance anaphors. They augment the Grosz and Sidner attentional state model by associating a discourse topic with each focus space that says what the state is about, and it can be a proposition, object, etc. In addition, the focus space could have associated with it a “Most Salient Entity” (MSE) which is the most important entity explicitly introduced in the segment. In the case of pronouns, an antecedent is first searched in the local context (if any), and then through past MSE's of open discourse segments.

Walker (2000) analysed 21 cases of long-distance resolution to support her claim that a cache is a better model of the attentional state than a stack. She supports Fox's proposal (1987) that lexical repetition can be used to signal a return pop. That is, a pronoun with a long-distance referent is often found in a sentence that has a similar verb to the space being referred to, and that this “informational redundancy” can serve as a trigger not to search the local segment but a past segment.

2.4 Clause-Based Centering

One of the major underspecifications of centering theory (Poesio *et al.*, 2000) is the notion of what constitutes an utterance. Most work in the field ignores the issue by treating a sentence as the minimal discourse unit. This is problematic because large, complex sentences could have clauses that interact with each other in a non-linear manner. Because of this problem, Kameyama (1998) developed theories on the interaction of clauses and updating centering constructs. In the clause-based centering proposal, sentences can be broken up into two classes: sequential and hierarchical. In the sequential decomposition, the sentence is broken up into several utterances whose centering output is the input to the following utterance. In hierarchical decomposition, each utterance does not necessarily affect the following utterance. It is possible that a combination of utterances affect the centering input state to another utterance or that an utterance will not affect any other utterance. Kameyama views this sentence decomposition as a tree.

3 Evaluation 1: Baseline Selection

Determining the usefulness of incorporating discourse information in reference resolution requires a large corpus annotated with coreference and clausal information, and a system to try different algorithms. In the following Sections, we discuss our corpus, our testbed system for extracting noun-phrase entities, and finally the algorithms and their results. After testing each algorithm on the same corpus, the best one is selected as the baseline algorithm. If discourse or clausal information is used correctly, we should see an improvement over this baseline algorithm.

3.1 Corpus Description

The test corpus was constructed by merging two different annotations of a subset of the Penn Treebank (Marcus *et al.*, 1993). The news articles cover such different topics as reviews of TV shows and the Japanese economy. The portion of the Treebank consists of 52 Wall Street Journal articles which includes 1241 sentences and 454 non-quoted third person pronouns that refer to NP entities.

Carlson *et al.* (2001) annotated those articles with rhetorical structure information in the manner of Mann and Thompson (1988) with very high annotator reliability. This annotation breaks up each discourse into clauses connected by rhetorical relations. So from this work there is a decomposition of sentences into a smaller units (a total of 2755 clauses) as well as a discourse hierarchy for each article and relations between pairs of segments.

The corresponding Penn Treebank syntactic structures for each sentence were also annotated with coreference information in the same manner as

Ge *et al.* (1998). This meant that all third-person pronouns were marked with a specific identification number and all instances of the pronoun's antecedent were also marked with the same id. In addition, the Penn Treebank includes annotations for the syntactic tree structures of each sentence, so syntactic attributes such as part-of-speech and number information were extracted. Also, each noun phrase entity was marked manually for gender information.

Finally, the RST corpus and the Penn Treebank coreference corpus were merged such that each discourse entity (in this case, only noun-phrases) had information about its syntactic status, gender, number, coreference, etc. and the following discourse information: the clause it is in, the depth of the clause in the RST tree, and the rhetorical relations that dominate the clause. The Penn Treebank data and only the clausal breakdown of each sentence are used in this evaluation. In the second evaluation, all of the RST data comes into play.

3.2 Algorithms

One of the problems with reporting the performance of a pronoun resolution algorithm is that researchers often test on different corpora so it is hard to compare results. For example, an algorithm tested on a news corpus may perform differently on a corpus of short stories. In this particular experiment, we have a common corpus to test different algorithms, with the goal of simply selecting the best one to use as a baseline for comparison with schemes that incorporate clausal information. We examine three well-known pronoun resolution methods: Left-Right Centering (Tetreault, 1999, 2001), the S-list algorithm (Strube, 1998), and Brennan *et al.*'s centering algorithm (1987), in addition to a naïve metric. The naive metric involves searching through a history list starting with the last mentioned item and selecting the first one that meets gender, number, and syntactic constraints. All four algorithms are primarily syntax-based. Because of this limitation they should not be expected to fare too well in interpreting pronouns correctly since proper interpretation requires not only syntactic information but also semantics and discourse information.

Left-Right Centering (henceforth LRC) is based on Centering Theory in that it uses salience (calculated from grammatical function) to choose which entities should be the antecedents of anaphoric entities. However, the algorithm does not use Rule 2 or the notion of a Cb in determining referents. LRC works by first searching the current utterance left-to-right for an antecedent that matches number, gender and syntactic constraints. If one is not found, then it searches past Cf-lists left-to-right (in which the entities are ranked from most salient to least salient) until an antecedent is found.

The S-list algorithm differs from LRC in two ways: first, it maintains only one list (the Salience list) that is updated incrementally as the sentence is processed; second, the S-list is sorted not by grammatical function but by Prince's familiarity scale (1981) which classifies entities by whether they are new to the discourse, mediated (inferable) or hearer-old. Relying solely on syntax does not do the S-list justice since determining the status of certain entities (such as inferrables in the mediated category) requires a combination of syntactic and semantic knowledge. So our implementation is an approximation of the original formulation.

The final method is the original centering algorithm by Brennan *et al.* (henceforth BFP) which makes full use of the centering rules and transitions. The algorithm works by first generating all possible Cb-Cf combinations, filtering out combinations that violate syntactic constraints, and then ranking the remaining combinations by which transition makes the discourse the most coherent. We follow Walker (1989) by using the LRC algorithm to search for intrasentential antecedents for a pronoun first (since BFP does not specify an intrasentential algorithm) then use BFP if an intersentential search is required.

3.3 Results

Each algorithm was tested on the corpus in two different variations (see Table 1): the first is the conventional manner of treating sentences as the smallest discourse unit (S); the second involves splitting each sentence into clauses specified by the RST annotations (C), so clauses are now utterances.

Algorith m	**Right (S)**	**% Right (S)**	**Right (C)**	**% Right (C)**
LRC	376	80.8	347	76.4
S-List	333	73.4	318	70.0
BFP	270	59.5	221	48.7
Naïve	230	50.7	254	56.0

Table 1: Pronoun Resolution Algorithms over (S)entences and (C)lauses

The (S) results accord with the larger study of the same algorithms in Tetreault's 2001 paper – that the LRC performs better than the other two algorithms and that on a new corpus, one would expect the algorithm to resolve 80% of the pronouns correctly.

The (C) results are a first stab at the problem of how to incorporate clausal structure into pronoun resolution algorithms. The result is negative since each algorithm has a performance drop of at least 3%.

3.4 Error Analysis

We performed an error analysis of the 87 pronouns that the leading metric, LRC (S) resolved incorrectly. Below, the 87 errors are broken up into seven main cases (in bold). For each item, the number of errors are listed as well as a description of the error and examples. In the examples, pronouns and their antecedents are listed in italics.

(24) Minimal S. The sentence has several embedded sentential (S) clauses and the pronoun refers to the subject of the minimal S that both NP's are located in. This happens a lot for sentences such as “the committee said the *company* reneged on *its* obligations” – sentences where the verb is *think*, *consider*, *said*, *proposed*, etc.

(21) Localized errors. The antecedent of the pronoun is very local (smaller than the minimal S). Usually the pronouns are found in PP's or complements that modify an NP, which the pronoun refers to. For example (words in italics are the entities in consideration):

(1) “...to get a major *customer*'s 1,100 parcel-a-week load to *its* doorstep”

(15) Preposed Phrase. LRC has the strategy that any preposed phrase (words before the manually marked subject of the sentence) is processed after the main clause has been processed (see (Tetreault, 2001: Section 5.1) for details). This means that the subject of the sentence will skip any possible referents in the preposed phrase and search immediately in the previous sentence. If there are any pronouns in the preposed phrase that refer outside the phrase, it will search the subject first and the main clause of the sentence before moving on to previous sentences. This was a metric meant to handle cases such as:

(2) “Although *he* was really tired, *John* managed to drive 10 hours without stopping.”

However, there are three cases in which the pattern doesn't always hold in which the subject refers to an element in the preposed phrase:

(9) Reference from the matrix clause to the preposed phrase.

(3) “Until 1980 when *Japan* joined the $10,000 per capita club of the advanced countries, *it* was a model developing nation.”

In this case, Japan is the referent for it but it gets “skipped” and a non-gendered singular referent in the previous sentence is selected.

(5) Reference from an object to a subject within a preposed phrase. The current LRC method ignores this case since the pronoun only has the matrix clause and past utterances as its domain of accessibility. In the following preposed phrase:

(4) “Because *drivers* aren't permitted to load *their* own vehicles...”

the elements of the phrase are demoted to the end of the utterance’s Cf-list and so search the subject of the sentence before looking at their own subject.

(1) Reference outside a preposed phrase. There was one case where an element of a preposed phrase refers to an element in a previous sentence (Addison is a city):

(5a) "*Addison* is no stranger to cans of worms"
(5b) "After *its* previous mayor committed suicide last year, an investigation disclosed that town officials regularly voted...."

(12) Parallelism. Cases in which the clauses containing the pronoun and antecedent exhibit parallel structure, such as parallel sentences or parallel verb phrases.

(6) "It more than doubled *Federal*'s long term debt to $1.9billion, thrust the *company* into unknown territory – heavy cargo – and suddenly expanded *its* landing rights to 21 countries from 4."

The above is an example of conjoined verb phrases with reference between subjects.

(11) Competing Antecedents. There were other entities ranked higher in the Cf-list and thus without any semantic or lexical knowledge, which were selected before the actual antecedent was considered. For example:

(7) "The weight of *Lebanon*'s *history* was also against him, and *it* is a history..."

The correct antecedent of *it* is *Lebanon*'s *history* but *weight* has a higher grammatical function so LRC selects it as the referent.

(4) Cases of plural pronouns. Referring to companies which are marked as singular. Here the *Ministry* is considered a singular entity so *they* is incorrectly resolved to *conferences*.

(8a) "The *Ministry of Construction* spreads concrete throughout the country and boasts in international conferences that Japan's roadway per capita is the longest in the world."
(8b) "But *they* seldom think of the poor commuters."

(2) Genitive Errors. Incorrect reference to the possessed entity instead of the possessor.

(8) "Mr. Richardson wouldn't offer specifics regarding *Atco*'s proposed British *project*, but he said *it* would compete for customers with..."

LRC incorrectly finds the *project* as the referent for *it* when it should be *Atco*.

The main point to take from this analysis is that all of the errors that LRC made were with pronouns that had antecedents in the same or previous sentence. This suggests that a strategy that takes clausal information into account to break up complex sentences with embedded segments could improve performance.

Also of note was that LRC made no long distance pronominalization errors. There were only 10 pronouns that had a long-distance antecedent (defined as two or more sentences away from anaphor) and in most of these cases, the antecedent had no competing antecedents due to gender constraints. Thus an approach that simply goes back in the discourse history through past Cf-lists is bound to be successful. As stated earlier, there are several instances in natural language in which a pronoun with a long-distance antecedent has competing antecedents in the local context. One would expect that discourse structure would be most useful in these cases, to deter a search of the local context by blocking previous utterances from consideration. Since LRC resolves all 10 cases correctly, the best one can hope for in this evaluation here is to not worsen performance.

4 Evaluation 2: Incorporating Clausal Information

The next step in our evaluation is to see if we can improve upon the baseline of 80.8% from the LRC algorithm. In the following sections we present several different methods that use clausal and RST information to better that figure.

4.1 Clausal Algorithms

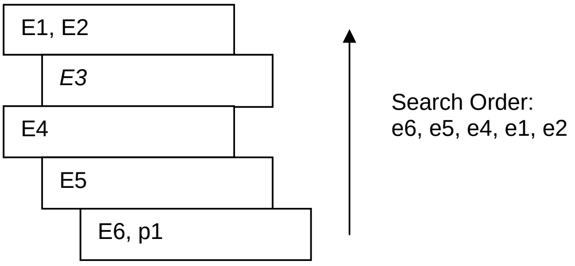


Figure 2: Grosz and Sidner accessibility

Two types of augmentations to LRC were developed for this experiment. The first uses the hierarchical structure placed on the clauses to determine whether or not an antecedent is accessible. Based on Grosz and Sidner's pushing and popping discourse structure, we work under the simple assumption that an entity is inaccessible if it is more embedded in the RST tree than the referring entity, meaning if we were explicitly tracking the attentional state, that embedded utterance would have been popped before our current utterance was processed.

Thus, the Grosz and Sidner approximation (henceforth GS) works only by considering the depth of past clauses. The algorithm works as follows: for each pronoun the attentional stack is constructed on the fly since we have perfect information on the structure of the discourse. The search works by looking through past clauses that are only the same or lower depth of the previous clause visited. The reasoning is that embedded segments that are of greater depth are not related to the entities that follow them. If they were, they would share the same embedding. Figure 2 shows how this works. Assume that clauses closest to the left are the closest to the root of the tree (lower depth). When searching for an antecedent for a pronoun in the last clause, first search all entities in it, if one is not found, then go back clause by clause until one is found. So the search order would be E6, E5, E4, then skip over E3 since it is more embedded than the current clause E4. E1 and E2, however, are accessible since it is the same depth as E4.

We evaluated three different versions of this metric: 1. using the strategy only for intrasentential searching (GS-intra); 2. using it only for intersentential searching (GS-inter); 3. using it for both (GS-both). For the first two versions, the default LRC algorithm was employed. So for version 1, if an antecedent could not be found intrasententially that met the constraints imposed by the hierarchy of clauses, then previous Cf-lists are searched using the original LRC method.

The second main metric uses the clausal structure to rank entities within the current utterance when one searches intrasententially. This method is meant to mirror Kameyama's intrasentential centering theories by modifying the ranking of the entities already processed in the current utterance. Many of Kameyama's centering constraints are difficult to automate so a direct implementation of her theories will not be evaluated here. Rather, we take a naive implementation of her theories. One of the main points of her work is that one may be able to better account for intrasentential references (and thus intersentential ones) by appealing to the clausal structure of utterances. In our study, we approximate sequential clause centering and try out different metrics for weighting the salience of clauses to approximate hierarchical decomposition.

There are variations of this strategy: depth-sort, nucleus-sort, and left-right-sort. For all three variations, the intersentential component from the original LRC algorithm is used (search each past utterances' Cf-list from most salient to least salient).

The depth-sort method works by sorting each entity by its depth in the RST tree for that discourse. The closer a clause (and thus the entities in the clause) are to the root of the tree, the more salient we assume they are. In the case of clauses being at the same depth, the clause closer to the head of the sentence is deemed more salient. Likewise, in nucleus-sort, each entity is sorted by whether it is a nucleus or satellite. So all entities in nucleus clauses are searched first before any entities in satellite clauses are searched. The order within both groupings is determined by grammatical function. The idea here is that nuclei are more salient in the discourse than satellites. The difference between depth-sort and GS-intra is that in depth-sort, all clauses (within the current utterance) are accessible, but just ranked differently to reflect their relative salience. In GS-intra, clauses that are at a lower embedding than the current clause are not searched, so this approach is more constraining and potentially more efficient since it prunes entities from consideration.

In left-right-sort (lr-sort), all intrasentential searches are done mimicking the intersentential search method except we consider clauses instead of sentences. So, instead of searching entities left-to-right in the current utterance, the most recent clause is searched first left-to-right. If an entity is not found that meets constraints, then we move to the next most recent clause, and so forth.

The results for all 6 metrics are seen in Table 2.

Algorithm	Right	% Right	Clausal	LRC	Neither
GS-both	335	73.6	36	62	51
GS-intra	347	76.4	36	54	51
GS-inter	335	73.6	36	62	51
Depth-sort	364	80.2	6	9	81
Nucleus-sort	354	78.0	11	24	76
LR-sort	265	80.4	31	33	56

Table 2: Pronoun Resolution Algorithms with Basic Clausal Metrics

The last three columns represent the overlap in errors between the clausal algorithm and LRC. The “Clausal” column is the number of pronouns that the clausal-based algorithm gets right but LRC gets wrong. Since LRC fared better than the algorithms tested here the numbers will be low in this column. The “LRC” column is the opposite – the number of pronouns the baseline metric gets right that the clausal one doesn't. The last column is the number of pronouns both resolve incorrectly.

The main result from these tests is that basic approaches do not improve performance, though the GS algorithms do have the advantage of lowering the search time. However, the “Clausal” column indicates that these algorithms do get some pronouns right that LRC does not, though fewer than the converse. But just because approximated versions of the theory do not improve accuracy does not entail that theories regarding structure are incorrect. It means that better implementations are required for validation.

One major source of error for the GS algorithms was that many antecedents in the previous sentence were deemed inaccessible due to their clausal structure. Centering theories and other focus-based theories all hold that entities in the previous utterance are very salient since they often contribute to cohesion in the discourse. The poor performance of the GS algorithms can be traced to the approximation of discourse segments. Since a RST tree is used, discourse segments are always at most two utterances long (exempting the case of multi-nuclear nodes). In addition, an entity may not be able to search its sister node, if the sister is also a tree. This means that satellites in these cases cannot refer to the nucleus they are in a relation with. In short, the segmentation entails that clauses in the local context may be incorrectly removed from consideration. The figures above automatically include using the previous utterance, even if it is deemed inaccessible. This results in a gain of 2% for the GS algorithms.

The depth-sort algorithm performs most closely to LRC because the grammatical ordering of the sentence and the clausal order imposed by depth are similar. The cases that depth-sort gets right involve sentences in which the verb phrase is considered more salient than the subject clause and the pronoun refers to an entity in the verb phrase, as opposed to the subject. A similar trend takes place with nucleus-sort – that the nucleus of a multi-clausal sentence is often the subject so both algorithms order the search the same way. However, there are a few cases where the antecedent is not the subject, but rather an entity in a nucleus class in the middle of the sentence.

The interesting trend in the overlap of LRC's and lr-sort's errors is that even though they perform roughly the same, there is not as much cross-over in errors as in the depth and nucleus instantiations. The same trend was found in the GS algorithms. It turns out this sort of the intrasentential clausal structure is very close to the Hobbs (1977) algorithm in that it traverses the parse tree right-to-left, doing a breadth-first search of sub-trees. In (Tetreault, 2001), the Hobbs algorithm and LRC were compared and found to perform roughly the same despite the opposing directions for searching through the current utterance. The majority of lr-sort's errors came from the antecedent being in the subject position but a competing referent was found before the search hit the subject clause. As expected, most of the errors that lr-sort (and the GS algorithms) got right that LRC didn't were pronouns that had localized antecedents.

4.2 Veins Algorithm

Veins Theory (Cristea *et al.*, 1998; Ide *et al.*, 2000) is an extension of Centering Theory from local to global discourse. The empirically tested method makes use of discourse structure (RST trees) to determine the accessibility of referents. The theory assumes that only a subset of the clauses preceding the anaphor are actually relevant to successfully interpreting the anaphor. This subset (domain of referential accessibility, or DRA) is determined by the interaction of the tree hierarchy and whether a clause is a nucleus or a satellite. As a result of this pruning effect, the theory has the advantage over syntax-based approaches presented in this paper by searching only a few select utterances instead of all past utterances in the discourse.

Using RST as the basis for their discourse representation, terminal nodes in the binary tree represent the clauses of the discourse, and non-terminal nodes represent the rhetorical relations. The DRA for a clause is computed in two steps. First, the "head" of each node is computed bottom-up by assigning a number to each terminal node. Non-terminal nodes are labelled by taking the union of the heads of its nuclear children. The second step, computing the "vein," is a top-down method. First, the vein of the root of the tree is the head. For every nuclear node, if it has a left sibling that is a satellite, its vein is the union of the head of the child and its parent vein, otherwise it inherits its parents vein only. For every satellite node, if the node is the left child of its parent then its vein is the union of its head with the parent's vein. Otherwise, its vein is the union of its head with the parent's vein but with all prior left satellites removed.

Finally, the DRA for a clause is simply all the nodes in the clause's vein that precede it. Intuitively, if a node has parents that are all nuclei, it will be more accessible to other entities since it is highly salient according to Veins Theory (VT). However, satellites serve to restrain accessibility. In short, the Veins notion of accessibility takes into account more information than the GS approximation previously discussed. The G&S method was based solely on the relation of the clauses in the discourse tree; the Veins approach also incorporates the nucleus information.

The Veins algorithm had to be changed slightly to run on the corpus: our corpus is not entirely binary – there were a few cases of multinuclear nodes. For these extra nodes, we simply repeated the rules for the right child. We performed two tests – first, run the algorithm over the corpus to see if it betters 81%, the baseline accuracy; second, check how many referents are actually found in the DRA of a pronoun.

The first test involves running the Veins algorithm over our corpus to compare with the other tests. The original formulation is a metric of accessibility not resolution since it does not specify how to search the DRA or how to search clauses within the DRA. Thus, Veins was implemented as follows: for every pronoun, search the clauses of its DRA from most recent to least recent, from left to right. The current clause was searched using LRC's intrasentential method. If no antecedent is found in the current clause or sentence, then past clauses which are within the entity's DRA are searched. Cases in which the antecedent is outside the pronoun's DRA are considered wrong. We also implemented a version of Veins merged with LRC (Veins++)

which takes care of referents found outside the DRA. So if the original Veins algorithm does not find an antecedent in the DRA, the LRC algorithm is used to find a suitable antecedent. This approximation makes sense given the VT claim that referents outside the DRA incur a higher processing load on the interpreter. This processing load is mirrored by having to run a comprehensive search of the discourse to find an antecedent. The results are found in Table 3 with the square brackets indicating the number of pronouns the method was unable to resolve. The sentence (S) version Veins means that the LRC intrasentential method is used to search the entire sentence the entity is in.

Algorithm	**Right (S)**	**% Right (S)**	**Right (C)**	**% Right (C)**
Veins	294 [71]	64.8	264 [87]	58.2
Veins++	358	78.9	346	76.2

Table 3: Veins Resolution Algorithms

Although the original Veins approach is more efficient than other methods we have discussed in this paper, it does not perform as well since the DRA ends up being too constrictive. As expected, Veins++ fared better since it can go back through the discourse to find referents for a pronoun that didn't have a referent in its DRA. However, even with this boost it still does not perform better than our baseline of 80.8%. Our second test provides some explanation for this result.

Error analysis showed that out of the 454 pronouns, 349 (or 76.9%) had referents in the DRA specified, but that still leaves almost a quarter with referents outside the DRA. The Veins Theory does state that entities outside the DRA are possible and that reference to them requires a higher processing load. However, we found only 10 cases of inaccessibility in which the referent was found not in the current or previous sentence. One would expect that most of the errors in this approach would come from entities that were several utterances away, not from the local context in which entities would be more accessible. Ide and Cristea note that all exceptions to accessibility in their corpus analysis come from pronouns and antecedents in attribution relations (such as “he said...”). Of the 105 inaccessible cases, 40 had both the pronoun and antecedent in attribution relations. An additional 15 had just the pronoun in such a relation and 18 of the 105 had just the referent in such a relation. However, the remaining 32 remain unaccounted for. The two most common reasons for inaccessibility were 1. that there was an intervening satellite node that blocked access to the referents even though it was one utterance or less away; and 2. there were cases of attribution-like relations.

In terms of long-distance pronominalization, the original Veins formulation was unable to resolve 6 of the 10 cases when treating sentences as the minimal discourse unit, and when considering clauses, was unable to resolve 9 of the 10 cases. All of these were pronouns and antecedents in attribution relations.

Ide and Cristea claim that a way of dealing with this attribution problem is to merge together clauses in an attribution relation. We believe this heuristic would improve performance since a significant portion of the errors Veins Theory make are due to attribution relations.

4.3 Corpus Transformations

The previous results showed that using the RST tree in the Grosz and Sidner approach and in the Veins produced too fine a segmentation and thus many clauses are deemed unfairly inaccessible. To counter this, we developed three transformations to a RST tree to flatten the tree and make more clauses accessible. The first involves replacing multi-clausal sentences with one clause in the RST tree; and the second involves merging all subtrees that have a satellite leaf in a relation with a subtree consisting of all leaves, one of which is a nucleus.

The intuition with the first transform (SENT) is that many of the errors in the original approximation of G&S based on RST are intrasentential. By merging the clauses together, the tree becomes flattened, and all entities within a sentence are accessible. An example of this transform is in Figure 3 in which one assumes the leaves are three clauses of a RST subtree and are constituents of one sentence. Doing the SENT transform yields the result in the second half of Figure 3, a subtree that is now a leaf of the sentence reconstructed.

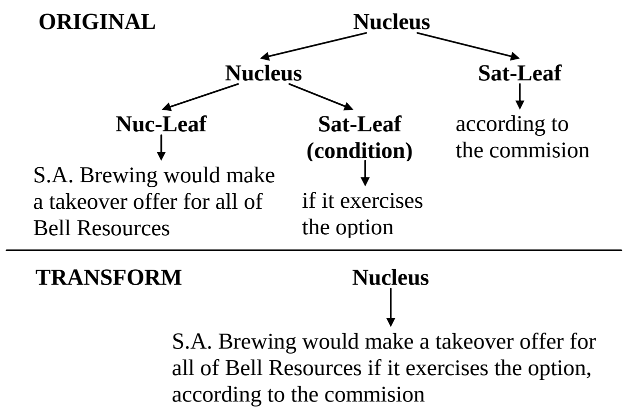


Figure 3: SENT transform example

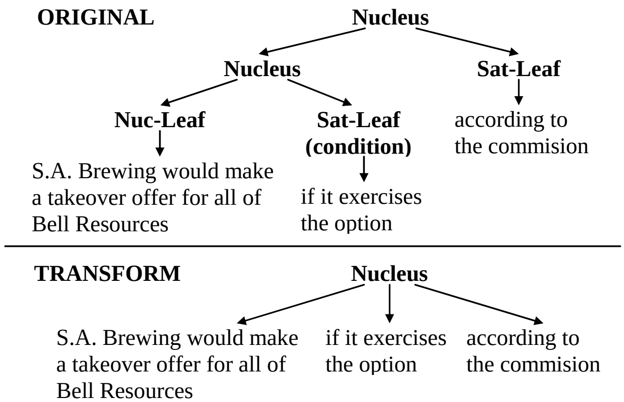


Figure 4: SAT transform example

The intuition with the second transform (SAT) is that satellite leaves that modify a nucleus subtree are very closely related to the content of the nucleus leaf of that subtree, as well as the satellite leaf of that subtree. By merging them, the tree is flattened, and pronouns within the original satellite leaf can refer to clauses in the subtree since they are now at the same depth. Figure 4 provides an illustration of the satellite transformation on (1). The side-effect of this transformation is that the RST tree is no longer binary. Finally, a hybrid transform (SENT-SAT) involves using both of the transforms on the corpus to flatten the tree even more.

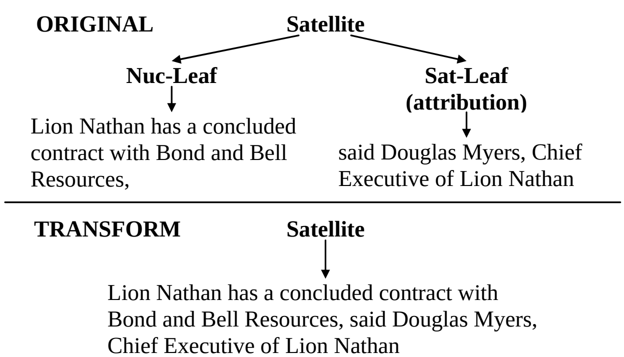


Figure 5: ATT transform example

Another transformation, ATT, is used to counter the attribution relation problem in Veins by simply merging leaves that stand in attribution relations. So if a subtree has two leaves in an attribution relation, it is replaced by a leaf with the text of the two original leaves merged. This process is similar to SENT. See Figure 5 for an example.

Both algorithms were run over the original RST corpus, ATT (attribution merge) and SAT (satellite merge) transformation of our original corpus, SENT (sentence merge) and SENT-SAT hybrid (see Table 4). The (S) version means that the LRC intrasentential search was used over the entire sentence, not just the clause that the pronoun occupies (C). Intersententially, the clause is the minimal discourse unit for both (S) and (C) methods. The (*) signals that the algorithm does not search the previous clause as a default.

Because the SENT transformations created unbalanced RST trees, the Veins algorithm could not be tested with that transform. The results show that the Grosz and Sidner algorithm fares best over the SENT and SENT-SAT transforms using the last-seen metric. However, it still performs the same as our baseline, so no improvement was seen by the transforms. It should be noted that the ATT transform did improve the Veins algorithm as suggested by its authors.

Transform	Veins (S)	Veins (C)	GS (S*)	GS (S)	GS (C)
Original	78.9	76.7	72.3	78.9	73.6
ATT	79.3	78.2	73.7	79.3	76.3
SAT	78.9	76.4	73.6	79.1	73.9
SENT	N/A	N/A	78.5	80.8	N/A
SENT-SAT	N/A	N/A	79.7	80.8	N/A

* signals that the algorithm does not search the previous clause as a default.

Table 4: Transform Results (% correct)

Without the attribution transform, the Veins Algorithm (S) gets only 6 of the 10 long-distance pronouns resolved correctly. The GS algorithms do about as well without segmentation. With the transformations, all the algorithms resolve all 10 cases correctly. However, it should be noted that the original LRC algorithm also resolves all correctly. This success rate is due to the fact that 9 of the 10 pronouns are either “he” or “him” and there are no other candidates with masculine gender in the discourse up to that point. So a simple search through a history-list would resolve these correctly. The other long-distance pronoun is a plural (“their”) and again there are no competing antecedents.

We also found that the Veins Algorithm was slightly more efficient than its competitors in that it considered less entities in its search. However the savings were marginal and are more than offset by the loss in performance.

5 Discussion

Discourse decomposition can be evaluated in two ways: intrasentential breakdown (clausal level) and intersentential breakdown (discourse level). In the intrasentential case, all the algorithms performed better when using the (S) method, that is, when the intrasentential search called for searching the sentence the pronoun is in, as opposed to just the clause the pronoun is in. This indicates that ordering clauses by their depth within the sentence or by the Veins information does not improve intrasentential performance, and thus one is better off searching based on grammatical function than incorporating clausal information.

One can evaluate the intersentential decomposition by testing whether the pronouns with long-distance antecedents are resolved correctly. Determining global discourse structure involves finding the middle ground between strict segmentation (using the exact RST tree) and under-segmenting. Too strict a segmentation means that antecedents can be deemed incorrectly inaccessible; very little segmentation means that too many competing antecedents become available since referents are not deemed inaccessible. In our corpus, evaluating intersentential decomposition is difficult because all of the long-distance pronouns have no competing antecedents, so no discourse structure is required to rule out competitors. Therefore it is hard to draw concrete conclusions from the fact GS on the SENT and SENT-SAT transforms performs the same as LRC algorithm. However, it is promising that this metric does get all of them right, at least it is not overly restrictive. The only way to check if the method under-segments or is a good model is by testing it on a corpus that has long-distance pronouns with competing potential referents. Currently, we are annotating a corpus of dialogs for coreference and rhetorical structure to test this method. It should also be noted that even if an intersentential decomposition method performs the same as knowledge-poor method, it has the advantage of at least decreasing the search space for each pronoun.

Finally, we developed an algorithm for VT that constrains the initial search for a referent. If one is not found, LRC is used as a default. As suggested by the VT authors, we merged clauses in attribution relations, and this improved performance slightly, but not enough to better 80.8%. VT run on the SAT transform offered no performance enhancement since the theory already makes the nucleus subtrees accessible to satellite leaves.

In conclusion, this study evaluates the theory that clausal segmentation should aid in pronoun resolution by testing two algorithms based on two leading theories of discourse segmentation. Both approaches have the promise of improving pronoun resolution by 1. making search more efficient by blocking utterances or classes from consideration, thus speeding up the search for an antecedent and 2. making search more successful by blocking competing antecedents. We use resolution accuracy for all pronouns and accuracy over long-distance pronominalizations as metrics of success. Our results indicate that basic metrics incorporating discourse structure does not improve performance, and in most cases can actually hurt performance. However, due to the composition of long-distance pronouns in the corpus, it is necessary to retest the GS algorithm on the SENT and SAT transforms before drawing a definitive conclusion on the theory.

6 Future Work

There are two main experiments to try in future research. One is to test whether merging clauses in an attribution relation would improve Veins performance since the algorithm does appear promising. This seems promising given how a majority of its errors stem from this phenomenon. The second is to implement the proposals made by Poesio and DiEugenio (2001) for merging Grosz and Sidner's discourse theory with RST.

The LRC error analysis in Section 3.4 shows that most gains in overall performance will come from resolving pronouns in complex, multi-clausal sentences correctly. In many cases, determining the coherence relations as Kehler suggests (2002) (such as detecting parallelism between sentences or within sentences) could improve interpretation. In addition, many errors stem from competing antecedents in which incorporating knowledge of the verbs and the entities discussed would of course prove invaluable.

The work here has focused on making one large blanket method for improving performance. This may not be the best way to go when improving on the baseline. For example, Suri *et al.* (1999) and Tetreault (2001) showed that isolating different errors and developing algorithms to treat them specifically resulted in improvements. Tetreault's (2001) improvements to treating preposed phrases and genitives accounted for an increase of 7% in the LRC algorithm. Another big boost could come from isolating intrasentential errors by differentiating between pronouns that refer locally versus to the subject of the sentence. This was a major source of error difference between LRC and lr-sort and the basic GS algorithms. The drawback of this type of work is that it does not lend itself to a general theory of pronoun resolution, unless the theory is that there are different modules in a system. Tailoring an algorithm too much could mean that it becomes too dependent on the corpus and does not generalize to other domains.

One way to skirt around the issue of domain independence is statistical evaluation. Strube *et al.* (2002) and Mueller *et al.* (2002) did such experiments to predict what the main factors are in pronoun resolution and weight them accordingly. On a corpus similar to the one presented here, they were able to correctly resolve 90% of the pronouns using syntax and light semantics. By including weighted information about clausal structure, relations, etc. with the usual syntax and semantics information, resolution algorithms could be improved. Another statistical avenue is genetic algorithms, though they have been used in this field with mixed success (Byron & Allen, 1999; Barbu, 2002).

Finally, our research here has assumed perfect knowledge of discourse structure. Ultimately, the goal is to be able to incrementally build discourse structure while processing a sentence. For this to occur, one has to take into account forms of referring expression, cue words, changes in tense, etc. There has been some work in this area such as (Hahn and Strube, 1997), thar developed an algorithm for building a discourse hierarchy incrementally from changes in theme and centered entities. Also, Marcu and Echihabi (2002) developed a statistical method for automatically detecting basic discourse relations such as elaboration and contrast.

Acknowledgements

I would like to thanks James Allen, Donna Byron, Massimo Poesio, Craige Roberts, and Michael Strube for their insight and comments. Partial support for the research reported in this paper was provided by the DARPA research grant no. F30602-98-2-0133 and the ONR grant no. N00014-01-1-1015 both to the University of Rochester.

References

Barbu, C. 2002. "Genetic Algorithms in Anaphora Resolution." *4th Discourse Anaphora and Anaphora Resolution Colloquium (DAARC 2002)*, 7-12.

Brennan, S., M. Friedman and C. Pollard 1987. "A Centering Approach to Pronouns." *Proc. of the 25th Annual Meeting of the ACL (ACL'87).* Stanford, CA. 155-162.

Byron, D. and J. Allen 1999. "Applying Genetic Algorithms to Pronoun Resolution." *Proc. of the 16th National Conference on Artificial Intelligence (AAAI-99)*, 957-967.

Byron, D. and J. Tetreault 1999. "A Flexible Architecture for Reference Resolution." *Proc. of the 9th Conference of the European Association for Computational Linguistics*, 229-232.

Carlson, L., D. Marcu and M. Okurowski 2001. "Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory." *Proc. of the 2nd SIGDIAL Workshop on Discourse and Dialog, Eurospeech 2001.*

Fox, B. 1987. *Discourse Structure and Anaphora*. Cambridge: Cambridge University Press.

Ge, N., J. Hale and E. Charniak 1998. "A Statistical Approach to Anaphora Resolution." *Proc. of the 6th Workshop on Very Large Corpora*.

Grosz, B., A. Joshi and S. Weinstein 1995. "Centering: A Framework for Modeling the Local Coherence of Discourse." *Computational Linguistics* 21(2).203-226.

---------- and C. Sidner 1986. "Attention, Intentions, and the Structure of Discourse." *Computational Linguistics* 12(3).175-204.

Hahn, U. and M. Strube 1997. "Centering Segmentation: Scaling up the Centering Model to Global Discourse Structure." *Proc of the 35th ACL*. 104-111.

Hitzeman, J. and M. Poesio 1998. "Long Distance Pronominalization and Global Focus." *Proc. of the 36th ACL*.

Hobbs, J. 1977. “Resolving Pronoun References.” *Lingua*. 44:3.311-338.

Ide, N. and D. Cristea 1998. “Veins Theory: An Approach to Global Cohesion and Coherence.” *Proc. of COLING*.

---------- and ---------- 2000. “A Hierarchical Account of Referential Accessibility.” *Proc. of the 38th Annual Meeting of the ACL.*

Kameyama, M. 1998. “Intrasentential Centering: A Case Study.” *Centering Theory in Discourse*. M. Walker, A. Joshi and E. Prince (eds.) Oxford University Press.

Kehler, A. 2002. *Coherence, Reference, and the Theory of Grammar*. Stanford, CA: CSLI.

Mann, W. and S. Thomson 1988. “Rhetorical Structure Theory: Towards a Functional Theory of Text Organization.” *TEXT*. 8:3.243-281.

Marcu, D. and A. Echihabi 2002. “An Unsupervised Approach to Recognizing Discourse Relations.” *Proc. of the 40th Annual Meeting of the ACL.*

Marcus, M., B. Santorini and M. Marcinkiewicz 1993. “Building a Large Annotated Corpus of English: The Penn Treebank.” *Computational Linguistics*, 12:2.313-330.

Mitkov, R. 2000. “Towards a More Consistent and Comprehensive Evaluation of Anaphora Resolution Algorithms and Systems.” *Proc. of the 2nd Discourse Anaphora and Anaphora Resolution Colloquium*. 96-107.

Moser, M. and J. Moore 1996. “Towards a Synthesis of Two Accounts of Discourse Structure.” *Computational Linguistics.*

Mueller, C., S. Rapp and M. Strube 2002. “Applying Co-training to Reference Resolution.” *Proc. of the ACL 2002*.

Poesio, M. and B. di Eugenio 2001. “Discourse Structure and Accessibility.” *ESSLI 2001 Workshop on Information Structure, Discourse Structure and Discourse Semantics.*

Prince, E. 1981. “Towards a Taxonomy of Given-New Information.” *Radical Pragmatics*. P. Cole (ed.) Academic Press.

Strube, M. 1998. “Never Look Back: An Alternative to Centering.” *ACL*, 1251-1257.

Strube, M., S. Rapp and C. Mueller. 2002. “The Influence of Minimum Edit Distance on Reference Resolution.” Proc of the *EMNLP 2000*. 312-319.

Suri, L., K. McCoy and J. DeCristofaro. 1999. “A Methodology for Extending Focusing Frameworks.” *Computational Linguistics*. 25:2.173-194.

Tetreault, J. 1999. “Analysis of Syntax-Based Pronoun Resolution Methods.” *Proc. of the 37th Annual Meeting of the ACL*. 602-605.

---------- 2001. “A Corpus-Based Evaluation of Centering and Pronoun Resolution.” *Computational Linguistics*. 27:4.

Walker, M. 2000. “Toward a Model of the Interaction of Centering with Global Discourse Structure.” *Verbatim.*

---------- 1989. “Evaluating Discourse Processing Algorithms.” *Proc. of the 27th Annual Meeting of the ACL*. 251-261.

A Lightweight Approach to Coreference Resolution for Named Entities in Text

Marin Dimitrov, Kalina Bontcheva, Hamish Cunningham and Diana Maynard
University of Sheffield

This paper presents a lightweight approach to pronoun resolution in the case when the antecedent is a named entity. It falls under the category of the so-called “knowledge poor” approaches that do not rely extensively on linguistic or domain knowledge. We provide a practical implementation of this approach as a component of the General Architecture for Text Engineering (GATE). The results of the evaluation show that even such shallow and inexpensive approaches provide acceptable performance for resolving the pronoun anaphora of named entities in texts.

1 Introduction

Anaphora resolution and the more general problem of coreference resolution are very important for several fields of Natural Language Processing such as Information Extraction, Machine Translation, Text Summarization and Question Answering Systems.

Because of its importance, the problems are addressed in various works and many approaches exist (for an overview see e.g., (Mitkov, 2002)). The approaches differ in the technology that they use for the implementation (symbolic, neural networks, machine learning, etc.), in the domain of the texts that they are tuned for, in their comprehensiveness (e.g. whether only pronominal anaphora is considered) and in the results achieved.

This work falls under the class of “knowledge poor” approaches to pronominal resolution, which are intended to provide inexpensive (in terms of the cost of development) and fast implementations that do not rely on complex linguistic knowledge, yet work with sufficient success rate for practical tasks (e.g., (Mitkov, 1998)).

Our approach follows the salience-based approach in (Kameyama, 1997), which perform resolution following the steps:

- inspecting the context for candidate antecedents that satisfy a set of consistency restrictions
- assigning salience values to each antecedent based on a set of rules and factors
- choosing the candidate with the best salience value

The approaches that influenced our implementation were focused on anaphora resolution of a certain set of pronouns in technical manuals (Lappin & Leass, 1994; Mitkov, 1998; Barbu & Mitkov, 2001). The goal of our work is resolution of pronominal anaphora in the case where the antecedent is a named entity – a person, organization, location, etc., with respect to the coreference task as defined by the Message Understanding Conference (MUC, see (Chinchor, 1997; Hirschman & Chinchor, 1997)). The implementation relies only on the part-of-speech information, named entity recognition and orthographic coreferences existing between the named entities. The lightweight approach we present achieves acceptable performance without using syntactic structure information, focus identification, centering theory methods as well as world-knowledge based approaches, which shows that even simple heuristic rules identified from analysis of the text can be sufficient for simple coreference resolution functionality. The texts that we used for the evaluation were newswire articles in English, part of the Automatic Content Extraction (ACE) programme training corpus (ACE, 2000).

We provide an implementation of the approach, available as a component integrated with the General Architecture for Text Engineering (GATE) – a Language Engineering framework and set of tools developed by the University of Sheffield (Cunningham *et al.*, 2002).

2 Related Work

In (Lappin & Leass, 1994) a syntax-based approach for identifying the antecedent of 3rd person pronouns and lexical anaphors is presented. The algorithm relies on syntactic information for the text being processed and contains a set of factors assigning the salience value of each candidate antecedent. The implementation also contains a pleonastic *it*[1] identification component. The authors report 86% successfully identified antecedents in a corpus containing technical manuals. A modification of the algorithm that does not employ deep syntactic parsing is proposed in (Kennedy & Boguraev, 1996). The authors report 75.5% success in resolution on a corpus containing texts of different genres.

Baldwin (1997) presents a simple yet effective algorithm for pronominal resolution. The algorithm does not rely on heavy syntax parsing but instead employs a set of 6 rules that assign proper salience values to candidate antecedents. The authors report 92% precision and 64% recall on a corpus containing The Wall Street Journal articles.

Mitkov (1998) reports a knowledge poor approach to pronominal resolution. The algorithm does not employ syntactic information but relies on a set of indicators (rules) such as definiteness, heading, collocation, referential distance, term preference, etc. The indicators assign salience values to the antecedents. The author reports success rate of 89.7% on a corpus of technical manuals.

---

[1] Pleonastic pronoun is the case when a pronoun (usually *it*) does not refer to any particular antecedent.

A comparative evaluation of the algorithms of (Kennedy & Boguraev, 1996), (Baldwin, 1997) and (Mitkov, 1998) is performed in (Barbu & Mitkov, 2001) over the same data set. The study reports success rates that are lower but more reliable than the original ones.

3 *Corpus Analysis*

We used the ACE test corpora, which are divided into three different types:

- **broadcast news programs** (BNEWS), ground truth versions of texts generated by speech recognition (ASR) systems, from news programs of ABC News, CNN, VOA and PRI (approx. 60,000)
- **newspaper** (NPAPER), ground truth versions of texts generated by optical-character recognition (OCR), mostly from "The Washington Post" articles. Contains 61,000+ words
- **newswire** (NWIRE). Contains 66,000+ words

We analysed these texts in order to have a better understanding of the specific issues related to each type of corpus. First we made an analysis of the pronoun distribution in the texts, and later an analysis of pleonastic *it* occurrences was performed. Not all pronouns were included in the analysis, only the following categories:

- **personal pronouns** – *I, me, you, he, she, it, we, they, me, him, her, us, them*
- **possessive adjectives** – *my, your, her, his, its, our, their*
- **possessive pronouns** – *mine, yours, hers, his, its, ours, theirs*
- **reflexive pronouns** – *myself, yourself, herself, himself, itself, ourselves, yourselves, themselves, oneself*

There were cases in which a pronoun can be classified in more than one category. For example *his* and *its* may be possessive pronouns or possessive adjectives, however we rely on GATE's part-of-speech (POS) tagger to correctly assign the necessary category.

3.1 Total Pronouns

The percent of words that are pronouns reported in (Barbu & Mitkov, 2001) is 1.5% (422 pronouns out of 28,272 words). The average ratio we observed was almost three times higher. This is probably due to the specific differences in the domain of the analysed texts. The corpus in (Barbu & Mitkov, 2001) consists of technical manuals where specific grammatical constructs and language are being used. The ACE corpus consists of news articles and interviews where the number of named entities and the pronouns used to refer to them is unsurprisingly much higher.

The percentage of pronouns is shown in Table 1:

source	words	pronouns	pronouns (% of words)
npaper	61319	2264	3.7
bnews	60316	3392	5.6
nwire	66331	2253	3.4
TOTAL	187966	7909	4.2

Table 1. Number of pronouns and number of words in the ACE corpus

It is worth pointing out that the NWIRE and NPAPER part of the ACE corpus contain similar percentage of pronouns, while the percentage of pronouns in BNEWS is much higher. This is due to the fact that BNEWS contains mostly quoted speech dialogues, where pronouns are used more often than the names of the entities.

3.2 Distribution of Pronouns by Type

The relative distribution of pronouns by type is similar to the one reported in (Barbu & Mitkov, 2001). Again the most significant share is that of the personal pronouns, followed by the possessive pronouns while the share of reflexive pronouns is insignificant. This is shown in Table 2.

source	pronouns	pers.	pers. %	poss.	poss. %	refl.	refl. %
npaper	2264	1593	70.4	627	27.7	42	1.9
bnews	3392	2862	84.4	491	14.5	39	1.1
nwire	2253	1629	72.3	586	26.0	38	1.7
TOTAL	7909	6084	76.9	1704	21.5	119	1.5

Table 2. Distribution of personal, possessive and reflexive pronouns in the ACE corpus

The similarity between NPAPER and NWIRE corpora is observed again. The percentages for BNEWS are quite different from the rest and are closer to the ones reported in (Barbu & Mitkov, 2001).

Table 3 below shows the relative importance of the 10 most frequently observed pronouns in each corpus. There exists a significant difference in the distribution of certain pronouns in the different corpora. For example *I*, *you* and *we*, which are expected to indicate quoted speech presence, constitute around 13% and 19% of the pronouns in NPAPER and NWIRE respectively, while the percentage for BNEWS is almost twice as high – 32.6%.

Another fact of interest that is not shown in the table is the relative unimportance of possessive pronouns (*mine*, *yours*, etc.) in the text. There were only two such pronouns observed in the NPAPER corpus, constituting 0.1% of the pronouns, and there were no such pronouns in the BNEWS and NWIRE corpora. This implies that the coreference resolution algorithm may effectively ignore such pronouns because their successful handling will not significantly influence the overall performance.

npaper		bnews		nwire	
pronoun	%	pronoun	%	pronoun	%
he	18.3	it	18.9	it	18.9
it	16.8	I	11.6	he	16.5
his	12.0	you	11.6	his	11.0
its	8.6	he	10.5	I	8.2
they	8.0	they	10.1	they	8.1
I	6.5	we	9.4	its	6.7
we	6.4	his	6.1	we	6.7
she	4.8	its	3.1	you	5.0
her$	3.3	she	2.6	she	2.6
them	2.7	her$	2.0	her$	2.2

Table 3. Relative importance of the 10 most often observed pronouns in different parts of the ACE corpus, her$ *is the possessive adjective for* she*, while* her *is the corresponding object personal pronoun.*

The same holds for reflexive pronouns. They constitute about 1.5% of the pronouns in the three corpora, so their effective resolution is unlikely to contribute sufficiently to good performance.

3.3 Pleonastic It Statistics

Pleonastic pronouns are not considered anaphoric (since they don't have an antecedent) but identifying such occurrences is important so that the coreference resolution system will not try to resolve them. We analysed the three corpora for *pleonastic it* constructs. A full analysis for all non-anaphoric pronouns was out of the scope of this work. The results from the *pleonastic it* analysis are expected to contain some imprecision, since the texts were analysed only by one person. The percentage of *pleonastic it* occurrences we observed was very low compared with the percentages reported by other researchers, e.g. (Lappin & Leass, 1994) (7.7%). This difference is most likely a consequence of the different domain of the analysed texts – technical manuals vs. news articles and interviews. Table 4 contains the results from the pleonastic *it* analysis:

source	pronouns	*it*	pleonastic *it*	pleonastic *it* (% of pronouns)	pleonastic *it* (% of *it*)
npaper	2264	381	79	3.5	20.7
bnews	3392	642	105	3.1	16.4
nwire	2253	425	70	3.1	16.5
TOTAL	7909	1448	254	3.2	17.5

Table 4. Pleonastic it *occurrences as nominal value, percentage of all pronouns, percentage of the occurrences of* it

Note that the statistics for BNEWS and NWIRE are quite similar but they differ a lot from the ones for NPAPER. It is also worth pointing out that pleonastic *it* constitutes a large percent of the total number of occurrences of *it* so if pleonastic pronouns are ignored in the implementation of the resolution algorithm, the final results for the resolution of *it* are likely to be unsatisfactory. This observation is even more important if we consider that *it* constitutes about 19% of the pronouns in the three corpora.

4 Design of the Coreference Resolution Module

The analysis of the ACE corpora helped us clarify and prioritise the requirements for the implementation of the module.

The coreference module itself has modular structure – it consists of a main module and a set of submodules. This modular structure provides sufficient flexibility, so that the behaviour of the coreference module may be modified or tuned for specific tasks. Such specific tasks may require that the order in which submodules are executed be changed (unless there are interdependencies between them). For certain tasks it may not be feasible to load and execute some modules at all if they are unlikely to contribute much for the final result. This is the case with technical manuals, which do not usually contain quoted speech fragments, so the submodule identifying such fragments in the text will not be appropriate.

5 Architecture

At present the coreference module consists of three submodules:

- pronominal resolution submodule
- quoted text submodule
- pleonastic *it* submodule

The coreference module operates within the context of the GATE system, which provides an Information Extraction (IE) pipeline for named entity recognition (for an overview see (Cunningham, 1999)). The module depends on the results (provided as annotations) of various tasks performed by the IE pipeline such as tokenization, part-of-speech information, sentence boundary detection and Named Entity recognition.

The following sections present overviews of the implemented modules.

5.1 Quoted Speech Submodule

The quoted speech submodule identifies quoted fragments in the text being analysed. The identified fragments are used by the pronominal coreference submodule for the proper resolution of pronouns such as *I*, *me*, *my*, etc. that appear in quoted speech fragments.

The submodule does not handle perfectly all the possible constructs of quoted fragments, which degrades the performance of the pronominal submodule later. The main reason for that is the lack of correctly balanced quotation marks in the ACE corpora.

5.2 Pleonastic It *Submodule*

The pleonastic *it* submodule matches pleonastic occurrences of *it*. As we already discussed above, the number of pleonastic *it* occurrences observed was significantly smaller than the numbers reported by other researchers. Yet the relative share of pleonastic *it*, as a percentage of all the occurrences of *it*, makes identification of the former useful.

Previous work, such as (Lappin & Leass, 1994), reports that certain patterns about pleonastic *it* can be identified. The latter contains a set of rules for recognizing the pleonastic constructs most often observed in texts. Unfortunately we discovered that these patterns would not be sufficient for all the cases observed. The problems we identified are:

- Often a synonym or antonym of a modal adjective or a synonym of a cognitive verb appears in the construct.
- The patterns are not flexible enough and miss even small variations of the defined constructs (e.g. *It is likely that…* will be matched but *It is also likely that…* will be missed).
- There are constructs in the ACE corpus that will not be matched by these patterns (e.g. *It is mandatory that…* will be matched but similar statement like *It becomes mandatory that…* will be missed).

The first problem can be resolved by adding the proper synonyms and antonyms from WordNet (Miller *et al.*, 1990) or another lexical resource. The other problems have to be resolved by extending the base patterns.

With the help of WordNet and according to our observations of the ACE corpus, we extended the set of modal adjectives from (Lappin & Leass, 1994) to the following set (the original set of 15 adjectives appears in **bold**):

*acceptable, adequate, **advisable**, appropriate, bad, better, **certain**, clear, common, **convenient**, decent, **desirable**, **difficult**, doubtful, **easy**, **economical**, efficient, enough, essential, expected, fair, feasible, **good**, great, hard, **important**, illegal, imperative, implausible, important, impossible, impractical, improbable, inadequate, inadvisable, inappropriate, inconvenient, inefficient, inessential, insufficient, invalid, **legal**, **likely**, mandatory, **necessary**, obligatory, painless, plausible, **possible**, practical, probable, rare, reasonable, recommended, safe, sensible, **sufficient**, suggested, suitable, sure, tough, typical, unacceptable, unadvisable, unclear, undesirable, unexpected, unfair, unimportant, unlikely, unnecessary, unreasonable, unsafe, unsuitable, unsure, unusual, unwise, unworthy, **useful**, useless, usual, valid, wise, wonderful, worthy, wrong*

Of course such extension makes the pleonastic *it* module more computationally demanding and slows down the overall performance, so a balance should be made between extending the rules and processing with acceptable performance.

The cognitive verbs were extended to the set:

*accept, advise, **anticipate**, **assume**, **believe**, consider, demand, deny, estimate, **expect**, foresee, intend, **know**, predict, promise, propose, realize, **recommend**, recognize, report, require, suggest, **think***

In order to improve the flexibility of the patterns from (Lappin & Leass, 1994) and the number of pleonastic occurrences they identify we have provided extended versions, based on the data analysis performed over the ACE corpus. The extended patterns are:

1. It be (adverb) modaladj (conj01) S
2. It be (adverb) modaladj (for NP) to VP
3. It is (adverb) cogv-ed that S
4. It (adverb) verb01 (conj02 | to) S
5. NP verb02 it (adverb) modaladj (conj01 NP) to VP

We have dropped patterns 6 (*It is time to **VP***) and 7 (*It is thanks to **NP** that **S***) from the original paper, because they constituted less than one percent of the observed pleonastic *it* occurrences.

In the patterns above we have:

be = {*be*, *become*, *remain*}

adverb = {*highly*, *very*, *still*, *increasingly*, *certainly*, *absolutely*, *especially*, *entirely*, *simply*, *particularly*, *quite*, *also*, *yet*, *even*, *more*, *most*, *often*, *rarely*}

modaladj is the set of modal adjectives already discussed

conj01 = {*for*, *that*, *is*, *whether*, *when*}

conj02 = {*that*, *if*, *as*, *like*}

cogv-ed is the passive participle of the cognitive verbs defined above

verb01 = {*seem*, *appear*, *look*, *mean*, *happen*, *sound*}

verb02 = {*find*, *make*, *consider*}

The implementation of the pattern extends the rules so that:

- Different forms of the sets of verbs **be**, **verb01** and **verb02** are recognized (base, present 3rd person, present non-3rd person, past participle)
- Question forms are matched
- Modal verbs used with the above sets are matched
- Negation is matched

We identified one more pattern that was observed often in the ACE corpus, but we did not provide implementation for it because the pattern was not generic enough and depends on too many specific expressions. The pattern looks like

6. It be/take time-expr before/since S

where **time-expr** represents time expressions such as *two weeks*, *today*, *one month*, *a while*, *longer*, etc.

Table 5 lists the distribution of the pleonasms from each type observed in the ACE corpus together with the percentage of the occurrences correctly identified.

pattern	occurrences	% of *pleonastic it*	identified (%)
1	35	13.9	72.0
2	65	25.8	72.0
3	3	1.2	33.3
4	18	7.1	77.8
5	11	4.4	72.7
6	16	6.3	–
unclassified	104	41.3	–
total	252	100.0	37.7

*Table 5. Pleonastic-*it *statistics – the number of pleonastic occurrences per pattern, percentage of the occurrences per pattern, percentage of occurrences correctly identified*

Note that patterns 1 and 2 are observed most often and the percentage of pleonastic *it* constructs that were not matched by any pattern is very high – more than 40%. The precision (number of occurrences matched / all occurrences of this type) of the specific rules is relatively good and with the exception of one rule it is around 72% but the high number of unclassified occurrences degrades the overall performance.

5.3 Pronominal Resolution Submodule

The main functionality of the coreference resolution module is in the pronominal resolution submodule. This submodule uses the result from the execution of the quoted speech and pleonastic *it* submodules.

The module works according to the following algorithm:

1. For each pronoun:
   - Inspect the appropriate context for all candidate antecedents for this kind of pronoun;
   - Choose the best antecedent (if any).
2. Create the coreference chains from the individual anaphor/antecedent (this step is performed from the main coreference module).

Pronoun resolution (step 1) works as follows:

1. If the pronoun is *it* then a check is performed if this is a pleonastic occurrence and if so then no further attempt for resolution is made.
2. The proper context is determined. The context size is expressed in the number of sentences it will contain. The context always includes the current sentence (the one containing the pronoun) and one or more preceding sentences.
3. Depending on the type of pronoun a set of candidate antecedents is proposed. The candidate set includes the named entities that are compatible with this pronoun. For example, if the current pronoun is *she* then only the Person

annotations (as provided by GATE) with 'gender' feature equal to 'female' or 'unknown' will be considered as candidates. From all candidates one is chosen according to evaluation criteria specific for the pronoun.

The corpus analysis that we performed showed that certain pronouns are related to their respective antecedents in a similar manner. We divided the pronouns into three groups based on this similarity – group I includes the pronouns *she*, *her*, *her$*, *he*, *him*, *his*, *herself* and *himself*, group II includes *it*, *its* and *itself*, and group III includes *I*, *me*, *my* and *myself*.

The resolution of group I pronouns is based on the observation that these pronouns typically refer to persons. The characteristics of the resolution process are:

- Context that will be inspected is not very big – cases where the antecedent is found more than 3 sentences further back than the anaphor are very rare.
- The recency factor is heavily used – the candidate antecedents that appear closer to the anaphor in the text are scored better.
- Anaphora has higher priority than cataphora.[2] If there is an anaphoric candidate and a cataphoric one then the anaphoric one is preferred, even if the recency factor scores the cataphoric candidate better.

The resolution process for group I pronouns performs the following steps:

1. Inspect the context of the anaphor for candidate antecedents. Only annotations of type Person are considered as candidate antecedents – cases where *she*/*her* refers to inanimate entity (a ship for example) are not handled.
2. For each candidate perform a gender compatibility check – only candidates having 'gender' feature equal to 'unknown' or compatible with the pronoun are considered for further evaluation.
3. Compare each candidate with the best candidate so far:
    - If the two candidates are anaphoric for the pronoun then choose the one that appears closer.
    - The same holds for the case where the two candidates are cataphoric relative to the pronoun.
    - If one is anaphoric and the other is cataphoric then choose the former, even if the latter appears closer to the pronoun.

The pronouns of group II (*it*, *its* and *itself*) also share many common characteristics. The resolution process contains certain differences from the one for the previous set of pronouns.

Successful resolution for *it, its* and *itself* is more difficult because of the following factors:

- There is no gender compatibility restriction. In the case where there are several candidates in the context, the gender compatibility restriction is very useful for rejecting some of the candidates. When no such restriction exists, and with the

---

[2] Cataphora is the case when the pronoun precedes its antecedent in the text.

lack of any syntactic or ontological information about the entities in the context, the recency factor plays the major role for choosing the best antecedent.

- The number of nominal antecedents (i.e. entities that are referred to by something other than a name) is much higher compared with the number of such antecedents for *she*, *he*, etc. In this case trying to find antecedents only amongst named entities degrades the precision a lot.

We performed analysis of the occurrences of *it*, *its* and *itself* in the ACE corpus in order to determine the usefulness of the recency factor as the only factor applied for choosing the best antecedent. Our analysis showed that:

- In 52% of the cases, the most recent named entity (provided by GATE) of type Organization and Location was the correct antecedent.
- In 15% of the cases, the antecedent was a named entity that was not the most recent related to the anaphor.
- In 33% of the cases, the antecedent was nominal where the resolution will fail, as their resolution was left to be considered in future work.

The analysis shows that the recency factor all by itself offers some means of correct pronominal resolution. Further, we identified that half of the cases (7.3%) where the antecedent was not the most recent named entity were appositional. For example:

> ***Yamaichi Securities Co*$_1$**, *once* ***Japan*$_2$***'s largest securities house, officially closed* ***its*$_1$** *last offices today after authorities revealed the severity of its losses.*

In this example, if the best antecedent is chosen on the basis of recency then *its* will be incorrectly matched to *Japan*. If apposition was identified, then the most proper choice would have been the named entity to which the apposition itself refers (in this case *Yamaichi Securities Co*).

The resolution steps are similar to the ones for group I etc. with the following differences:

- Only entities of type Location and Organization are considered as candidate antecedents.
- Only recency is considered for choosing the best antecedent.
- Named entities that are cataphoric to the pronoun are not considered as candidate antecedents.

The resolution of the pronouns of the 3rd group – *I*, *me*, *my* and *myself* – is dependent on the results generated by the quoted speech submodule. One important difference from the resolution process of other pronouns is that the context is not measured in number of sentences but depends solely on the quote span. Another difference is that the context is not contiguous – the quoted fragment itself is excluded from the context because it is unlikely that an antecedent for *I*, *me*, etc. appears there.

The context itself consists of:

- The part of the sentence where the quoted fragment originates, which is not contained in the quote – i.e. the text prior to the quote.
- The part of the sentence where the quoted fragment ends, which is not contained in the quote – i.e. the text following the quote.
- The part of the sentence preceding the sentence where the quote originates, which is not included in another quote.

For example, the context for the following example is underlined:

Doctor Frank Plummer$_1$, scientific director of CNML, made a statement at the conference. "We're puzzled as to why we are finding this virus in about 40% of the cases. What I$_1$ can conclude is that we have a new epidemic and whether it's the entire cause of SARS... I$_1$'m not sure," he$_1$ said.

The underlined text in the next example is context for the second quoted fragment. The context for the first one will include the first part of the underlined text and parts (or the whole) of the preceding sentence (not shown).

Sen. John McCain$_1$ said on CBS's Face the Nation, "I'm optimistic that we can get this done by this summer"" Noting that the White House has complained, he$_1$ said, "I$_1$ think we may be well-positioned."

Another difference with resolution of pronouns of the first set (*he*, *she*, *his*, *him*, etc.) is that candidate antecedents are considered to be not only named entities (of type Person) but also the pronouns *he* and *she.*

We identified several patterns that classify the relation between the pronouns *I, me, my, myself* and their antecedents. The subset of the corpus that was analysed consisted of 40 documents containing 95 quoted fragments with 72 occurrences of the pronouns of interest. The patterns we identified for these occurrences are:

- The antecedent is the closest named entity in the text following the quoted fragment. This pattern is observed in 52% of the cases. An example is: *"**I$_1$** did not urge anyone to say anything that was untrue," **Clinton$_1$** told Lehrer.*
- The antecedent is found in the sentence preceding the sentence where the quoted fragment originates. If the preceding sentence also contains a quote then the antecedent is usually the named entity (or pronoun) that is closest to the end of the quote. This pattern was observed in 29% of the cases. An example is: *"**I$_1$** did not urge anyone to say anything that was untrue," **Clinton$_1$** told Lehrer. "That's **my$_1$** statement to you"*
- The antecedent is the closest named entity preceding the quote in the sentence where the quote originates. This pattern accounts for less than 3% of the cases. An example is: *U.S. officials said there was confusion about whether China would fulfil the contracts, but **Cohen$_1$** declared: "**I$_1$** believe we have assurances that such sales will not continue."*
- The antecedent is either nominal (13%) or a named entity in position where the patterns above will not identify it correctly (3%). These cases will not be

handled properly by the algorithm, which will induce degradation in recall and possibly degradation in precision (if the wrong antecedent is proposed).

It is worth noting that contrary to other pronouns, the antecedent for *I, me, my* and *myself* is most often cataphoric or, if anaphoric, it is not in the same sentence as the quoted fragment.

The resolution algorithm consists of the following steps:

1. Locate the quoted fragment description that contains the pronoun. If the pronoun is not contained in any fragment then return without proposing an antecedent.
2. Inspect the context of the quoted fragment (as defined above) for candidate antecedents.
3. Try to locate a candidate in the text following the quoted fragment (first pattern). If more than one candidate is present, choose the one closest to the end of the quote. If a candidate is found then propose it as an antecedent.
4. Try to locate a candidate in the text preceding the quoted fragment (third pattern). Choose the one closest to the beginning of the quote. If found then choose as an antecedent.
5. Try to locate antecedents in the unquoted part of the sentence preceding the sentence where the quote starts (second pattern). Give preference to the one closest to the end of the quote (if any) in the preceding sentence or closest to the sentence beginning.

6 Evaluation

We annotated manually a small subset of the ACE corpora in order to evaluate precision, recall and F-measure for the implementation (for an overview of these metrics see e.g. (Chinchor, 1992)). The subset consists of 21 randomly selected documents (7 from each corpus) containing 352 pronouns. The sample corpus represents 5% of the documents in the ACE corpus and contains 4.5% of the pronouns. No pronouns were excluded from the evaluation. Occurrences of the pronouns that the implementation does not handle yet degrade the recall. Nominal antecedents degrade the precision. The precision achieved on the evaluation corpus was 66% and the recall was 46%. These numbers are comparable with the performance of other knowledge-poor coreference resolution approaches (Barbu & Mitkov, 2001; Mitkov, 1998). The table below contains the results for each individual group of pronouns.

The results show that the resolution of pronouns such as *he, she, her*, etc. is relatively successful even with such simple heuristic patterns used and without incorporating any syntax or centering information. The precision is degraded by the ratio of nominal antecedents. The algorithm will also benefit from some syntactic information indicating the subject of the sentence, because the results show that the recency factor and the gender agreement are not sufficient.

pronoun group	precision (%)	recall (%)	f-measure (%)
I	79.3	77.2	78.2
II	43.5	51.7	47.2
III	77.8	62.2	70.0

Table 6. Precision, Recall and F-measure for the three groups of pronouns (1st group includes he*,* she*, etc., the 2nd group includes* it*,* its *and* itself*, the 3rd one includes* I*,* me*,* myself *and* my*)*

The resolution of pronouns such as *it*, *itself* and *its* is less successful. Apart from the nominal antecedent which have even greater impact for this group, additional degradation is induced from the low performance of the pleonastic *it* module, which although using rules that cover many more cases than the ones in (Lappin & Leass, 1994) still identifies only 38% of the pleonastic occurrences. It is worth noting that the pleonastic *it* module has very high precision and very low recall, so a further extension of its patterns will improve the recall and will have positive impact on the resolution of *it*.

The recall errors for resolution of *I*, *me*, etc. are mainly due to errors made by the quoted speech submodule. Additionally, the performance is negatively impacted by the specifics of the BNEWS corpus, where the quoted fragments are not marked in the text, and as a result no attempt for resolution of the pronouns of the 3rd group will be made on this corpus.

Finally, if we measure the performance of the coreference module, independently from the Named Entity Recogniser of GATE, i.e. in a “best case scenario” (against the same corpus but with manually annotated named entities) where all the named entities are correctly identified, then precision goes up to 73% and recall up to 53% for all pronouns, with the biggest improvement for group 3 where precision goes up to 86% and recall is 76%.

7 Conclusion

The novel aspects of this work are the detailed analysis of pronoun distributions and patterns in domain corpora on the basis of which we developed shallow techniques for anaphora resolution. This lightweight approach achieves acceptable performance without using any syntactic information or centering theory methods, which shows that even simple heuristic rules identified from analysis of the text can be sufficient for baseline coreference functionality.

As more named entity recognisers are becoming available for new languages (e.g., Romanian, Hindi – for details, see (Maynard & Cunningham, 2003)), it will be interesting to experiment with porting this approach to languages other than English. Some of the modules (e.g. quoted speech identification) are directly reusable, while others (e.g. pleonastic *it* patterns) are either inapplicable or need an adaptation to the new language following a detailed corpus study similar to the one carried out here.

Unfortunately further improvement in the precision and recall just by incorporating lightweight techniques is unlikely to be achieved. That is why we intend to incrementally extend the basic functionality with new features. In future work, we will address apposition identification, extending the set of handled pronouns, and integrate it with the new GATE module for resolving nominal anaphora.

References

ACE 2000. "Entity Detection and Tracking Phase 1 – ACE Pilot Study Task Definition". Available online from the ACE site: http://www.itl.nist.gov/iaui/894.01/tests/ace/index.htm.

Baldwin, B. 1997. "CogNIAC: High Precision Coreference with Limited Knowledge and Linguistic Resources". *Proc. of ACL/EACL Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*, 38-45.

Barbu, C. and R. Mitkov 2001. "Evaluation Tool for Rule-based Anaphora Resolution Methods". *Proc. of ACL 2001*, 34-41.

Kennedy, C. and B. Boguraev 1996. "Anaphora for Everyone: Pronominal Anaphora Resolution without a Parser". *Proc. of the 16th Int. Conference on Computational Linguistics* (COLING'96), 113-118.

Chinchor, N. 1992. "MUC-4 Evaluation Metrics". *Proc. of the 4th Message Understanding Conference*, 22-29.

---------- 1997. "Overview of MUC-7". In *Message Understanding Conference Proc.* http://www.itl.nist.gov/iaui/894.02/related_projects/muc/proceedings/muc_7_proceedings/overview.html

Cunningham, H. 1999. "Information Extraction – A User Guide". Research memo CS-99-07, University of Sheffield.

----------, D. Maynard, K. Bontcheva and V. Tablan 2002. "A Framework and Graphical Development Environment for Robust NLP Tools and Applications". *Proc. of the 40th Anniversary Meeting of the ACL*, 168-175.

Hirschman, L. and N. Chinchor 1997. "MUC-7 Coreference Task Definition". In *Message Understanding Conference Proc.* http://www.itl.nist.gov/iad/894.02/related_projects/muc/proceedings/muc_7_toc.html.

Kameyama, Megumi 1997. "Recognizing Referential Links: An Information Extraction Perspective". *Proc. of ACL/EACL Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*, 46-53.

Lappin, S. and H. Leass 1994. "An Algorithm for Pronominal Anaphora Resolution". *Computational Linguistics* 20.535-561.

Maynard, D. and H. Cunningham 2003. "Multilingual Adaptations of a Reusable Information Extraction Tool". *Proc. of the 10th Conference of the European Chapter of the ACL* (EACL 2003).

Miller, G., R. Beckwith, C. Fellbaum, D. Gross and K. J. Miller 1990. "Introduction to Wordnet: An On-line Lexical Database". *Int. Journal of Lexicography* 3.235-244.

Mitkov, R. 1998. "Robust Pronoun Resolution with Limited Knowledge". *Proc. of the 18th Int. Conference on Computational Linguistics* (COLING'98)/ACL'98 Conference, 869-875.

---------- 2002. *Anaphora Resolution*. London: Longman.

A Unified Treatment of Spanish *se*

Randy Sharp
Department of Linguistics, University of British Columbia

This paper investigates the referential properties of the Spanish clitic *se* and proposes a way to formalize these properties within a unification-based formalism, and to apply it to the problem of Spanish-English machine translation. It is argued that all instances of *se* can be categorized as reflexive, characterized as either *personal* or *nonpersonal*. Differences in the use and interpretation of *se* arise from the way that *se* discharges an argument from a verb's argument structure. Personal *se*, which occurs in referential, obligatory and aspectual reflexives, always discharges an internal argument, whereas nonpersonal *se*, which occurs in passive, impersonal, inchoative and middle reflexives, always discharges the external argument. In both cases, the clitic adjoins to the I(nflection) node in the structural representation and is bound by the subject, thereby giving rise to the reflexive relation. In the nonpersonal reflexive cases, in which the external argument is effectively suppressed, the subject is realized as expletive *pro*, which binds the reflexive clitic and thus maintains the reflexive relation. The translation of sentences containing *se* depends on how the verb's argument structure is affected, and on which argument has been discharged.

1 Introduction

Reflexives[*] are not generally considered a hard problem in anaphora resolution. Their antecedents are not only intrasentential, thus requiring no cross-sentential inspection whatsoever, they are even within the same clause as the reflexive. Hence, once an element is identified as a reflexive, its resolution should follow virtually deterministically.[1] For this reason, the treatment of reflexives has received little attention in the anaphora resolution literature.

The problem facing an anaphora resolution model, however, lies in actually identifying the reflexive. In Spanish, the clitic *se* is easily recognized as reflexive in sentences of the type in (1):

(1) a. Juan *se* vio. "Juan saw himself."

---

[*] Thanks to Rose-Marie Déchaine and Derek Carr of UBC for valuable comments on an earlier version of this paper. This research was partially supported by Social Sciences and Humanities Research Council of Canada (SSHRC) grant #752.1999.2082.

[1] Notable exceptions include English "picture" reflexives (i), in which the antecedent, although contained locally within the same clause as the reflexive, is ambiguous between the subject and object, and reflexives like Chinese *ziji* (ii) (Cole *et al.*, 2001:xiv), which may be either locally or long-distance bound by its antecedent:

(i) John$_i$ showed Bill$_j$ pictures of himself$_{i/j}$

(ii) Zhangsan$_i$ renwei Lisi$_j$ zhidao Wangwu$_k$ xihuan ziji$_{i/j/k}$
"Zhangsan thinks Lisi knows Wangwu likes self"

b. Juan *se* pegó. “Juan hit himself.”

However, its status as reflexive in impersonal constructions (2) is less obvious:

(2) a. *Se* habla español. “Spanish is spoken.”
b. *Se* venden casas. “Houses for sale.”

To account for the difference, some researchers have postulated two (or more) lexical entries for *se* (Otero, 1966; Perlmutter, 1971; Zubizarreta, 1982), one reflexive, one nonreflexive, while others have attempted to consolidate the entries under a single morpheme (Manzini (1986) for Italian *si*, Wehrli (1986) for French *se*). Still others (e.g. Bruhn de Garavito (1999), Sanz (2000)) have posited two (or more) structural representations in which *se* can occur. In this paper, I show that a single lexical entry and a single structural representation are adequate. Furthermore, I show that all instances of *se* can indeed be analysed as reflexive.[2] By this is meant that *se* is always bound by (i.e. has as its antecedent) the local subject of the clause containing it. The differences in interpretation evident in (1) and (2) arise from the way in which the reflexive interacts with the lexical properties of the verb, in particular whether it discharges an internal or an external argument, in the sense of Williams (1981). When *se* discharges an internal argument, a personal reflexive, as in (1), results; when it discharges the external argument, a nonpersonal reflexive results, as in (2).

2 Personal Reflexives

Examples of the two types of reflexives are presented below. The personal reflexives are presented in this section, and nonpersonal reflexives in the next.

The defining characteristic of all personal reflexives is that the reflexive clitic always agrees in person and number with the subject of the clause, even if the subject is missing under *pro*-drop. That is, changing the person value of the subject in (1) results in a corresponding change to the clitic; alternatively, changing the clitic requires a corresponding change to the subject, as illustrated in the full person/number paradigm in (3):

(3) a. (Yo) *me* ví. “I saw myself.”
b. (Tú) *te* viste. “You saw yourself.”
c. (Él/ella) *se* vio. “He/she saw himself/herself.”

[2] I ignore here the “spurious *se*” (Perlmutter, 1971), which replaces the first of two adjacent 3rd person clitics, as in *Yo se lo di* “I gave it to him/her/them”, in which *se* has replaced the dative clitic *le* or *les*. This *se*, which is clearly nonreflexive given the 1st person singular subject, has a phonological origin distinct from that of reflexive *se*. I will assume that it is handled at the PF-syntax interface, where *se* is effectively converted to *le/les* on input, and *le/les* is converted to *se* on output in the context of two adjacent 3rd person clitics.

d. (Nosotros) *nos* vimos. "We saw ourselves."
e. (Vosotros) *os* vistéis. "You saw yourselves."
f. (Ellos) *se* vieron. "They saw themselves."

Three types of personal reflexives are identified below.

2.1 Referential Reflexive

The examples in (1) above illustrate typical referential reflexives. The subject and the clitic object have independent referential thematic roles (θ-roles), e.g. AGENT and THEME. When the two thematic constituents refer to the same entity, the object is reflexive, as in (3); when they refer to distinct entities, the object is nonreflexive, as in (4):

(4) a. Juan vio el accidente. "Juan saw the accident."
b. Juan *lo* vio. "Juan saw it/him/*himself."

2.2 Obligatory Reflexive

Certain verbs obligatorily take a reflexive object; examples in English include *perjure oneself*, *pride oneself*, and *behave oneself*[3] (Levin, 1993:107). In Spanish, many verbs are considered obligatory reflexives, such as *quejarse* "complain", *arrepentirse* "repent", *abstenerse* "abstain", *suicidarse* "commit suicide", and many others. Like English reflexive verbs, these Spanish verbs may not occur without a reflexive agreeing with the subject; for example:

(5) a. Yo *(*me*) quejé del ruido. "I complained about the noise."
b. Tu *(*te*) quejaste del ruido. "You complained about the noise."
c. Juan *(*se*) quejó del ruido. "Juan complained about the noise."

(As is customary, the notation *(se) indicates that only if *se* is present is the sentence grammatical; if it is absent, it is ungrammatical.)

Note that verbs like *lavar* "wash" and *afeitar* "shave" are often considered reflexive verbs, as typified by the sentences in (6):

(6) a. Juan *se* lavó. "Juan washed (himself)."
b. Juan *se* afeitó. "Juan shaved (himself)."

My claim is simply that these verbs are normal obligatorily *transitive* verbs, but not necessarily reflexive:

(7) a. Juan afeitó a su hijo. "Juan shaved his son."
b. Juan lo afeitó. "Juan shaved it/him/*himself."
c. *Juan afeitó. "Juan shaved ?? ."

---

[3] Note that the object of *behave* may be omitted, in which case it is an understood reflexive. When present, the object is obligatorily reflexive, explaining the ungrammaticality of (i)b:

(i) a. John behaved. = John behaved himself.
b. *John behaved his children.

In (7a), the object may be replaced by any relevant entity, rendering the sentence nonreflexive. Similarly, a nonreflexive clitic may be substituted for the object, as in (7b). The obligatoriness of the object is shown by the ungrammaticality of (7c), where neither a full object nor clitic is present. The English verbs *wash* and *shave*, however, represent inherent reflexive verbs; if the object is not present, the predicate is interpreted as if it contained a reflexive object. That is, *John washed* means the same as *John washed himself*. But, unlike the understood obligatory reflexives like *behave*, if the object is present, it may or may not be reflexive, as appropriate.

2.3 Aspectual Reflexive

The difference between (8) and (9) below has been related to the telicity of the verb (Sanz, 2000), in that the examples in (9) indicate a completion of the event expressed in (8), i.e. a change to an ACCOMPLISHMENT event type (Vendler, 1967:Ch. 4):

(8) a. Juan comió la sopa. "Juan ate the soup."
b. Juan leyó el libro. "Juan read the book."
(9) a. Juan *se* comió la sopa. "Juan ate up the soup."
b. Juan *se* leyó el libro. "Juan read the book [to completion]."

The sentences in (8) are consistent with an ACTIVITY reading in which a modifying adverbial such as *por una hora* "for an hour" is licit, as in (10), whereas the sentences in (9) are infelicitous with *por una hora*, as shown in (11):

(10) a. Juan comió la sopa por una hora. "Juan ate the soup for an hour."
b. Juan leyó el libro por una hora. "Juan read the book for an hour."
(11) a. *Juan *se* comió la sopa por una hora.
b. *Juan *se* leyó el libro por una hora.

Replacing *por una hora* with an adverbial consistent with an ACCOMPLISHMENT reading, such as *en una hora* "in an hour", renders the sentence acceptable:

(12) a. Juan *se* comió la sopa en una hora. "Juan ate up the soup in an hour."
b. Juan *se* leyó el libro en una hora. "Juan read the book [to completion] in an hour."

In this respect, aspectual *se* can be said to affect the event structure of the predicate.

3 Nonpersonal Reflexives

The general characteristics that identify nonpersonal reflexives are the following. First, the logical subject is always suppressed. Second, the logical object, if there is one, may become the surface subject, or it may remain *in situ*, i.e. in its postverbal position. The postverbal object usually (but not always) triggers agreement with the verb. Finally, the clitic *se* is always present. Four types of nonpersonal reflexives are identified below: impersonal, passive, inchoative, and middle.

3.1 Impersonal Reflexive

In impersonal reflexives such as (2) above and (13) below, no agentive subject is present, neither explicitly nor as a dropped pronoun:

(13)a.	*Se* cree que mienten.	"It is believed that they lie."
b.	*Se* baila toda la noche.	"There is dancing all night."
c.	*Se* vende casas.	"Houses are sold."

The translation of impersonals into English commonly makes use of expletive subjects (e.g. *it* in (13a), *there* in (13b)) or passive voice if an object is present (13c). Note that in this latter case, the logical object *casas* "houses" does not agree in number with the 3rd person singular verb *vende* "sell", so it cannot be a case of the passive reflexive, described in the next subsection. (This particular usage of the impersonal is subject to regional variations in acceptability.) An alternative strategy in English is to use an impersonal subject such as *one*, *you*, *they*, or *people*.

3.2 Passive Reflexive

The most common of the nonpersonal reflexives is the passive reflexive, in which the logical object is always present, either preverbally, as in (14a), or postverbally, as in (14b):

(14)a.	Un temblor *se* registró.	"A tremor was registered."
b.	*Se* registró un temblor.	"A tremor was registered."

The logical object behaves like a subject in that it always agrees in person and number with the verb, as shown in (15):

(15)a.	Unos temblores *se* registraron.	"Some tremors were registered."
b.	*Se* registraron unos temblors.	"Some tremors were registered."

Furthermore, as subject, the logical object may pronominalize and thereby drop, as in (16):

(16)a.	*Se* registró.	"It was registered."
b.	*Se* registraron.	"They were registered."

The passive reflexive differs from a true passive by not involving a form of the auxiliary *ser* "be" followed by a past participle, and by not allowing the agentive subject to appear in a *por* "by" phrase. For example, the sentence in (17) uses the true passive form *fue registrado* "was registered", followed by the agentive subject headed by *por* "by", whereas (18) illustrates the unacceptability of the agentive *por* phrase with the passive reflexive:[4]

(17) Un temblor *fue registrado* ayer *por* el instituto de geofísica.
"A tremor was registered yesterday by the geophysical institute."
(18) *Un temblor *se* registró ayer *por* el instituto de geofísica.

As its name suggests, the passive reflexive is most often translated into English using the passive voice.

3.3 Inchoative Reflexive

The causative/inchoative alternation[5] occurs in both Spanish and English, as shown in (19), where the object *el agua* "the water" in the causative sentence in (19a) becomes the subject in the inchoative in (19b), and the causative subject *la cocinera* "the cook" is suppressed:

(19)a. La cocinera hirvió el agua. "The cook boiled the water."
b. El agua hirvió. "The water boiled."

The alternation is lexically determined; for example, *hervir* "boil" above allows the alternation, but *abrir* "open" does not, although English *open* does, as shown in (20):

(20)a. El viento abrió la puerta. "The wind opened the door."
b. *La puerta abrió. "The door opened."

To express the inchoative variant "the door opened" in Spanish, the reflexive *se* is introduced, as in (21):

(21)a. La puerta *se* abrió. "The door opened."
b. *Se* abrió la puerta. "The door opened."
c. *Se* abrió. "It opened."

As (21) shows, the inchoative reflexive in Spanish behaves in all respects like the passive reflexive. That is, the logical object may be preverbal or postverbal, it behaves like the subject in that it agrees with the verb, and it may pronominalize and drop.

---

[4] However, counterexamples such as the following appear to be attested:
(i) Este verbo *se* ha usado por buenos escritores.
"This verb has been used by good writers."
I have no satisfactory analysis for such constructions.
[5] I adopt Levin's (1993) and Haspelmath's (1993) terminology. Burzio (1981) uses the term "ergative" for the intransitive variant. Zubizarreta (1987) refers to the "causative/anti-causative" alternation, and Sanz (2000) to the "transitive/uncausative" alternation.

3.4 Middle Reflexive

The middle construction, typified in (22), also shows the “raising” of object to subject, with the Spanish variants showing concomitant subject-verb agreement:

(22)a. Estos libros *se* venden bien. “These books sell well.”
b. Un traje de seda *se* lava fácilmente. “A silk dress washes easily.”

Middles always require *se*, and are usually characterized by the presence of an adverbial of manner, and/or convey a modal notion of ability or possibility, among other properties (Fagan, 1992). In all other respects, Spanish middle reflexives, like the inchoative reflexives, behave like the passive reflexives.

4 Analysis

Given the assortment of constructions in which *se* appears, the challenge is to find a unifying analysis that explains the diversity in as perspicuous a manner as possible, rather than treating the constructions as isolated and independent phenomena. The proposal made here suggests a very simple relationship between clitics and the verbal argument structure. That relationship is based on the notion of argument discharge, where the hypothesis is that a clitic always discharges an argument. The nature of the argument discharged by the reflexive clitic *se* determines the type of reflexive construction.

4.1 Phrase Structure

The analysis is situated in a generative grammar perspective, effecting a hybridization of the Principles and Parameters framework (Chomsky, 1986; Webelhuth, 1995) and the more computationally elaborated theory of HPSG (Pollard & Sag, 1994). I adopt a nonmovement analysis in which syntactic structure directly reflects the parsed word order. I also adopt the notion of HEAD PROJECTION, where every lexical item in a sentence is treated as a head, projecting its own phrasal structure when another constituent attaches to it. The attachment follows one of the two basic binary-branching patterns in (23):

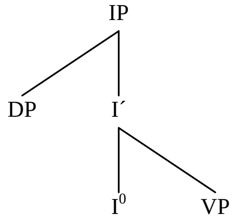
(23)

where X is the category of the projecting head, and Y is the attaching structure, which may be interpreted as either an argument to, or a modifier of, the head. The constituents are annotated with the binary features [±MAX, ±MIN] (rather than traditional bar-levels), indicating the constituent's status as a maximal or minimal projection (Kitagawa (1986), borrowing from a proposal by

Muysken (1982)). Using this notation, lexical heads are always annotated [+MIN] and maximal projections are always [+MAX]. The local head of the trees in (23) is nonmaximal ([–MAX]) and the local projection is nonminimal ([–MIN]). The attaching structure is always interpreted as a maximal projection (Jackendoff, 1977). These binary features [±MAX, ±MIN] are consistent with Chomsky's (1995) Bare Phrase Structure, where a nonprojecting lexical head is itself a maximal projection. For example, pronouns, clitics, unmodified adverbs and adjectives, etc., have the structural features [+MAX, +MIN]. However, for expository convenience I will use the more usual terms of XP for maximal projection ([+MAX]), X^0 for minimal projection ([+MIN]), and X′ for intermediate projection ([–MAX, –MIN]).

I adopt the now-standard DP analysis (Abney, 1987), in which determiners take NPs (and potentially other categories) as complements. Pronouns and clitics are also represented as lexical DPs. The clause, i.e. subject and predicate, is represented as a projection of I(nflection), containing tense and agreement features. The basic structure of the clause is represented as follows:

(24)



with DP the subject and VP the predicate.

The question arises as to where clitics are structurally positioned. That Romance clitics always precede tensed verbs (as proclitics) suggests a close connection to the inflectional category I, the locus of tense.[6] Furthermore, clitics are not necessarily semantically related to the tensed verb they precede, as occurs in clitic climbing:

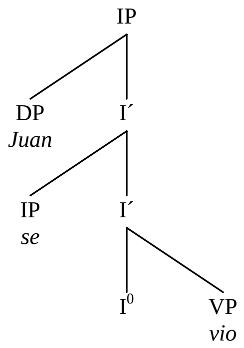
(25)a. Tú tienes que hacer*lo.* “You have to do it.”
b. Tú *lo* tienes que hacer. “You have to do it.”

In (25a), the accusative clitic *lo* “it” is semantically related to the infinitive verb *hacer* “do” as its THEME. It is not, however, semantically related to the tensed verb *tienes* “have”, in front of which the clitic has “climbed”. Thus, there is little motivation for suggesting that the clitic attaches to the verb directly;

[6] Clitics as suffixes (enclitics) may immediately follow infinitives, imperatives and present participles (and also tensed verbs in Old Spanish). Such enclitics are not treated in this paper; however the properties reported here, in particular with respect to argument discharge, also carry over to enclitics.

instead, I propose that clitics left-adjoin to the Inflection projection, as illustrated in (26) for the sentence *Juan se vio* “Juan saw himself ”:[7]

(26)



In (26), the clitic *se* (as with all proclitics) adjoins to I′. It does not adjoin to IP, since then it would precede the subject, which does not occur:

(27)a. **Se* Juan vio.
b. [$_{IP}$ *se* [$_{IP}$ Juan vio]]

Nor does the clitic adjoin to I^0. Given the theory advocated here that clitics always discharge an argument, no argument information is available at the I^0 node. Hence it cannot adjoin to I^0 but must adjoin higher in the tree, i.e. at I′, where argument information from the VP becomes available. This is discussed in the following section.

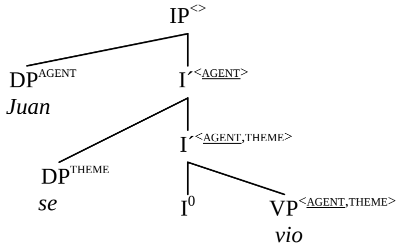
4.2 Argument Structure

The argument structure of a verb (or any predicate) indicates the number and nature of arguments required in order for a sentence containing the verb to be considered complete and coherent. Many notational devices have been proposed for representing argument structure; here, I use an angle-bracketed list of arguments bearing θ-roles, noting that the full set of subcategorization features must specify (explicitly or through some derivative process) syntactic and semantic restrictions. Furthermore, following Williams (1981), one argument will be distinguished as external, occurring outside the domain of the verb phrase (in particular, in subject position); all others are internal, i.e. within the projection of the verb. When a particular constituent satisfies (or unifies with) one of the predicate's argument specifications, that argument is discharged. The undischarged arguments in the list are percolated up the tree

[7] Many details of structure are being simplified here, such as ordering of clitics, the relation of clitics to negation and question formation, etc. A more articulated structure for the I projection may ultimately be required. I leave that prospect for future research.

until such time as they can be discharged and thereby saturate the subcategorization list (i.e. all arguments discharged). The two positions of interest where such discharge may occur are the subject position and clitic position.[8] The argument structure for the construction in (26), for example, is illustrated in (28):

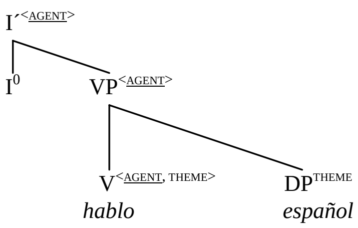
(28)



The verb *vio* "saw" specifies two arguments, AGENT and THEME; following Williams' convention, the external argument is marked by underlining. Since no arguments are directly discharged within the VP, the entire argument list is percolated up to the immediately dominating I′ node. From there, the clitic *se* discharges the THEME argument, thereby reducing the arguments in the argument list to just the AGENT, which because it is marked as external, is discharged by the constituent in subject position, *Juan*. The argument list in the IP is then fully saturated (< >), and the sentence is thus well formed.

In null subject (*pro*-drop) constructions, the subject is not phonetically realized and hence is not available to be integrated by the parser into the structure of the clause. For the sentence *Hablo español* "I speak Spanish", the syntactic structure that is generated by the parser is represented as follows:

(29)

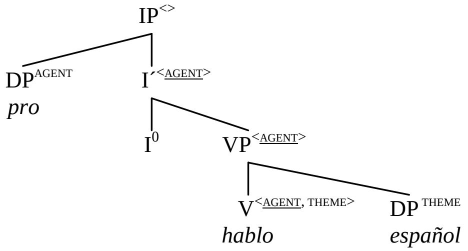


An unrealised external argument in the argument list at I' is always interpreted as the null subject ("little *pro*"), with person and number features

[8] Other positions where arguments may be discharged include the specifier of CP for *wh*-marked arguments and adjunctions to IP for topicalized arguments. I ignore these here.

matching those of the tensed verb, in this case 1st person singular.[9] At the appropriate level of representation (discussed below), the pronoun is inserted into the structure in accordance with the EXTENDED PROJECTION PRINCIPLE (Chomsky, 1982:10), which states that all clauses have subjects, so that (29) becomes analysed as the structure in (30):

(30)



With this basic architecture, the analysis of reflexive *se* can now be presented.

4.3 Reflexivization

All DPs have referential properties that affect their ability to enter into reflexive or nonreflexive relationships with other DPs. Following Bouchard (1983), I take these properties to include a referential index, agreement features (person, number and gender), and semantic features such as [±ANIM]. When two DPs are coreferential, they must necessarily have matching (=unifying) referential properties. If they are not coreferential, then at least the referential indices are distinct. Nouns receive a unique referential index inherently, and anaphora take their index value from their antecedents. Expletive pronouns such as *it* and *there* have a referential index value of zero.

A reflexive DP obeys certain strict morphological, syntactic, and semantic constraints, as elaborated by a theory of anaphora binding. For present purposes, the following REFLEXIVE PRINCIPLE can be taken as the basic syntactic requirement of reflexives:[10]

(31) REFLEXIVE PRINCIPLE:
A constituent X is a reflexive if and only if X is bound in its local binding domain.

The relevant definitions for the terminology used in (31) are as follows:

---

[9] In the case of infinitival constructions, the null subject's features are unified with those of the matrix subject or object, depending on whether the matrix verb subject-controls or object-controls the null subject.

[10] Cf. Haegeman's (1994:221) *Principle of Reflexive Interpretation*: "A reflexive X must be bound in the minimal domain containing X, X's governor, and an accessible subject/SUBJECT." My REFLEXIVE PRINCIPLE is a simplication thereof.

- α binds β iff α c-commands β, and α and β are coindexed.
- α c-commands β iff α and β are sister nodes, or α and γ are sisters and γ dominates β.
- α and β are coindexed iff they share the same referential properties (index, agreement, and semantic features).
- The local binding domain (in the context here) is the (finite) clause, i.e. the closest IP which dominates the anaphor in question and a subject.

Morphologically, reflexives in English are identified by the suffix -*self* or -*selves* on the pronoun.[11] In Spanish, reflexives are realized as clitics (aside from the prepositional reflexive object *sí*, as in *Solo piensa en sí* "He only thinks of himself", which is not treated here). The clitic may be reflexive or not, depending on its lexical properties. Thus, the accusative clitic *lo*, like dative *le* (and their plural and feminine counterparts) are marked nonreflexive [–REFL], whereas *se* is marked reflexive [+REFL]. 1st and 2nd person clitics (*me*, *te*, *nos*, *os*) may be either reflexive or nonreflexive, depending on the context, and so are disjunctively marked [±REFL].

When a clitic DP attaches to the I′ projection, it is in a position to discharge an argument from the argument list. When the DP discharges an internal argument, its configuration is that shown in (32a), where *i* and *j* are the indices of the external and internal argument, i.e. α and β, respectively. If the clitic discharges the external argument, the I projection will be either that in (32b) if an internal argument remains undischarged, or (32c) if no internal argument remains to be discharged:

(32) a. $[_{I'^{<\underline{\alpha}i>}}\ DP_j^{\beta}\ \ I'^{<\underline{\alpha}i,\ \beta j>}]$ b. $[_{I'^{<\beta j>}}\ DP_i^{\alpha}\ \ I'^{<\underline{\alpha}i,\ \beta j>}]$ c. $[_{I'^{<>}}\ DP_i^{\alpha}\ \ I'^{<\underline{\alpha}i>}]$

If the clitic DP in (32a) is marked [+REFL], the REFLEXIVE PRINCIPLE ensures that the clitic and its binder, the subject (=the external argument), share referential properties; i.e. *i=j*. The result is a personal reflexive construction. If the clitic DP is marked [–REFL], then the REFLEXIVE PRINCIPLE prevents α and β from being coindexed, and a nonreflexive clitic construction like that in (4) results. In (32b) and (32c), the clitic DP has discharged the external argument. The only clitic available that can discharge anything other than an internal argument is *se*. Since *se* is marked [+REFL], it requires a subject in order to

[11] I disregard the logophoric interpretation of –*self* pronouns, as in:
(i) That paper was written by Thomas and *myself*.
They clearly need to be addressed in a complete system for analyzing English reflexives. However, since Spanish does not appear to have such logophoric pronouns, it is not an issue here. See (Roberts, 1997:171-172) and (Huang, 2000) for discussion.

satisfy the REFLEXIVE PRINCIPLE. As there is no longer an external argument to become subject, something must happen in order to save the analysis. I propose that when *se* discharges an external argument, it confers externalhood on an undischarged internal DP argument, if there is one. That argument will then become the subject, binding the reflexive clitic. In (32b), for example, β will be externalised (if it is a DP), so that the constituent discharging β will be in subject position, and hence bind the reflexive clitic. If, however, there is no undischarged internal DP argument, as in (32c), then there is no argument that can appear in subject position. By the EXTENDED PROJECTION PRINCIPLE, an expletive *pro* is introduced into subject position. It is this 3rd person semantically null subject which binds the reflexive, which is necessarily *se* since it is the only 3rd person clitic that could have zero reference, and the REFLEXIVE PRINCIPLE is maintained.

5 Implementation

The analysis presented here has been implemented in a Spanish-English bidirectional machine translation system, running at the UNAM in Mexico City since 1994 (`http://triton.dgsca.unam.mx/traduce`). The system, coded in Prolog and employing a unification-based formalism, is called MMT (Multilingual Machine Translation), a descendant of the CAT2 framework (Sharp, 1991). The formalism provides a means of describing linguistic objects at various levels of representation as well as transformations between the levels. The rules for each language are defined independently of any target language, so that new languages can be added without affecting existing language descriptions. In addition, language-independent rules (i.e. principles), such as those instantiating the HEAD PROJECTION configurations (cf. (23)), are factored out and made available to all language-specific levels. Such “grammar sharing'” further contributes to the multilingual functionality of the system. The overall architecture is illustrated in Figure 1.

The syntactic structure of a Spanish text is labelled SLES. To accommodate translation, the structure is normalized to a flattened predicate-argument structure at the translational level TLES. Between the two is an interface level ILES; it is here that constituents are added or removed as part of the normalization process. For example, null subjects are inserted into ILES during analysis and removed during synthesis. The English grammar consists of analogous levels of representation. Translation is performed by bidirectional transfer rules between TLES and TLEN. The levels labelled SL, IL, and TL

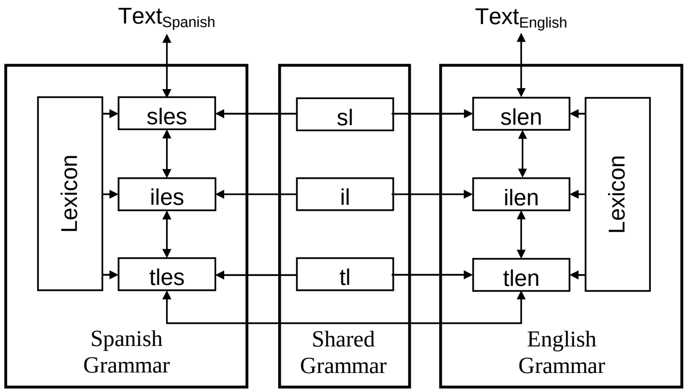


Figure 1: Architecture of MMT

comprise the shared grammar components for the language-specific syntactic, interface, and translational levels.[12]

The lexicon consists of a set of lexical entries, each represented as an attribute-value matrix. Each entry specifies its morphological, syntactic, and semantic properties, including its combinatorial properties with other constituents, such as arguments and/or modifiers. For example, a partial entry for the verb *ver* "see" is illustrated in (33) below:

(33)

$$\begin{bmatrix} \text{lex} & \text{ver} \\ \text{argstr} & \begin{bmatrix} \text{ext} & \begin{bmatrix} \text{sem} & \begin{bmatrix}\text{role} & \text{agent}\\ \text{nom} & [\text{anim} \;\; +]\end{bmatrix} \\ \text{syn} & [\text{cat} \;\; \text{d}] \end{bmatrix} \\ \text{arg1} & \begin{bmatrix} \text{sem} & [\text{role} \;\; \text{theme}] \\ \text{syn} & [\text{cat} \;\; \text{d}]\end{bmatrix}\end{bmatrix} \\ \text{syn} & [\text{cat} \;\; \text{v}] \end{bmatrix}$$

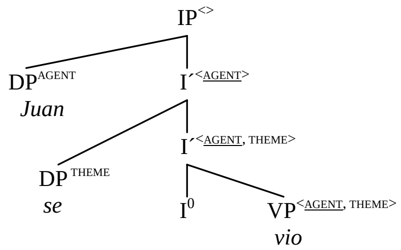
The argument structure specifies two arguments: `ext`, an external DP animate AGENT, and `arg1`, an internal DP THEME.

5.1 *Personal Reflexives*

Consider first the referential reflexives, such as *Juan se vio* "John saw himself", whose syntactic structure is illustrated in (28), repeated here:

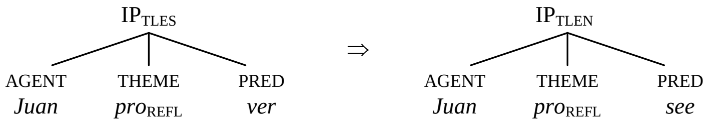
[12] A morphological level, whose details are not relevant here, is also present between the textual form and the syntactic level.

(28)



The clitic *se* discharges the THEME argument, as in (32a) above, leaving the external AGENT to be discharged by *Juan* in subject position. By the REFLEXIVE PRINCIPLE, the clitic, being reflexive, forces the AGENT and THEME to share referential properties, i.e. referential index, agreement features, and semantic nominal features. The interface structure neither inserts nor deletes any constituents, so the ILES and TLES representations have the same form, which is directly transformed to the equivalent English TLEN, shown in (34):

(34)



The THEME reflexive pronoun on the English side takes the same semantic properties and θ-role as the Spanish pronoun, and by activation of the REFLEXIVE PRINCIPLE for English it is generated as the reflexive pronoun *himself*, yielding the sentence *Juan saw himself*. Note that in the reverse translation from English TLEN to Spanish TLES, the REFLEXIVE PRINCIPLE will also ensure that the reflexive pronoun is generated as *se* by virtue of the 3rd person feature on the external argument *Juan*, unifying with the reflexive pronoun.

Obligatory reflexives have the property that one argument has the distinguished θ-role REFL. Because it is an internal argument, its discharge will also give rise to a personal reflexive. For example, the obligatory reflexive verb *quejarse* “complain” has the lexical entry in (35).

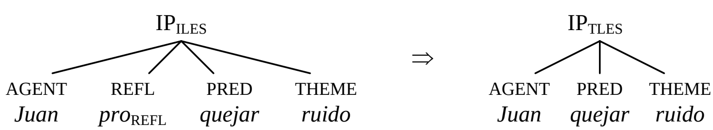
The `ext` definition specifies the referential properties of the external argument. The `arg1` specification provides for a *de*-DP THEME argument, as in *Se queja del ruido* “He complains about the noise”. The `arg2` specification defines the reflexive argument with referential properties unified with those of the external argument.

(35)

$$\begin{bmatrix} \text{lex} & \text{quejar} \\ \text{argstr} & \begin{bmatrix} \text{arg1} & \begin{bmatrix} \text{sem} & \begin{bmatrix} \text{role} & \text{theme} \end{bmatrix} \\ \text{syn} & \begin{bmatrix} \text{cat} & \text{p} \\ \text{pform} & \text{de} \end{bmatrix} \end{bmatrix} \\ \text{arg2} & \begin{bmatrix} \text{sem} & \begin{bmatrix} \text{role} & \text{refl} \\ \text{nom} & \text{NOM} \\ \text{ref} & \text{REF} \end{bmatrix} \\ \text{syn} & \begin{bmatrix} \text{cat} & \text{d} \\ \text{arg} & \text{AGR} \\ \text{refil} & + \end{bmatrix} \end{bmatrix} \\ \text{ext} & \begin{bmatrix} \text{sem} & \begin{bmatrix} \text{role} & \text{agent} \\ \text{nom} & \text{NOM} \\ \text{ref} & \text{REF} \end{bmatrix} \\ \text{syn} & \begin{bmatrix} \text{cat} & \text{d} \\ \text{agr} & \text{AGR} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{syn} & \begin{bmatrix} \text{cat} & \text{v} \end{bmatrix} \end{bmatrix}$$

Since an argument bearing the θ-role REFL has no independent semantic content, it is eliminated in the transformation from ILES to TLES, as illustrated below:

(36)



The TLES structure is now isomorphic to the English equivalent. When translating in the opposite direction, from English to Spanish, the reflexive pronoun is inserted into the ILES structure from TLES, triggered by the designated θ-role REFL in the argument structure of the predicate *quejar*.

The aspectual reflexive, as in *Juan se comió la sopa* “Juan ate up the soup” (=(9a)), presents a challenge to the model proposed here, as the reflexive clitic does not obviously discharge a thematic argument. The solution lies in introducing an (internal) event argument, which *se* can discharge, as does (presumably) English *up*. The content of this event argument defines the aspectual properties of the event in terms of (at least) the two binary features [±TELIC, ±PUNCTUAL] (Sanz, 2000). The four combinations that these two features yield coincide with the four Vendler event types, as indicated below:

(37)a.	[+TELIC, +PUNCTUAL]	ACHIEVEMENT
b.	[+TELIC, –PUNCTUAL]	ACCOMPLISHMENT
c.	[–TELIC, +PUNCTUAL]	ACTIVITY
d.	[–TELIC, –PUNCTUAL]	STATE

A verb phrase acquires event features in the course of its derivation, in combination with a delimiting or nondelimiting direct object, if any, and under modification by adverbials, such as *in an hour, for an hour*, etc. The event argument is represented within the argument structure as follows:

(38)

$$\begin{bmatrix}\text{argstr} & \begin{bmatrix}\text{e} & \begin{bmatrix}\text{telic} & \pm \\ \text{punctual} & \pm\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

so that e is percolated up the tree along with the thematic arguments. When the argument structure indicates an ACCOMPLISHMENT ([+TELIC,–PUNCTUAL]), the event argument may be discharged by *se*:

(39)

$$\text{I}'^{\langle\underline{\alpha}\rangle}$$

DP *se* $\quad$ $\text{I}'^{\langle\underline{\alpha},\ \text{e} = [+\text{TELIC},-\text{PUNCTUAL}]\rangle}$

This configuration of discharge ensures that only ACCOMPLISHMENT events may cooccur with aspectual *se*, and the other event types may not (Sanz, 2000:52):

(40)	ACTIVITY:	*Juan *se* condujo el coche.	"Juan *se* drove the car."
	STATE:	*Juan *se* odia las acelgas.	"Juan *se* hates chard."
	ACHIEVEMENT:	*Juan *se* empezó a correr.	"Juan *se* started running."

As Sanz points out, non-ACCOMPLISHMENT events may become ACCOMPLISHMENT events given an appropriate context. In such cases, aspectual *se* may occur. For example, the atelic ACTIVITY in (41a) may become the telic ACCOMPLISHMENT in (41b) when augmented with a delimiting adverbial:

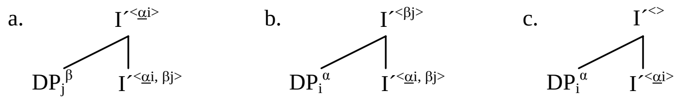
(41)a.	*Juan *se* visitó Alemania.	"Juan visited Germany."
b.	Juan *se* visitó Alemania de norte a sur.	"Juan visited Germany from North to South."

When the event argument is discharged, the thematic argument structure is unaffected, and all undischarged thematic arguments are percolated up the tree unchanged. In particular, the external argument of the verb must still be discharged, as shown in the dominating node I′ in (39). Since an internal (event) argument has been discharged by reflexive *se*, a personal reflexive construction results.

5.2 Nonpersonal Reflexives

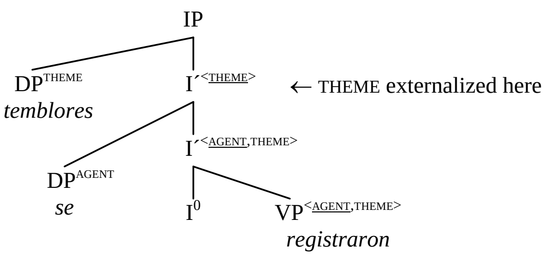
All nonpersonal reflexive constructions share one characteristic in common: the external argument has been discharged by *se*. Recall from (32), repeated here, the configurations where *se* can discharge either an external or internal argument:

(32)



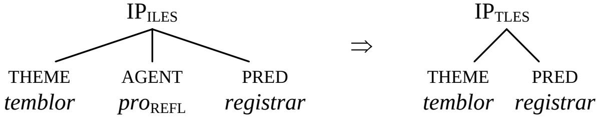
When *se* discharges an internal argument, as in (32a), the external argument remains undischarged and a personal reflexive is generated, as in *Juan se vio* "Juan saw himself". However, *se* can just as easily discharge the external argument, leaving the internal argument undischarged, as in (32b). When *se* discharges the external argument in this way, it also transfers externalhood to the undischarged argument, as long as it is a DP and hence suitable as a subject. An example is the sentence *Temblores se registraron* "Tremors were registered", where the logical object *temblores* has become the surface subject, agreeing in person and number with the verb. The structure of this sentence is shown below:

(42)



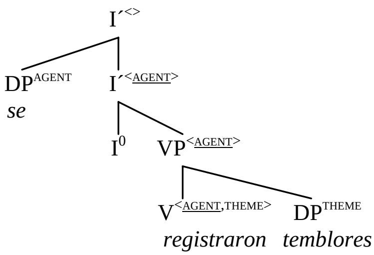
At ILES, all thematically marked constituents are normalized, but in the transformation to TLES, the reflexive pronoun that discharged the external argument is removed. This is illustrated in (43), yielding the English translation "Tremors were registered". All instances of passive, inchoative, and middle reflexives where the logical object has been fronted into subject position are modelled exactly in this way.

(43)



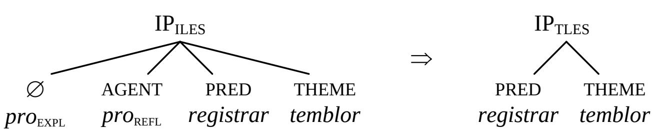
Consider now (32c). This covers the cases where *se* discharges the external argument, and no undischarged arguments remain to be discharged. Such a condition arises in one of two situations: Either in nonpersonal reflexives where the logical DP object has already been discharged, as in *Se registraron temblores* “Tremors were registered”, or in impersonal reflexives such as *Se sabe que mintieron* “It is known that they lied”, where an internal clausal argument has already been discharged, or where the verb is intransitive, as in *Se bailaba* “There was dancing”. The first case has the following structure:

(44)



Although *se* discharges the external argument, it is not in subject position. For one, Spanish, unlike French, has no subject clitics. Second, if *se* were a pronominal in subject position, it could be deleted under *pro*-drop, but the sentence Ø *registraron temblores* can only mean the active voice “They registered tremors”. So, the subject has not yet been introduced into the structure in (44). As pointed out earlier, the solution is to provide the subject as an expletive *pro*, which is then in a position to bind reflexive *se*, thus satisfying the REFLEXIVE PRINCIPLE. From ILES to TLES, the expletive is deleted, since it bears no semantic content and has zero reference. Furthermore, since the clitic *se* also has zero reference, by virtue of being bound by the expletive, it too is deleted. This is illustrated in (45):

(45)



The structure of the IP at TLES is equivalent to that in (43), aside from constituent order, which is irrelevant at the translational level. The English generation module uses any of the available devices for synthesizing an appropriate translation when no referential subject is present. In this case, since a logical object is available, “tremors”, it also produces the passive sentence “Tremors were registered”.

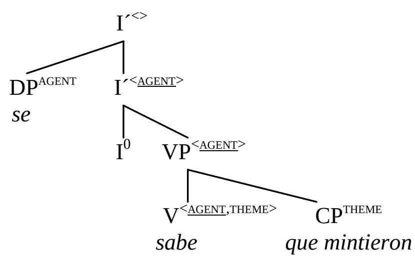
An unresolved question surrounds the agreement of the verb with the internal object. In (44), the verb *registraron* appears to agree with the object *temblores*, both being plural in number. Normally, in Spanish as in English, subject-verb agreement prevails, but object-verb agreement does not. The only exception in English appears to be in *there*-constructions, such as:

(46)a. There goes my favorite car.
b. There go my favorite cars.

If *there* is an expletive occupying subject position, then it is mysterious why the verb should agree with the object. One analysis is to assume that precisely the expletive *there* somehow mediates the agreement between the verb and the associate of *there*. Thus, in (46a) the verb *goes* agrees with the singular object *car*, whereas in (46b), *go* agrees with the plural *cars*. I assume that the same phenomenon happens in Spanish; the expletive *pro* mediates the agreement of the verb with the object, thus giving rise to the distinction between *Se registraron temblores* "Tremors were registered" and *Se registró un temblor* "A tremor was registered".

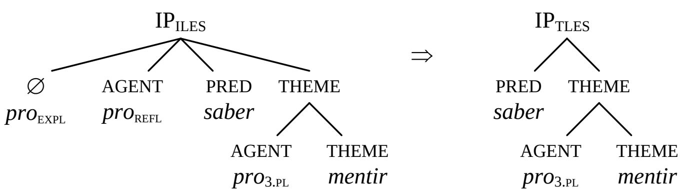
Returning now to impersonal reflexives, we see that the configuration in (32c) again comes into play, where no argument is available for discharge in subject position. For example, the sentence *Se sabe que mintieron* "It is known that they lied" has the SLES structure in (47):

(47)



As in (44) above, an expletive *pro* is inserted into subject position at ILES to satisfy both the EXTENDED PROJECTION PRINCIPLE and the REFLEXIVE PRINCIPLE. Both the expletive and the reflexive AGENT are removed at TLES, yielding the structures:

(48)



In such configurations where the THEME is clausal and there is no external argument, the English generation module inserts expletive *it* at ILEN to produce the translation “It is known that they lied”.

The examples above have illustrated the structures involved in passive reflexives and impersonal reflexives. In that the inchoative reflexives and middle reflexives behave like the passive reflexives, as pointed out in subsections 3.3 and 3.4, these same structures and the processes that create them also carry over to the inchoative and middle reflexives.

Summarizing, the reflexive clitic *se* discharges the external argument in all nonpersonal reflexive constructions. A DP object can be externalized and thus realized in subject position, or it can remain *in situ*, with expletive *pro* being introduced in subject position. The analysis and implementation of nonpersonal reflexive constructions is considerably simplified: either externalize an undischarged DP if there is one, otherwise, introduce expletive *pro*.

6 Translation

The theory of translation behind the present system is based on the principle of semantic homomorphism, i.e. the semantic content at one translational level corresponds to the semantic content at all translational levels. For this reason, only the semantically relevant constituents in a clause are present at the translational levels; they have been stripped of expletive pronouns, obligatory reflexive pronouns, and reflexive AGENTs, as well as the grammatical formatives particular to a specific language. These constituents are fully recoverable in synthesis, based on the argument structure of the target predicate, in conjunction with the fulfillment of the EXTENDED PROJECTION PRINCIPLE and the REFLEXIVE PRINCIPLE. This entails that whatever contentful thematic constituents are present at TLES, the translations of those constituents and none others are required to be present at TLEN, and vice versa.

In addition to maintaining thematic content, the principle of semantic homomorphism also applies to the semantic feature content of the constituents at the translational levels. This is embodied in the following general transformation rule between any two constituents at the translational level:

(49) `[sem = X]` ⇔ `[sem = X]`

By this rule, all semantic features must unify universally. In particular, θ-roles at the translational levels are retained across languages, which is why expletive pronouns and obligatory reflexives, which are language-specific, are absent from the translational level. Also retained across languages by the rule in (49) are semantic nominal features, such as the referential index and animacy.

Ambiguities are inherent in translation, particularly sentence-based translations without the benefit of context or world/common knowledge. The same applies to the translation of sentences involving the reflexive clitic *se*. This is particularly acute in null subject contexts. For example, the sentence in (50a) with an explicit subject is unambiguous, but (50b) is not:

(50)a. Él se ha infectado. "He has infected himself."
b. Se ha infectado. "He has infected himself." / "He/she/it has been infected."

Similarly, the Spanish sentence in (51) can have a number of possible translations:

(51) Se va a matar! a. "He's going to kill himself!"
b. "Someone's going to kill!"
c. "He's going to be killed!"

In (51), there are three potential argument slots which *se* could conceivably discharge: the external argument of *va* (*ir*), the external argument of *matar*, and the (optional) internal argument of *matar*. In (51a), *se* discharges the internal argument of *matar* "kill", whose external argument is controlled by the subject of the matrix verb. This is a standard case of a referential reflexive. The null subject (external argument) of the matrix clause is recovered at ILES as a 3rd person singular pronoun, which then determines the referential properties of the reflexive clitic. (The gender of the pronoun depends on its extrasentential reference, the determination of which will depend on other anaphoric resolution methods discussed in this volume.) Assuming masculine gender by default, the corresponding English subject pronoun *he* and the reflexive pronoun *himself* are generated.

In (51b), *se* discharges the external argument of *va* "going to", resulting in an impersonal reflexive. The subject is realized as expletive *pro*, which controls the external argument of *matar*, essentially giving arbitrary reference to both subjects. The internal argument of *matar* is not realized.

In (51c), *se* discharges the external argument of *matar*, but this time the object of *matar*, a pronoun, becomes externalized. The matrix subject (*he*) is now controlling this externalized subject, giving rise to the passive reflexive interpretation. Thus, the three separate translations are not the result of three separate hand-tailored rules, but simply a consequence of *se* discharging either the internal or external argument.

7 Conclusion

The unified treatment of *se* entails a single lexical entry for *se*, a clitic lexically marked as a reflexive by containing the feature [+REFL]. Like all preverbal clitics in Romance, it left-adjoins to I(nflection), from which it is able to discharge a verbal argument. When it discharges an internal argument (thematic or event), a personal reflexive construction results. Discharging the external argument results in a nonpersonal reflexive.

Further research is required into the uses of Spanish *se* not covered here in order to determine the adequacy of the present proposal to those cases. For example, three issues that still need to be addressed include the phenomenon of clitic climbing as in (52), the effect of clitic doubling on personal reflexives as in (53), and the treatment of ethical or possessive datives, as in (54):

(52) Juan *se* tuvo que arrepentir de los pecados mortales.
"Juan had to repent his mortal sins."

(53) Juan *se* dio un regalo *a sí mismo*.
"Juan gave himself a present."

(54) Juan *se* rompió la pierna.
"Juan broke his leg [accidentally]."

Extensions to reflexive uses in other Romance and non-Romance languages follow.

References

Abney, S. 1987. *The English Noun Phrase in its Sentential Aspect*. Ph.D. thesis, MIT.

Borer, H. and Y. Grodzinsky (eds.) 1986. *Syntax and Semantics 19: The Grammar of Pronominal Clitics*. New York: Academic Press.

Bouchard, D. 1983. *On the Content of Empty Categories*. Dordrecht: Foris.

Bruhn de Garavito, J. 1999. *The Syntax of Spanish Multifunctional Clitics and Near-native Competence*. Ph.D. thesis, McGill University.

Burzio, L. 1981. *Intransitive Verbs and Italian Auxiliaries*. Ph.D. thesis, MIT.

Chomsky, N. 1982. *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, MA: MIT Press.

---------- 1986. *Knowledge of Language*. New York: Praeger.

---------- 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Cole, P., G. Herman and C.-T. Huang (eds.) 2001. *Syntax and Semantics 33: Long-Distance Reflexives*. New York: Academic Press.

Fagan, S. 1992. *The Syntax and Semantics of Middle Constructions*. Cambridge: Cambridge University Press.

Haegeman, L. 1994. *Introduction to Government & Binding Theory*, 2nd ed. Oxford: Blackwell.

Haspelmath, M. 1993. "More on the Typology of Inchoative/Causative Verb Alternations." B. Comrie and M. Polinsky (eds.) *Causatives and Transitivity*, 87-120. Amesterdam: John Benjamins.

Huang, Y. 2000. *Anaphora: A Cross-Linguistic Study*. Oxford: Oxford University Press.

Jackendoff, R. 1977. *X′-Syntax*. Cambridge, MA: MIT Press.

Kitagawa, Y. 1986. *Subjects in Japanese and English*. Ph.D. thesis, University of Amherst, Massachusetts. Published as: Kitagawa. 1994. *Subjects in Japanese and English*. New York: Garland Publishing.

Levin, B. 1993. *English Verb Classes and Alternations*. Chicago: University of Chicago Press.

Manzini, R. 1986. "On Italian *si*." See (Borer & Grodzinsky, 1986), 241-262.

Muysken, P. 1982. “Parametrizing the Notion ‘head’.” *Journal of Linguistic Research* 2.57-75.

Otero, C. 1966. “El otro *se*.” *Letras I*.49-57.

Perlmutter, D. 1971. *Deep and Surface Structure Constraints in Syntax*. Holt, Rinehart & Winston.

Pollard, C. and I. Sag 1994. *Head-Driven Phrase Structure*. Stanford, CA: CSLI.

Roberts, I. 1997. *Comparative Syntax*. London: Arnold.

Sanz, M. 2000. *Events and Predications. A New Approach to Syntactic Processing in English and Spanish*. Amesterdam: John Benjamins.

Sharp, R. 1991. “CAT2: An Experimental EUROTRA Alternative.” *Machine Translation* 6.215-228.

Vendler, Z. 1967. *Linguistics in Philosophy*. Ithaka, NY: Cornell University Press.

Webelhuth, G. (ed.) 1995. *Government and Binding Theory and the Minimalist Program*. Oxford: Blackwell.

Wehrli, E. 1986. On Some Properties of French Clitic *se*. See (Borer & Grodzinsky, 1986), 263-283.

Williams, E. 1981. “Argument Structure and Morphology.” *The Linguistic Review* 1(1).81-114.

Zubizarreta, M. 1982. *On the Relationship of the Lexicon to Syntax*. Ph.D. thesis, MIT.

---------- 1987. *Levels of Representation in the Lexicon and in the Syntax*. Dordrecht: Foris.

Section II

Theoretical, Psycholinguistic and Cognitive Issues

Binding and Beyond: Issues in Backward Anaphora*

Eric Reuland and Sergey Avrutin
Utrecht Institute of Linguistics OTS, Utrecht University

In this article, we investigate backward anaphora in English and Dutch from a theoretical and comparative point of view. We particularly focus on what appear to be cases of backward binding, with concomitant violations of c-command and/or weak crossover. A perspective on the encoding of binding relations is developed which obviates this problem.

1 Introduction

In this article, we will focus on principles of anaphora resolution taking certain issues in backward anaphora as our starting point. Backward anaphora has received relatively little attention in the recent theoretical literature (although there is a rather extensive literature in computational linguistics). In the sixties through eighties, the subject received more systematic attention, for instance in works such as (Langacker, 1966), or (Lakoff, 1968), and subsequently in the linking approach to binding developed by Higginbotham (for instance, (Higginbotham, 1983)), and in particular Reinhart (1983), who presents an extensive theoretical study.[1] The diminishing attention in recent years may well be due to the tacit assumption that for one part, backward anaphora trivially instantiates co-reference, and for some other part, equally trivially instantiates reconstruction. In this article, we will show that the conditions on backward anaphora are in fact not so trivial.

We will see evidence that the patterns observed result from the interplay between precedence and structural dependence. The former belongs to the system of use, the latter to the computational system of human language. Languages differ in the extent to which certain interpretive dependencies are grammatically encoded. As we will see, systems of use kick in precisely where grammar leaves matters open. This results supports the proposal that the division of labour between different cognitive components involved in language is governed by general economy principles (Reuland, 2001): in case of competition between systems of use and grammatical computation, the latter

---

* We would like to express our gratitude to António Branco, Denis Delfitto, and Martin Everaert for their very helpful and stimulating comments.

[1] We will be using "backwards anaphora" instead of the somewhat stilted term "kataphora". In line with this, we will often be using the term "antecedent" even where it follows the pronoun dependent on it, and hence, strictly speaking, is a postcedent rather than an antecedent.

wins. Note that Reinhart (1983)'s study of backward anaphora focussed on English. Our results are compatible with her findings that in English precedence does not play a systematic role in standard cases of pronominal binding.

2 The issue

We will be investigating asymmetries in the interpretation of referentially dependent elements in adverbial clauses preceding or following their main clause. Consider to this end the following case of English and its Dutch counterpart:

(1) a. Every student$_i$ turned in an assignment before he$_i$ went home.
    b. Before he$_i$ went home, every student$_i$ turned in an assignment.
(2) a. Iedere student$_i$ leverde een opdracht in voor hij$_i$ naar huis ging.
    b. Voor hij$_i$ naar huis ging, leverde iedere student$_i$ een opdracht in.

(1) and (2) show that at least some form of bound variable dependency is insensitive to the relative position of pronoun and binder. In fact, what is even more striking is that in (1b/2b) *he* en *hij* can apparently be bound by *every student* and *iedere student* respectively, although under standard versions of c-command, *every student* does not c-command *he* (henceforth, we will omit specific reference to Dutch where both languages work the same).[2] Prima facie, in order to accommodate the pattern two ways are open: i) assume that c-command works in such a way that *he* in (1b) is in the c-command domain of *every student*; ii) assume that the *before*-clause is in fact moved from the position in (1a) in which *he* is c-commanded by *every student* and the binding is licensed in the original position by reconstruction.[3]

Reinhart (1983) argues convincingly that the correct definition of c-command should be stated in such a way that at least certain preposed constituents are in the c-command domain of the subject. Specifically, this applies to preposed VP-material, witness the contrast between (3a) and (3b):[4]

---

[2] The currently standard version of c-command is given in (ii) and illustrated in the configuration of (i):
(i) a [$_c$ …b …]
(ii) *a* c-commands *b* iff *b* is contained in the category *c*, *c* a sister of *a*
If *a* is adjoined to *c*, *c* will not qualify as a category, and if there is no category *d* contained in *c* and containing *b*, *b* will command *a*.

[3] Reconstruction is illustrated by the fact that in order to interpret the anaphor in (ia) we compute its binding possibilities from the direct object position, indicated by the trace, by putting it back as in (ib):
(i) a. himself, John didn't admire t
    b. – John didn't admire himself

[4] Reinhart (1983:23) assumes the following definition of c-command:
(i) Node A c(onstitutent)-commands node B iff the branching node α_1 most immediately dominating A either dominates B or is immediately dominated by a node α_2 which dominates B, and α_2 is of the same category type as α_1.

(3) a. In Ben$_i$'s office, he$_i$ is an absolute dictator.
b. *In Ben$_i$'s office, he$_i$ placed his$_i$ new brass bed.

Assuming a structure in which a PP preposed from within VP stays within the sentential domain, under Reinhart's definition, c-command does obtain in (3b), but not in (3a). Under current implementations distinguishing between categories and segments, the effect is explained if the PP in (3b) is adjoined to IP, and the one in (3a) adjoined higher up. Under either approach, the subject c-commands *Ben* in (3b) but not in (3a), leading to a condition C violation in the former, but not in the latter case.[5] However, this line cannot be adopted for (1) and (2). If *he* is c-commanded by *every student* in (1), this implies that (4) has the same status as (3b):

(4) Before John$_i$ went home, he$_i$ turned in his assignment.

This result is incorrect, as (4) is perfectly well formed. This shows that the preposed adverbial clause must be outside the c-command domain of the subject. From this paradoxical result, we can draw the conclusion that alternative takes on c-command do not provide an answer, unless we are willing to assume quite different structures for (1b) and (4), for which there is no motivation.

As already noted, the alternative line is that the *before*-clause has been moved from a position where it is c-commanded by the subject, and gets reconstructed for interpretation. This requires that reconstruction is optional, and the reconstructed version is selected or not on the basis of interpretive need. In (1b), reconstruction is necessary in order for *he* to be bound, hence the reconstructed version is interpreted; in (4), reconstruction would lead to a condition C violation, hence the non-reconstructed version is interpreted. Regardless of the desirability of such optionality, it cannot be all there is, as is shown by the fact that (5) is also grammatical, in contrast to (6):[6]

(5) Before every student$_i$ went home, he$_i$ turned in his assignment.
(6) *He$_i$ turned in his assignment before every student$_i$ went home.

The subject clearly does not c-command the adverbial clause in (5), as *he* does not cause a condition C violation as there is in (6). How, then, can binding obtain in one direction in (1b) and in the other in (5)? Suppose, in a

---

In Reinhart's analysis S and S' (corresponding to current IP/TP and CP) were taken to be of the same category type. S' was in turn dominated by E which was of a different category type.

[5] Condition C states that a referential expression cannot be bound, as illustrated by the impossibility to construe *John* as dependent on *he* (i):
(i) **He* thought that *John* would come.

[6] António Branco (p.c.) informs us that the corresponding construal is unavailable in Portuguese. Cross-linguistic variation in this area is clearly a matter of further investigation.

Chierchia-style analysis, we propose that *every student* in (5) undergoes Quantifier Raising, and is adjoined higher up, high enough to c-command *he*. If so, the same reasoning can be applied to *every student* in (1b).[7] (Note, that if the quantifier moves up, one would expect a weak crossover configuration to arise. For the moment, we will ignore this, coming back to it in Section 5.) For the cases discussed so far, this works. But is it indeed enough? That it is not, will be shown by the properties of a different type of dependency, originally observed in (Evans, 1977:491):

(7) Every logician$_i$ was walking with a boy near that logician's$_i$ house.

As Evans observed, in these cases there is an interpretive dependency between that *logician* and *every logician*. This dependency is reminiscent of a bound variable dependency despite the fact that *that logician* is not a pronoun. This type of dependency has been recently discussed by Noguchi (1997), Ueyama (1998), and Hara (2002).

If we put this type of dependency in the paradigm of (1/2) a contrast obtains, as in (8) and (8'):

(8) a. *Iedere student$_i$ leverde zijn opdracht in, voor die student$_i$ naar huis ging.*
"Every student turned in his assignment before that student went home"
b. **Voor die student$_i$ naar huis ging, leverde iedere student$_i$ zijn opdracht in.*
"Before that student went home every student turned in his assignment"

In Dutch, (8a) is marginally acceptable, but the dependency represented in (8b) is impossible throughout. The situation in English is somewhat different, since for a group of speakers both (8'a) and (8'a) are fine. For present purposes, it suffices that there is a variety of English in which (8'a) is fine, but in which (8'b) is as impossible as Dutch (8b).[8] For the remainder of this paper we will base ourselves on that variety.

(8') a. Every student$_i$ turned in an assignment before that student$_i$ went home.
b. *Before that student$_i$ went home, every student$_i$ turned in an assignment

[7] If we take AdvP to stand for the adverbial clause, QP for the quantifier phrase, CP for the matrix clause and pro for the pronominal, depending on the details of QR and c-command, for (iv) either (i) or (ii) would do. We will not elaborate this issue here.

(i) $[_{CP}$ $[_{AdvP}$ QP $[_{AdvP}$ …]] $[_{CP}$ … pro …]]
(ii) $[_{CP}$ QP $[_{CP}$ $[_{AdvP}$ …] $[_{CP}$ … pro…]]]

[8] Thanks to David Pesetsky, Bill Philip, Sharon Unsworth and Amy Wall for discussion. We have no quantitative data on the pervasiveness of the contrast in English. It should be noted, however, that even in the variety of English in which (8'b) is acceptable, the relation cannot be simply one of standard bound variable anaphora since the contrast does surface with a different choice of quantifier:

(i) a. (?)No student$_i$ turned in an assignment before that student$_i$ went home.
b. *Before that student$_i$ went home, no student$_i$ turned in an assignment.

Again, people differ as to whether (ia) is acceptable, but we have found no speaker for whom (ib) is acceptable.

Thus, unlike (1b), (8'b) is impossible under the dependent interpretation. Regardless of whether adverbial clauses reconstruct or not, and regardless of issues around c-command and QR, the contrast in (8') shows that these cannot solve the problem at hand. If reconstruction were available, (8'b) should be as well-formed under the given interpretation as (1b). But it clearly is not. There is no reason to assume that the *before*-clause in (8'b) is in a different position than the one in (1b). Why, then, would reconstruction be unable to save (8'b) if it can save (1b)? The same holds true of QR. If QR can bring *he* under the scope of *every student* in (1b), why cannot it do so to *that student* in (8'b)? Let it suffice for the moment to conclude that there is a serious puzzle here.

2.1 The nature of bound variable dependencies: from the canonical view to a minimalist perspective

In order to put this puzzle in a theoretical perspective, we will discuss some general issues in the theory of bound variable dependencies. Canonical bound variable anaphora we understand as the relation extensively discussed in works such as (Reinhart, 1983), (Heim, 1998), (Reinhart, 2000) and many others: a pronominal depends for its interpretation on a possibly quantified antecedent in a manner that reflects an operator-variable relation in a logical syntax notation, as in (9).[9]

(9) a. Every lawyer wondered whether Mary would leave after seeing the mess in his office
b. Every lawyer (λx (x wondered [whether Mary would leave [after seeing the mess in x's office]]))

The theoretical status of the logical syntax notation has through the years mostly been left implicit. Is it a notation of convenience, or does it reflect a linguistically significant "level of representation"? This issue gets sharpened in terms of the minimalist program (Chomsky, 1995), which we adopt here. The minimalist program effects a principled restriction on operations within narrow syntax, i.e. C_{HL}. Operations in C_{HL} are subject to the inclusiveness condition. That is, they are restricted to rearranging morpho-syntactic objects (feature bundles) that are part of the numeration. The numeration itself can also only contain morpho-syntactic objects. Operations cannot insert material in the derivation that is not in the numeration, nor can they replace one type of expression by another. At that level, dependencies are exclusively encoded by processes of chain formation: copy, agree, triggered by mechanisms of feature checking. Such a dependency is just a by-product of independent syntactic

[9] For reasons discussed in (Reinhart, 2000), the concept of co-reference must be broadened so as to also allow for co-valuation of variables bound by non-referring expressions. We will henceforth leave this issue aside.

operations blindly applying. Given the encoding mechanisms, such dependencies are also characterized by locality. As discussed in (Reuland, 2001), dependencies between simplex anaphors like Dutch *zich*, Icelandic *sig*, etc. and their antecedents are indeed instantiated as syntactic chains, which are subsequently interpreted in terms of variable binding.

It is easily seen that, by contrast, the mapping of (9a) onto (9b) cannot be done within C_{HL}. This mapping involves replacing pronominals by variables, and inserting a λ (or an equivalent expression). These are precisely among the operations that have no place within C_{HL}. Thus, (9b) cannot be brought about by syntactic operations, and consequently, it cannot be a syntactic level (such as an old-style "LF"). Also, syntactic mechanisms of the type involved in syntactically binding anaphors such as *zich* in the theory of Reuland (2001) cannot apply here. For current purposes it suffices that this must be so since variable binding as instantiated in (9) is not subject to standard syntactic island constraints (Wh-islands, CED).

The dependency in (9) must be represented at some level. Clearly, the language system contains an interpretive component to relate linguistic expressions to however knowledge, believes, etc. are represented in our system of thought. Let us call this component SEM (Chomsky, 1995, 1998, 1999). We will be assuming that the inclusiveness condition only pertains to computations within C_{HL} proper. We already know that the mapping to expressions at the PF interface violates the inclusiveness condition (see the discussion in (Chomsky, 1995, 1998, 1999)). Interpreting an expression by necessity also involves extra-linguistic systems, such as the system of thought. Thus, also the mapping to representations at the Conceptual-Intentional (C-I) interface cannot obey inclusiveness. We will be assuming that expressions reflecting λ-abstraction or some equivalent mechanism are in principle available at the C-I interface and that (9b) is a proper representation at that level.[10] Thus, at the C-I interface chain-type dependencies are represented as variable binding, but variable binding may also reflect dependency relations that are not syntactically encoded.

According to Reinhart (1983) and subsequent work, variable binding of β by α obtains under the following conditions:

(10) i) α and β are co-indexed
ii) α c-commands β
iii) in logical syntax β is translatable as a variable

[10] Kayne (2001) explores the possibility of encoding all bound-variable type dependencies using the mechanisms of 'narrow syntax'. However, in order to be able to do so, additional latitude is introduced into the system, and certain distinctions must be reintroduced that would otherwise be lost. The empirical consequences remain to be explored (see (Reuland, 2003) for some discussion).

Note that some proviso must be made for the use of indices in this formulation. Indices are not morpho-syntactic objects and hence cannot be part of the numeration. Therefore, co-indexing cannot be represented as such. Thus, instead or having indices in syntactic representation, the dependencies must be introduced by the choice of variables in the translation from syntactic expressions into expressions at the C-I interface. But with this proviso, the implementations of Heim (1998) and Reinhart (2000) can be used with just some straightforward changes. The details will not concern us here. See (Reuland, 2001) for some discussion and footnote 5 for a summary of some of the issues.[11]

Clearly, the interpretive system must also allow us express *co-reference*. We are able to assess whether certain different descriptions pertain to the same referent, as in cases like (11):

(11) The robber had entered the vault. John’s accuser swore that he had taken the diamonds.

Here, John may be identical to the robber or not, *he* may refer to the enterer, but need not, since it could also, though less probably, refer to the accuser. Clearly, although one may speak of a dependency between, for instance, *he* and *the robber*, in such cases the dependency is not grammatically encoded. Could we say it is just a matter of accessing a general knowledge base? At least that much is clear at the moment, that if it is a general knowledge base, it is at least to some extent structured according to linguistic principles. That is, referents are accessed via representations that bear markings such as “masculine”, “feminine”, “neuter”, etc., that may be quite arbitrary from the perspective of the properties of the individuals that are denoted. As convincingly shown by

---

[11] Pursuing this route would necessitate the following refinement of the relation between C_{HL} and the language of thought (Chomsky, 1998:3). Chomsky defines a language L to be a device that generates expressions EXP=<PHON, SEM> where PHON provides the instructions for sensori-motor systems, and SEM for systems of thought. As stated in (Reuland, 2001), C_{HL} should in fact be defined as generating expressions EXP' = <PHON', SEM'> that obey the inclusiveness condition. The C-I interface, then, would contain a translation procedure T, which maps SEM' onto SEM (at the sensori-motor side some similar procedure may be assumed). For the sake of concreteness, one may assume the following procedures (Reuland, 2001):

i. An expression *DP[...t....]* is translated as DP (λx (...x...)), or any alphabetic variant (in general, vbl_{DP} is the variable of the λ-expression derived by translating DP)

ii. If an expression is just a bundle of Φ-features (pronouns and SE-anaphors (Reinhart & Reuland, 1993)) it is translated as a variable (vbl_{PRO}), or, if it is free, it undergoes QR and is treated as a DP.

iii. If a dependency is syntactically encoded, i.e. XP is CHAIN-linked to DP their Φ-features are copies. If so, vbl_{XP}=vbl_{DP} (in particular, vbl_{SE}=vbl_{DP}).

If the dependency is not syntactically encoded, binding obtains by choosing the right alphabetic variant upon translation of the pronominal and subsequently the binder. As Denis Delfitto (p.c.) points out, some morpho-syntactic property does enter the procedure, since agreement (matching in phi-features) is required in order for an arbitrary set S of pronouns to be translated into the SAME variable. However, this is more like a filter than an encoding in the sense intended.

Ariel (1990), accessibility of referents, in addition to 'recency of mention', also reflects much more linguistic properties such as 'topicality', and properties reflecting the richness of the linguistic description by which they have been introduced. We will refer to this part of the interpretive system as 'discourse structure'. For concreteness sake, we will be assuming that discourse structure is part of the C-I interface.[12]

Note, however, that bound variable anaphora even if put outside of narrow syntax, along minimalist lines, still is a very "syntactic" procedure. Variables are simply the C-I interface representations of pure bundles of phi-features, pronominals, anaphors and traces left by QR, not of determiners with deictic force. Binding relations can be assessed with no recourse to any other information than formal properties of the expression at hand: c-command and identity of variables. No access to information beyond what is given in the expression resulting from the translation itself is necessary.

Clearly, as it has been known for many years (e.g. (Evans, 1980)) there are other linguistic dependencies that go beyond canonical variable binding. In fact, there are quite a few: donkey-anaphora, pay-check sentences, dependent readings of indefinites, etc. So, the question comes up whether it is justified to single out bound variable anaphora along the lines described here. Wouldn't it be more parsimonious to simply have one interpretive system taking care of this whole class of dependencies?

This is in fact the line taken in *Dynamic Semantics* (e.g. (Chierchia, 1995, 2001), (Elbourne, 2001) and others). In this type of approach, generalized interpretation functions cover a range of dependencies varying from bound pronouns to E-type pronouns, and dependent properties of indefinites, etc. We will henceforth refer to this class of approaches as "Generalized binding" approaches. Even intersentential connections as in (12) are brought under the same broad class of mechanisms.

(12) A bear is walking in the forest. He hums.

In such an analysis the connection between *a bear* and *he* is handled via λ-conversions that "dynamify" LF. Informally, one can say that the interpretation of the sentence containing *a bear* is enriched with the latter's context change potential (Chierchia, 1995). The effect is that the anaphoric expression *he* is brought under the scope of the indefinite. The classical Heim examples such as ****Every soldier** has a gun. Will* **he** *shoot?* where *he* cannot be

[12] We note in passing that discourse structure may be even more linguistic than generally assumed, given a recent discovery reported in (Cole *et al.*, 2000) that Chinese dialects spoken in Hong Kong vary in the discourse parameters involved in logophoricity.

brought under the scope of *every soldier* are accounted for in terms of scope restrictions on *every*.[13]

Chierchia conceives of his theory of dynamic binding as an explicit theory of discourse structure. Thus, the mechanisms enabling *he* in (12) to be brought under the scope of *a bear* constitute an explicit theory of how to form and access what one may informally call the discourse storage.

2.2 Binding and pseudo-binding

We have now prepared the ground for putting (8'), repeated below, in the proper perspective: What is the nature of the dependency observed, and how should it be accounted for?

(8') a. Every student_i turned in an assignment before that student_i went home.
b. *Before that student_i went home, every student_i turned in an assignment.

Note first that (8'a) really poses a puzzle by itself: Why is a "bound" reading possible at all? *That student* clearly depends for its interpretation on *every student*; but can it appropriately be called a variable? Cases like (8'a) are extensively discussed in (Noguchi, 1997). He argues that they do indeed instantiate bound variable anaphora. He assumes that D's such as *that* can be substituted by a variable: $[[_D\ \textit{that}]\ \textit{student}]] \rightarrow [[_D\ x]\ \textit{student}]]$. In the form it is given, Noguchi's proposal is not innocuous, however. The substitution of a variable for the demonstrative cannot be part of C_{HL}. First, it violates the inclusiveness condition: the variable cannot be part of the numeration. However, one could argue that this process is in fact not part of C_{HL} but of an interpretive process, hence not subject to inclusiveness (Noguchi appears to assume his proposal is part of syntax, but this is of course not crucial; nevertheless, some care is warranted). The second reason is that it violates the "Principle of Recoverability of Deletions" (PRD), since *that* has deictic force, and hence is more than just a bundle of phi-features. Note that this is a more fundamental problem. Relegating the substitution to an interpretive component does not help, since the PRD would still be violated.[14]

Even more importantly, in this form, his proposal doesn't help for (8'b): if $[_D\ x]$ is a variable in (8'a), why isn't it a variable in (8'b) that can be bound under the same conditions as the pronoun/variable in (1)? We think that this

---

[13] More precisely, conjunction/juxtaposition is treated in terms of function composition, with the restrictions on *no/every* coming out of the fact that negation cannot be defined in terms of function composition and that the same holds true for determiners, such as *every*, that can be defined in terms of negation (see (Delfitto, 2003) for precise discussion).

[14] As pointed out by Denis Delfitto (p.c) some caution may be required since fact-resuming anaphora like *that* in *A factory exploded. That caused great damage*, are not strictly deictic. Nevertheless, construing *that* as a just a bunch of phi-features would be overly simplistic for such cases as well.

contrast between (1) and (8') is in fact an important one which points toward fundamental differences in the way in which "interpretive dependencies" can be realized in language.

Thus, the relation between *every student* and *that student* cannot be one of bound variable binding. It appears to instantiate a rather different type of relation, which is in some sense similar to a binding-type dependency effected by more general discourse based principles. In order to express that, it is not standard variable binding, one may call it *pseudo-binding*.

First, we will deal with the empirical issues raised by (8'), then we will present a more thorough discussion of the theoretical issues it raises. In fact, we will see that if approached from the proper perspective, empirically (8') needs no special proviso at all. A simple assumption about incremental interpretation will do.

2.3 An incremental (left-to-right) procedure for anaphora resolution

What we need as a starting point is just the natural assumption that the interpretation of DPs in a sentence proceeds in a temporal/linear order from left to right. Thus, *every student* in (8'a) will first introduce a set of students.[15] In order to get the dependent reading, we need to assume no more than that *that student* is able to pick out any suitable individual from within that set. In particular, there is an assignment function *fs* that assigns to *that student* any arbitrary individual *a* from the set introduced by *every student*. If so, the sentence can be interpreted as verified, iff for any such *a* one may select from the set of students, *a* went home.[16] In that sense, *that student* receives a variable interpretation without having the status of a formally "bound" variable.[17]

Given that interpretation takes place incrementally from left to right, it follows that in (8'b) a dependency between *that student* and *every student* cannot be established. Without additional assumptions *that student* will be assigned a discourse referent as soon as it enters interpretation. If so, without

[15] In accordance with most of the literature, I will be assuming that it does, given the possibility to refer to such a set by a plural pronominal as in ***Every*** *soldier had a gun.* ***They*** *were dangerous*, although, as Denis Delfitto (p.c.) points out, the facts may not be entirely straightforward.

[16] Staying very informal, one could say that the interpretive effect obtains by instantiating the universal quantifier as free-choice *any*, with *that student* being interpreted as co-valued with the equivalent of *any student*.

[17] A similar usage is not possible with *the* and *this*. The question is why? We will be assuming that this usage is governed by accessibility in the sense of Ariel (1990). According to Ariel, *that* is a relatively low accessibility marker as compared to *the* and *this*. Apparently, the relevant value is not too readily accessible, and hence a relatively low accessibility marker like *that* is required. If so, this indicates that the semantic force of *that* is really necessary. As pointed out by Denis Delfitto (p.c.) it would be important to determine how precisely such differences are featurally encoded. Such an investigation would carry us beyond the scope of this study however.

back-tracking, it cannot pick out an arbitrary individual from within a set of students that is introduced later. In spirit, though not in execution, this is akin to a Condition C effect (Reinhart, 1983, 2000).

Not saying anything special about *that student* implies that something has to be said about *he* in (1b), repeated as (13) below, in order to accommodate the fact that it can be backward bound.

(13) Before he_i went home, every $student_i$ turned in an assignment.

Any account will have to rest on the fact that *he* is independently 'special', being a pure bundle of phi-features. The crucial stipulation, then, is that a true pronominal need not be interpreted immediately, given the fact that it lacks independent lexical content. That is, unlike *that student*, *he* can remain locally free and unvalued; or again, intuitively, put "into storage" for subsequent interpretation.[18]

The next step is to determine whether more than a stipulation is involved. This brings us to the more general issue of how interpretive dependencies are encoded.

3 Encoding interpretive dependencies

In linguistic theorizing, there is always a tension between two tendencies: one is the tendency of unification. From that perspective, a theory that can handle a broad spectrum of phenomena has a bonus. On the other hand, if certain types of phenomena are really different it is important that the theory expresses these differences. To the extent that a broad theory would have to deal with differences on the basis of stipulations and articulated sub-theories would express the differences naturally, theoretical differentiation gets the bonus. As already shown by the very existence of the contrast between (1) and (8'), there are limits to the possibility of identifying all interpretive dependencies. We now want to make the following point very explicitly: although, formally, canonical bound variable anaphora can be understood as a limiting case of generalized

---

[18] Hajime Hoji (p.c.) pointed out to us that this discussion is much in the spirit of Ueyama (1998). Ueyama distinguishes between two types of dependency. One, her "Formal dependency" requires c-command, and is very much akin to canonical bound variable anaphora. The other, "Indexical dependency", requires precedence, and is akin to E-type dependencies. Carrying out an extensive comparison would lead us beyond the scope of the present article. Note that it is crucial for this analysis that in the variety of English under discussion (and in Dutch), *that* is more than just a bundle of phi-features. That is, in addition to phi-features, it carries the lexical information distinguishing it from *the*, or *this*. This information is what prevents it from being translated as a true variable that can be stored, but requires a more round-about procedure resulting in 'pseudo-binding'. What we have to say about variants of English that allow backward 'binding' of *that student* is that in such variants the meaning of *that* has been subject to bleaching. Without proper investigation of such variants, their further properties and issues of semantic change it is hard to be more specific than that.

binding, within the language system it is handled by a different computational mechanism. From the perspective of an optimal organization of the language system this may look like an imperfection, but it is not. It is simply the application of form-based *syntactic* operations as part of SEM ***where applicable***. To put it differently, the types of operations allowed by C_{HL} are also available at the C-I interface. The C-I interface is, therefore, not a barrier for operations. If so, within the domain of interpretive principles one may expect certain cases in which there is competition between formal operations applying on expressions as they are entered from C_{HL} into the C-I interface, and operations that can only apply at a later interpretive stage after various enrichments have taken place. We will take it that generalized binding principles are of the latter type. Where form-based operations and generalized binding compete, the former win, just like variable binding wins over co-reference where they compete (Reinhart, 1983; Grodzinsky & Reinhart, 1993; Reinhart, 2000; Reuland, 2001). That is, there is a division of labour between components of the language system: even to that extent that if a syntactic strategy to an interpretive dependency is in principle available, and fails, you are not allowed to bypass this result by invoking an alternative strategy.[19] As we will see, there are intriguing restrictions on *backward binding* which can only be explained if one assumes that the choice between generalized binding and bound variable binding is governed by such principles. We will base our argument on a number of contrasts to be discussed in the next section.

4 Restrictions on backward binding

In this section, we will be discussing a number of contrasts that show up in cases of backward binding where the quantificational antecedent has no lexical restriction.

So far, our model for backward binding was provided by the case of (1b):

(1) a. Every student$_i$ turned in an assignment before he$_i$ went home.
    b. Before he$_i$ went home, every student$_i$ turned in an assignment.

However, in this particular case, the quantificational expression has a lexical restriction, namely *student*. If the restriction is dropped, the backward bound case is considerably less acceptable, as illustrated in (14):

(14) ?*Before he$_i$ went home, everyone$_i$ turned in an assignment.

Similar cases are equally degraded:

---

[19] As Reinhart puts it: "…if a certain interpretation is blocked by the computational system, you would not sneak in precisely the same interpretation for the given derivation by using machinery available for the systems of use."

(15)a. ?*Before he_i watches TV, $everyone_i$ eats an apple.
b. ?*Although/if he_i watches TV, $everyone_i$ eats an apple.

In Dutch and German, the following cases are fine without restriction, in contrast to English:

(16) *Voordat hij_i TV kijkt, eet $iedereen_i$ een appel.* (Dutch)
"before he watches TV, everyone eats an apple"
(17) *Bevor er_i fernsieht, isst $jeder_i$ einen Apfel.* (German)
"before he watches TV, everyone eats an apple"

Here, the subordinating conjunction is temporal. However, with non-temporal conjunctions, a clear contrast shows up between restricted and unrestricted quantification. Whereas the sentences in (18) and (19) are perfectly fine, those in (20) and (21) are once more heavily degraded:[20]

(18) *Hoewel/Indien hij_i TV kijkt, eet iedere $student_i$ een appel.* (Dutch)
"Although/if he watches TV, every student eats an apple"
(19) *Obwohl/Falls er_i fernsieht, isst jeder $student_i$ einen Apfel.* (German)
"Although/if he watches TV, every student eats an apple"
(20) *?*Hoewel/Indien hij_i TV kijkt, eet $iedereen_i$ een appel.* (Dutch)
"Although/if he watches TV, everyone eats an apple"
(21) *?*Obwohl/Falls er_i fernsieht, isst $jeder_i$ einen Apfel.* (German)
"Although/if he watches TV, everyone eats an apple"

Note that in all of these cases a referential post-cedent is fully accepted.[21]

The last contrast to be mentioned is that between backward binding of subject versus object pronominals in Dutch. Again, backward binding of an object pronominal by a restricted quantifier is fine, but a non-restricted quantifier is once more awkward:

(22)a. *Voordat Jan hem_i opgemerkt had, was iedere $student_i$ vertrokken.*
"Before John had noticed him, every student had left"
b. *?*Voordat Jan hem_i opgemerkt had, was $iedereen_i$ vertrokken.*
"Before John had noticed him, everyone had left"

The latter contrast is particularly interesting since it effectively rules out reconstruction of adverbial clauses as the instrument for backward binding.

[20] We are using the term 'degraded' rather than 'ungrammatical', since there is a certain amount of variation among speakers. However, all speakers find cases like (20) markedly worse than cases like (16) and (18).

[21] As pointed out by Denis Delfitto (p.c), the contrast between (1) and (15) does not show up in Italian with the corresponding control structures: *Prima di guardare la TV, ogni studente/ognuno mangiava una mela*. It would be interesting to investigate whether such a difference between control structures and finite clauses also shows up in the other languages investigated. This will have to wait for another occasion, however.

Besides reconstruction, it also rules out any approach in which binding obtains by just QR-ing *iedereen* so as to have scope over the adverbial clause.

Within the overall contrast between restricted and unrestricted quantifiers, we now have three sub-contrasts to account for: i) the contrast between English on the one hand, and Dutch and German on the other; ii) the contrast between temporal and non-temporal adverbial clauses in Dutch and German, and iii) the contrast between subject versus non-subject pronominals. In the next Section, we will discuss how these contrasts can be explained

5 Towards an explanation

As we noted at the beginning, one of the puzzles about backward binding is that it apparently avoids creating a weak crossover configuration. As already noted in (Chomsky, 1982), D-linking voids weak crossover effects. One could argue that restricted quantifiers can be D-linked, and unrestricted quantifiers cannot. But by itself this is not an explanation. It has never been made clear why D-linking has this effect. D-linking is just a name for the process more specifically described by current mechanisms of generalized binding yielding 'pseudo-binding' effects in our terms. And, as we noted, the general mechanisms of generalized binding should still apply to null- or virtually null-restrictions as their limiting case. Moreover, this cannot be all there is to say anyway, since subject/non-subject or temporal/non-temporal contrasts would be entirely unexpected. Weak crossover should rule out all backward binding by unrestricted quantifiers.

We will be claiming that there is another process involved in backward binding, which makes use of truly syntactic means. A dependency between elements in subordinate and matrix clause can be established through a syntactic dependency in the Tense system obviating WCO.

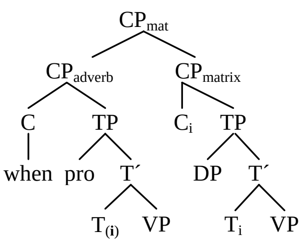
We will now discuss the necessary steps for the dependency to be derived. For concreteness sake, we will be assuming that the adverbial clause is adjoined to the matrix CP.[22]

In structure (23), the T's of matrix and adverbial clause do not c-command each other, but the C-position of the matrix clause c-commands both (assuming a category-segment distinction in which segments are c-commanded out of). We will assume that the matrix clause has a null C. Note that the presence of a C in the structure is necessary from the perspective of the theory of Enç (1997), where matrix C is the locus of Speech time. We know independently that C and

[22] We will abstract away from many details of the functional structure of the left periphery, since for current purposes such details do not appear to matter. We take CP for the sake of concreteness, but in fact any X/XP in the functional structure higher than TP would do.

T are able to share features. Shared features include not only Tense but also Agreement (see the rich literature on complementizer agreement, for instance (Haegeman, 1992)). We can abstract away from whether this feature sharing is implemented by Move or Agree. Let's simply take it that the possibility of feature sharing extends to C_{matrix} and T_{matrix}. Given this, there is a syntactic dependency between C_{matrix} and T_{matrix} indicated by the shared index *i* (which has no other status here than being shorthand for sharing of features). C_{matrix} does not only c-command T_{matrix}, it also c-commands T_{adverb}. Thus, configurational requirements for T_{adverb} to be part of a syntactic dependency are satisfied. What is further needed is the substantive assumption that T_{adverb} may indeed share features with C_{matrix} in temporal clauses.

(23)



Evidence for such a dependency comes from contrasts like the following:

(24) *Toen/nadat Jan de trein miste/gemist had/*gemist heeft, at hij een broodje.*
"When/after John the train missed/missed had/missed has, he ate a sandwich"

Note that, unlike in English, the present perfect *gemist heeft* is an unmarked expression of PAST in Dutch.[23] Yet, in this particular case, it is impossible. Instead, either a simple PAST or a Pluperfect has to be used. To put this in a proper perspective, note that languages differ in the way in which temporal dependencies are encoded. Russian, for instance, does not show what is known as *sequence of tenses*. In Russian, one says 'John said that he eats an apple', not 'John said that he ate an apple' if the saying and the eating are simultaneous. Similarly, Russian has no morphosyntactic expression of the Pluperfect. 'John ate' and 'John had eaten' are not necessarily formally distinguished. Yet, interpretively, simple past and pluperfect are as distinct in Russian as in English. What is different is not the interpretation, but whether or not certain interpretive dependencies between tenses are morpho-syntactically encoded. In English or Dutch they are, in Russian they are not. In this vein, we wish to claim that what is crucial about (24) is that a certain option to express the PAST

[23] *Jan heeft gisteren de trein gemist* 'Jan has yesterday missed the train' is the unmarked way of expressing PAST in Dutch even in the presence of a past time adverbial.

is blocked. There is no a priori reason why the present perfect form could not be used to express it. Rather, a morpho-syntactic dependency, which is, for instance, not found in Russian, is instrumental in blocking it (see Reuland & Avrutin, in prep. for further discussion of Russian).

If instead of a temporal conjunction, a concessive conjunction such as *hoewel* 'although', is used the contrast disappears:

(25) *Hoewel Jan de trein miste/gemist had/gemist heeft, at hij een broodje.*
"Although John the train missed/missed had/missed has, he ate a sandwich"

All three possibilities are realized, although with different interpretations. The one that concerns us here, with the present perfect, takes today's perspective on a past event. Although this reading is not easily accessible in (25), the explicit use of deictic temporal expressions brings out the contrast very clearly:

(26)a. **Toen/nadat Jan gisteren de trein gemist heeft, at hij vandaag toch gewoon een broodje.*
"When/after John yesterday the train missed has, he ate today yet simply a sandwich"
b. *Hoewel Jan gisteren de trein gemist heeft, at hij vandaag toch gewoon een broodje.*
"Although John yesterday the train missed has, he ate today yet simply a sandwich"

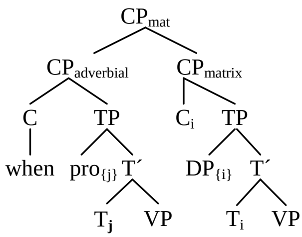
The conclusion is that in the case of temporal adverbs there is indeed a morpho-syntactic restriction on how PAST is expressed, prohibiting the present perfect, whereas in the case of concessive adverbs there is no such restriction. The present perfect is possible; the relevant interpretation can be brought out by a proper choice of indexicals. Thus, T_{adverb} does indeed share morpho-syntactic features with C/T_{matrix} in temporal clauses, but not in concessive clauses.[24]

Given this, let's consider the structure of (16) and (20) in more detail, using the structure of (23) as a starting point.

(16) *Voordat hij$_i$ TV kijkt, eet iedereen$_i$ een appel.*
"Before he watches TV, everyone eats an apple"
(20) *?*Hoewel/Indien hij$_i$ TV kijkt, eet iedereen$_i$ een appel.*
"Although/if he watches TV, everyone eats an apple"

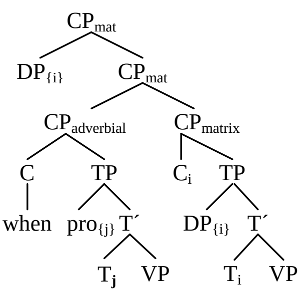
[24] A discussion of PAST conditionals with their intriguing properties would lead us too far for present purposes.

(27)



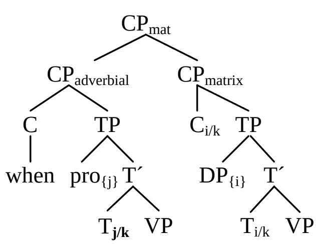
In order for the bound reading to be possible, the matrix subject DP should have scope over the pronominal subject of the adverbial clause. Let's assume for the sake of concreteness, that this is done by QR-ing DP and adjoining it to the CP root, as in (28).

(28)



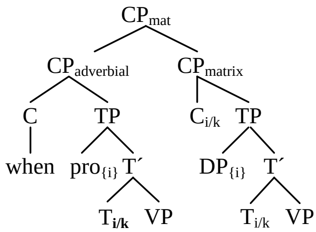
However, as noted earlier, the structure is expected to violate WCO. Observe, though, that as it is given, (27) does not represent the temporal dependency between adverbial and matrix T. Let's do so by adding an index *k*, as in (29).

(29)



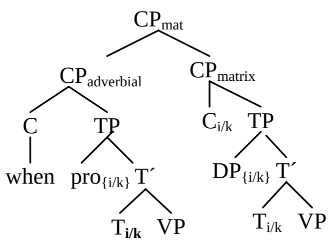
Since in principle it is possible for *hij* to be bound by *iedereen*, it is among the options that *iedereen* and *hij* have the same index. Implementing this option yields (30):

(30)



The last step is provided by the conception of nominative case presented by Pesetsky and Torrego (2001). They argue that nominative reflects an unvalued Tense feature on DPs. Thus, nominative assignment is tantamount to valuing this feature. Implementing this yields (31):

(31)



What we see now is that Spec-$T_{adverbial}$ (=pro) fully agrees (= shares all its phi-features) with $T_{adverbial}$. $T_{adverbial}$ agrees with C_{matrix} which in turn agrees with T_{matrix}. The latter agrees with Spec-TP= DP_{matrix}. Thus, linking all these agreement relations, there is full agreement between Spec-$T_{adverbial}$ (=pro) and Spec-TP= DP_{matrix}.

Given that variables are the translations at the C-I interface of bundles of phi-features, it follows that the variable resulting from *pro* and the variable resulting from QR-ing DP_{matrix} are fully identical. By itself, that does not guarantee binding but, slightly modifying Reinhart's theory of WCO (Reinhart, 1987), it does entail that QR-ing DP_{matrix} to a position where it has scope over *pro* does not give rise to a WCO violation, since the dependency can be read off the syntactic configuration independent of this movement.[25] This is precisely what we set out to derive.

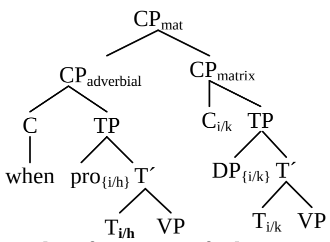
Note that this route is only open to bare quantifiers under the view developed in (Chomsky, 1995) and subsequent work that the restriction of a

[25] Reinhart (1987) argues that weak crossover violates the requirement that at S-structure the antecedent c-command the bound element from an A-position. The extension we would require is that at S-structure identity of index/variables is structurally encoded.

non-bare quantifier stays behind under A'-movement (only the quantificational part moves). Agree/head movement, on the other hand, can only affect the phi-features. Hence, this syntactic process cannot involve a DP with a non-null restriction.

The contrast between temporal and non-temporal adverbial clauses also follows. As we saw, in the case of *hoewel*, no temporal dependency is encoded. Hence the situation is as depicted in (32):

(32)



Here the features of the quantificational DP and pro fail to be fully identified, yielding a WCO violation if the DP attempts to bind the pronoun.

This derives both the contrast between temporal and non-temporal clauses, and the fact that non-subject pronominals in preposed clauses can never be backward bound, as shown by (22b) repeated here as (33):

(33) *?*Voordat Jan hem$_i$ opgemerkt had, was iedereen$_i$ vertrokken.*
"Before John had noticed him, everyone had left"

Clearly, since *hem* does not participate in the agreement with T, a syntactic dependency with *iedereen* cannot be formed.

How is the difference between temporal and non-temporal conjunctions encoded? Is it just their semantics that blocks a dependency, is it a difference in attachment site, or is it a combination of both (one property induces the other)? Here we cannot provide a final answer. Nevertheless, it is tempting to offer at least some speculation.

In a Cinque type hierarchy (Cinque, 1999), adverbs, and also adverbial clauses, have different attachment sites, temporal adverbs being lower in the structure than e.g. concessive adverbs. But this does not immediately transfer to preposed clauses without duplicating structure. However, if one takes the line of Nilsen (2003) that the interaction of adverbs with their environment, including their order, can in principle be derived from their semantic properties, duplicating structure is not necessarily implied. Since, independently, tense and phi-features go together (as on the finite verb), one may postulate that phi-features go up to a functional projection in the C-domain that relates to a temporal adjunct, but not to a "flavor" of C that relates to a concessive or conditional adjunct.[26] If so, the contrasts between backward binding in temporal and non-temporal adverbial clauses follow from independent properties of the left periphery of the clause.

Whereas this line of thought thus offers a further perspective on the variation in backward binding observed so far, it also offers a perspective on cross-linguistic variation of a more subtle nature. For instance, there are certain differences between Dutch and German, on the one hand, and English, on the other. In English even the counterpart of (16)/(17) is degraded:

(34) ?*Before he_i watches TV, $everyone_i$ eats an apple.

Under the approach developed here, it seems quite reasonable to relate this degradedness to the impoverished character of the English agreement system and the rudimentary status of the T-C dependency (as shown by the restricted appearance of V2). If one of the links cannot be established, there is no possibility for a WCO redemption. Yet, before firm conclusions can be drawn, other languages with impoverished agreement systems will have to be included in the investigation. Similarly, if a certain language formally encodes relations between T_{adverb} and higher up clauses by expressing mood (for instance in terms of subjunctive), one would accordingly not be very surprised if, in such a language, backward binding would be facilitated also for concessive and conditional clauses. All this will have to wait for other occasions, however.

6 Conclusion

We have seen that in a very restricted set of cases, bare quantifiers can backward bind a subject pronominal of a preposed clause. We have seen that this possibility is due to the fact that in these cases the dependency can be syntactically pre-encoded. We have also seen that the presence of a restriction on the quantifier makes binding possible throughout. This led us to an approach based on competition between subsystems.

It is crucial for this line of reasoning that the cases of 'backward binding' by a restricted quantifier we discussed are indeed not to be subsumed under canonical bound variable anaphora. And, in fact, the dependencies observed in these cases are subject to further restrictions that are incompatible with canonical BV status. For instance, replacing *iedere N* by *geen N*, or *every N* by *no N*, for that matter, makes establishing a dependency entirely impossible,

[26] I used the term "flavor of C" in analogy of what are called "flavors" of little v, leaving open whether there are just different values of C, or distinct functional categories in the left periphery (in so far as such a distinction makes sense at all).

providing further evidence that what we have here is really a different mechanism from canonical bound variable anaphora.[27]

(35)a. *Hoewel hij$_i$ TV keek, at geen student$_i$ een appel.
b. *Although he$_i$ was watching TV, no student$_i$ ate an apple.

To summarize the issue: intuitively it is reasonable to say that a bare quantifier will never be salient enough to establish a discourse set in the sense intended for D-linking. However, the moment one starts formalizing D-linking in terms of the general mechanisms of generalized dynamic binding, these appear rich enough to also make available E-type links also for bare quantifiers. Of course, it is possible to artificially incorporate the necessary restrictions by stipulation. However, ideally, the applicability of generalized binding mechanisms should be restricted in a principled manner. Here we propose that this can be achieved not by restricting generalized binding mechanisms, but by investigating how they compete with form-based syntactic operations where available. With bare quantifiers, variable binding can be brought about by form based syntactic operations manipulating phi-feature bundles, making copies, chains, etc. alone. But only under the conditions sketched. Binding by a bare quantificational antecedent thus reflects a purely structurally determined binding relation, as an instantiation of "logical syntax". Given the way labor is divided, where form based syntactic operations would in principle be available, but fail, they cannot be bypassed by a discourse based mechanisms. Thus, applying a discourse based mechanism treating a null restriction as a limiting case of a restriction will indeed be blocked by our conception of how labour is divided. It is an important question how far a purely form based logical syntax extends. This is yet another matter for further research.

References

Ariel, M. 1990. *Accessing Noun Phrase Antecedents*. London: Routledge, Croom Helm.
Avrutin, S. 1999. *Development of the Syntax-discourse Interface*. Dordrecht: Kluwer.
Chomsky, N. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
---------- 1982. *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, MA: MIT Press.
---------- 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
---------- 1998. "Minimalist Inquiries: The Framework." *MiTWPL 15*. Cambridge, MA: MIT Press.
---------- 1999. "Derivation by Phase." *MiTWPL 18*. Cambridge, MA: MIT Press.
Chierchia, G. 1995. *Dynamics of Meaning*. Chicago: Chicago University Press.

[27] Note that (i), which has a temporal connective, is quite OK. So, it is really the nature of the connection between main clause and subordinate clause that counts:
(i) *Terwijl hij$_i$ TV keek at niemand$_i$ een appel.*
"While he was watching TV no one ate an apple"

---------- 2001. “Long Distance Indefinites and Some of their Anaphoric Properties.” *GLOW Workshop*. Porto.

Cinque, G. 1999. *Adverbs and Functional Heads – A Cross-linguistic Perspective*. Oxford: Oxford University Press.

Cole, P., G. Hermon and C. Lee 2000. “Grammatical and Discourse Conditions on Long Distance Reflexives in Two Chinese Dialects.” P. Cole, G. Hermon and C. T. Huang (eds.) *Long-Distance Reflexives*, 1-44. *Syntax and Semantics* 33. New York: Academic Press.

Delfitto, D. 2003/to appear. “Results in Semantics.” *Lingue e Linguaggio* 2/2003.

Dowty, D. 1982. “Tenses, Time Adverbs and Compositional Semantic Theory.” *Linguistics and Philosophy* 5.23-55.

Elbourne, P. 2001. “A Unified Semantics for Pronouns.” *GLOW Workshop*. Porto.

Enç, M. 1987. “Anchoring Conditions for Tense.” *Linguistic Inquiry* 18.633-657.

Evans, G. 1980. “Pronouns.” *Linguistic Inquiry* 11.337-362.

Giorgi, A. and F. Pianesi 1991. “Toward a Syntax of Temporal Representations.” *Probus* 2.187-213.

---------- 1997. *Tense and Aspect. From Semantics to Morphosyntax*. New York: Oxford University Press.

Grodzinsky, Y. and T. Reinhart 1993. “The Innateness of Binding and Coreference.” *Linguistic Inquiry* 24.69-101.

Haegeman, L. 1992. *Theory and Description in Generative Syntax. A Case Study in West Flemish*. Cambridge: Cambridge University Press.

Hara, T. 2002. *Anaphoric Dependencies in Japanese*. Utrecht: LOT Dissertation Series.

Hatav, G. 1997. *The Semantics of Aspect and Modality: Evidence from English and Biblical Hebrew*. Amsterdam: Benjamins.

Heim, I. 1998/1993. “Anaphora and Semantic Interpretation: A Reinterpretation of Reinhart's Approach.” *SfS-Report-07-93*. Tübingen: University of Tübingen. Reprinted in: The Interpretive Tract, Uli. Sauerland and O. Percus (eds.). *MIT Working Papers in Linguistics* 25. Cambridge, MA: MIT Press.

Heny, F. 1982. “Tense, Aspect and Time Adverbials II.” *Linguistics and Philosophy* 5.109-154.

Higginbotham, J. 1983. “Logical Form, Binding and Nominals.” *Linguistic Inquiry* 14(3).395-421.

Hornstein, N. 1977. “Towards a Theory of Tense.” *Linguistic Inquiry* 8.521-557.

---------- 1990. *As Time Goes By*. Cambridge, MA: MIT Press.

Kamp, H. and U. Reyle 1993. *From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Dordrecht: Kluwer Academic Publishers.

Kayne, R. 2001. *Pronouns and their Antecedents*. Ms. NYU.

Lakoff, G. 1968. *Pronouns and Reference*. Bloomington, Indiana: Indiana University Linguistics Club.

Langacker, R. 1966. “On Pronominalization and the Chain of Command.” D. Reibel and S. Schane (eds.), *Modern Studies in English*, 160-186.

Nilsen, Ø. 2003. *Eliminating Positions: Syntax and Semantics of Sentential Modification*. Utrecht, UiL OTS: LOT Dissertation Series.

Noguchi, T. 1997. “Two Types of Pronouns and Variable Binding.” *Language* 73.770-797.

Partee, B. 1973. “Some Structural Analogies Between Tenses and Pronouns in English.” *The Journal of Philosophy* LXX, N0. 18.

Pesetsky, D. and E. Torrego 2001. “T-to-C Movement: Causes and Consequences.” M. Kenstowicz (ed.), *Ken Hale: A Life in Language*. Cambridge, MA: MIT Press.
Reinhart, T. 1983. *Anaphora and Semantic Interpretation*. London: Croom Helm.
---------- 1987. “Specifier and Operator Binding.” E. Reuland and A. ter Meulen (eds.) *The Representation of (In)definiteness*. Cambridge, MA: MIT Press.
---------- 2000. “Reference Time.” Talk given at the Language in Use Symposium. Utrecht University: UiL OTS.
Reuland, E. 2001. “Primitives of Binding.” *Linguistic Inquiry* 32.439-492.
---------- 2003. “Anaphoric Dependencies: A Window into the Architecture of the Language System.” State-of-the-article. GLOT International 7.1/2., 2-25.
---------- and S. Avrutin. In prep. *Backward Anaphora and Tense Interpretation*. Ms. Utrecht: UiL OTS.
Stowell, T. 1995. “The Phrase Structure of Tense.” J. Rooryck and L. Zaring (eds.) *Phrase Structure and the Lexicon*. Dordrecht: Kluwer.
Ueyama, A. 1998. *Two Types of Dependency*. Doctoral Dissertation. University of Southern California. Los Angeles: GSIL Publications, USC.
Zribi-Hertz, A. 1989. “A-type Binding and Narrative Point of View.” *Language* 65.695-727.

Modelling Referential Choice in Discourse: A Cognitive Calculative Approach and a Neural Network Approach*

André Grüning[1] and Andrej A. Kibrik[2]

[1]*Max Planck Institute for Mathematics in Sciences*

[2]*Institute of Linguistics, Russian Academy of Sciences*

In this paper, we discuss referential choice – the process of referential device selection made by the speaker in the course of discourse production. We aim at explaining the actual referential choices attested in the discourse sample. Two alternative models of referential choice are discussed. The first approach of Kibrik (1996, 1999, 2000) is the **cognitive calculative** approach. It suggests that referential choice depends on the referent's current activation score in the speaker's working memory. The activation score can be calculated as a sum of numeric contributions of individual activation factors, such as distance to the antecedent, protagonisthood, and the like. Thus, a predictive dependency between the activation factors and referential choice is proposed in this approach. This approach is cognitively motivated and allows one to offer generalization about the cognitive system of working memory. The calculative approach, however, cannot address non-linear interdependencies between different factors. For this reason we developed a mathematically more sophisticated **neural network** approach to the same set of data. We trained feed-forward networks on the data. They classified up to all but 4 instances correctly with respect to the actual referential choice. A pruning procedure allowed to produce a minimal network and revealed that out of ten input factors five were sufficient to predict the data almost correctly, and that the logical structure of the remaining factors can be simplified. This is a pilot study necessary for the preparation of a larger neural network-based study.

1 Introduction

We approach the phenomena of discourse reference as a realization of the process of *referential choice*: every time the speaker needs to mention a referent s/he has a variety of options at his/her disposal, such as full NPs, demonstratives, third person pronouns, etc. The speaker *chooses* one of these options according to certain rules that are a part of the language production system. Production-oriented accounts of reference are rarer in the literature than comprehension-oriented; for some examples see (Dale, 1992; Strube & Wolters, 2000).

Linguistic studies of referential choice often suffer from circularity: for example, a pronominal usage is explained by the referent's high activation,

---

* This article results from two papers delivered at DAARC: the talk by Kibrik at DAARC-2000 in Lancaster, and the joint talk by Grüning and Kibrik at DAARC-2002 in Lisbon. Andrej Kibrik's research has been supported by grant 03-06-80241 of the Russian Foundation for Basic Research.

while the referent is assumed to be highly activated because it is actually coded by a pronoun in discourse. In a series of studies by Kibrik (1996, 1999, 2000), an attempt to break such circularity was undertaken. The main methodological idea is that we need an account of referent activation that is entirely independent of the actual referential choices observed in actual discourse. There are a variety of linguistic factors that determine a referent's current activation, and once the level of activation is determined, the referential option(s) can be predicted with a high degree of certainty. This approach includes a quantitative component that models the interaction of activation factors yielding the summary activation of a referent. As it will be explained below, the contributions of individual factors are simply summed, and for this reason we use the shorthand *cognitive calculative approach*. This approach is outlined in Section 2 of this paper.

The cognitive calculative approach, however, has some shortcomings; in particular, its arithmetic nature could not allow addressing non-linear interaction between different factors. It is for this reason that we propose an alternative approach based on the mathematical apparatus of *neural networks.* In Section 3, computer simulations are reported in which we attempt to find out whether neural networks can help us to overcome some shortcomings of Kibrik's original approach. As the available data set is quite small (102 items) and large annotated corpora are not so easily obtained, we decided to design this study as a pilot study, rather than putting weight on statistical rigor.

2 The cognitive calculative approach

2.1 General assumptions underlying the cognitive calculative approach

In this paper, we approach discourse anaphora from the perspective of a broader process that we term referential device selection or, more simply, *referential choice*. This term differs from "discourse anaphora" in the following respects.

1) The notion of "referential choice" emphasizes the dynamic, procedural nature of reference in discourse. In addition, it is overtly production oriented: referential choice is the process performed by the speaker/writer. In the course of each act of referential choice, the speaker chooses a formal device to code the referent s/he has in mind. In contrast, "anaphora" is usually understood as a more static textual phenomenon, as a relationship between two or more segments of text.

2) Unlike "discourse anaphora", "referential choice" does not exclude introductory mentions of referents and other mentions that are not based on already-high activation of the referent.

3) The notion of referential choice permits one to avoid the dispute on whether “anaphora” is restricted to specialized formal devices (such as pronouns) or has a purely functional definition.

These three considerations explain our preference for the notion of referential choice. Otherwise the two notions are fairly close in their denotation.

A number of general requirements towards the cognitive calculative approach to referential choice were adopted from the outset of the study. The model must be:

(i) speaker-oriented: referential choice is viewed as a part of language production performed by the speaker;
(ii) sample-based: the data for the study is a sample of natural discourse, rather than heterogeneous examples from different sources;
(iii) general: all occurrences of referential devices in sample must be accounted for;
(iv) closed: the proposed list of factors cannot be supplemented to account for exceptions;
(v) predictive: the proposed list of factors aims at predicting referential choice with maximally attainable certainty;
(vi) explanatory and cognitively based: it is claimed that this approach models the actual cognitive processes, rather than relying on a black box ideology;
(vii) multi-factorial: potential multiplicity of factors determining referential choice is recognized; each factor must be monitored in each case, rather than in an ad hoc manner, and the issue of interaction between various relevant factors must be addressed;
(viii) calculative: contributions of activation factors are numerically characterized;
(ix) testable: all components of this approach are subject to verification;
(x) non-circular: factors must be identified independently of the actual referential choice.

2.2 The cognitive model

Now, a set of more specific assumptions on how referential choice works at the cognitive level is in order. Recently a number of studies have appeared suggesting that referential choice is directly related to the more general cognitive domain of working memory and the process of *activation in working memory* (Chafe, 1994; Tomlin & Pu, 1991; Givón, 1995; Cornish, 1999; Kibrik, 1991, 1996, 1999). For cognitive psychological and neurophysiological accounts of working memory, see (Baddeley, 1986, 1990; Anderson, 1990; Cowan, 1995; Posner & Raichle, 1994; Smith & Jonides, 1997). The claim that referential choice is governed by memory processes is compatible with psycholinguistic frameworks of such authors as Gernsbacher (1990), Clifton and Ferreira (1987), Vonk *et al.* (1992), with the cognitively-oriented approaches of the Topic continuity research (Givón, 1983), Accessibility theory (Ariel, 1990), Centering theory (Gordon *et al.*, 1993), Givenness hierarchy

(Gundel *et al.*, 1993), and Cognitive grammar (van Hoek, 1997), as well as with some computational models covered in (Botley & McEnery, 2000). Thus, the first element of the cognitive model can be formulated as follows:

- ➢ The primary cognitive determiner of referential choice is activation of the referent in question in the speaker's working memory (henceforth: WM).

Activation is a matter of degree. Some chunks of information are more central in WM while some others are more peripheral. The term *activation score* (AS) is used here to refer to the current referent's level of centrality in the working memory. AS can vary within a certain range – from a minimal to a maximal value. This range is not continuous in the sense that there are certain important thresholds in it. When the referent's current AS is high, semantically *reduced* referential devices, such as pronouns and zeroes, are used. On the other hand, when the AS is low, semantically *full* devices such as full NPs are used. Thus the second basic idea of the cognitive model proposed here is the following.

- ➢ If AS is above a certain *threshold*, then a semantically reduced (pronoun or zero) reference is possible, and if not, a full NP is used.

Thus at any given moment in discourse any given referent has a certain AS. The claim is that AS depends on a whole gamut of various factors that can essentially be grouped in two main classes:

- *properties of the referent* (such as the referent's animacy and centrality)
- *properties of the previous discourse* (distance to the antecedent, the antecedent's syntactic and semantic status, paragraph boundaries, etc.)

These factors are specified below in Sections 2.3 and 2.4. Now the third basic point of the model can be formulated:

- ➢ At any given point of discourse *all relevant factors interact* with each other, and give rise to the integral characterization of the given referent (AS) with respect to its current position in the speaker's WM.

In other words, such oft-cited factors of referential choice as distance to the antecedent, referent centrality, etc., affect the referential choice not directly but through the mediation of the speaker's cognitive system, specifically, his/her WM. Therefore, these factors can be called *activation factors*.

The actual cognitive on-line process of referential choice is a bit more complex than is suggested by the three postulates formulated above. Some work on referential choice (see e.g. (Kibrik, 1991)) has been devoted to the issue of ambiguity of reduced referential devices. In the process of referential choice, a normal speaker filters out those referential options that can create ambiguity, or *referential conflict*. Thus, it is possible that even in case of high activation of a referent, a reduced referential device is still ruled out. The referential conflict filter is outside of the focus of this paper, but consider one illustrative example from the Russian story discussed in the following section, in an English translation.

(1) The mechanic started, but immediately returned – *he* began to dig in the box of instruments; they were lying in their places, in full order. *He* pulled out one wrench, dropped it, shook *his* head, whispered something and reached in again. Fedorchuk now clearly saw that the mechanic was a coward and would never go out to the wing. The pilot angrily poked **the mechanic** at the helmet with his fist <…>

The referent of interest here is “the mechanic”; all of its mentions are underlined, and the pronominal mentions are also italicised. The point in question is the boldfaced mention of this referent. “The mechanic” is very highly activated at this point (see Section 2.3 below), therefore, the pronominal mention ***him*** can be expected here. However, in the Russian original text (as well as in its English translation) such pronominal mention does not really fit. The reason is that, in spite of the extremely high activation of the referent, there is also at least one other referent, “Fedorchuk”, that is equally activated and therefore can be assumed by the addressee to be the referent of the pronoun. Using a pronoun to refer to “the mechanic” would cause a referential conflict. Normally speakers/writers filter out the instances of potential referential conflict, by using disambiguation devices – from gender-specific pronouns to full NPs, as in example (1) (for details see (Kibrik, 1991, 2001)).

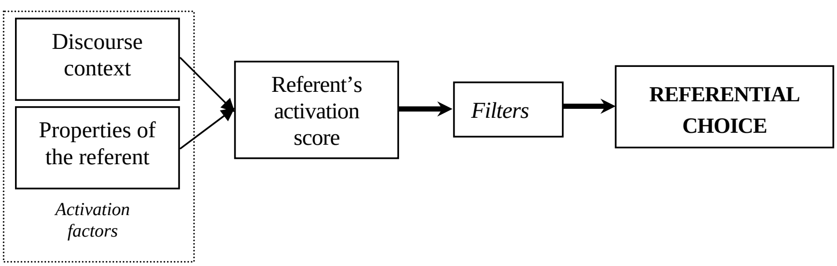


Figure 1: The cognitive multifactorial model of reference in discourse production

The cognitive model outlined above is summarized in the chart in Figure 1. The “filters” component implies, in the first place, the referential conflict filter, as well as some other filters, see (Kibrik, 1999).

This cognitive model is proposed here not only in a declarative way; there is also a mathematical, or at least quantitative, or *calculative* component to it. Each activation factor is postulated to have a certain numeric weight that reflects its relative contribution to the integral AS value. The general model of referential choice outlined above is assumed to be universal but the set of activation factors, especially their relative numeric weights, and thresholds in the AS range are language-specific. In this article, two studies are reported that have been conducted for Russian (Section 2.3) and English (Section 2.4) written narrative discourse, with the explanation of the quantitative component of the model.

Both of the presented studies are based on small datasets, especially by standards of modern computational and corpus linguistics. However, it must be made clear that the original purport of these studies was of theoretical, rather than of computational, character: to overcome two major stumbling blocks common for the studies of reference. To reiterate, these two stumbling blocks are:

- *circularity*: Referential choice is explained by the level of activation (or another quasi-synonymous status), and the judgment on the level of activation is obtained from the actual referential form employed
- *multiplicity of factors*: Suppose factor A is of central importance in instance X, and factor B in instance Y. It often remains unclear what, if any, is the role of factor A in instance Y, and of factor B in instance X.

So, the goal of the proposed approach is to explore the following issue: is it possible to construct a system of activation factors that, first, are determined independently of actual referential choice, and, second, predict and explain referential choice in a cognitively plausible way?

As it will become clear from the exposition of the calculative component of this approach, this component is extremely time- and effort-consuming, and inherently must have been restricted to a small dataset. We believe this does not call into question the theoretical result: a system of interacting activation factors can indeed be constructed.

2.3 The Russian study

In this study (for details, see (Kibrik, 1996)), a single sample of narrative prose was investigated – a short story by the Russian writer Boris Zhitkov “Nad vodoj” (“Over the water”). This particular sample discourse was selected for this study because narrative prose is one of the most basic discourse types,[1] because written prose is a well-controlled mode in the sense that previous discourse is the only source for the recurring referents, and because Boris

---

[1] There is an unresolved debate in reference studies on whether referential processes are genre-dependent. Fox (1987a) proposed two different systems of referential choice, depending on discourse type. Toole (1996) has argued that the factors of referential choice are genre-independent. We do not address this issue in this article, but assume that in any case referential choice in narrative discourse must be close to the very nuclear patterns of reference, since narration is among the basic functions of language, is attested universally in all languages and cultures, and provides a very favorable environment for recurrent mention of referents in successive discourse units.

Zhitkov is an excellent master of style, with a very simple and clear language, well-motivated lexical choices, and at the same time with a neutral, non-exotic way of writing. This specific story is a prototypical narrative describing primarily basic events – physical events, interactions of people, people's reflections, sentiments, and speech. The story is written in the third person, so there are no numerous references to the narrator.

The sample discourse comprised about 300 discourse units (roughly, clauses). There are about 500 mentions of various referents in the sample, and there are some 70 different referents appearing in the discourse. However, only a minority of them occurs more than once. There are 25 referents appearing at least once in an anaphoric context, that is in a situation where at least a certain degree of activation can be expected.

The fundamental opposition in Russian referential choice is between full NPs and the third person pronoun *on*. Discourse-conditioned referential zeroes are also important, but they are rarer than *on* (for further details, see (Kibrik, 1996)).

Several textual factors have been suggested in the literature as directly determining the choice of referential device. Givón (1983, 1990) suggested that *linear distance* from an anaphor to the antecedent is at least one of the major predictors of referential choice. Givón measured linear distance in terms of clauses, and that principle turned out to be very productive and viable. In many later studies, including this one, discourse microstructure is viewed as a network of discourse units essentially coinciding with clauses.[2]

Fox (1987a: Ch. 5) argued that it is the rhetorical, hierarchical structure of discourse rather than plain linear structure that affects selection of referential devices. Fox counted *rhetorical distance* to the antecedent on the basis of a rhetorical structure constructed for a text in accordance with the Rhetorical Structure Theory (RST), as developed by Mann and Thompson (Mann *et al.*, 1992). According to RST, each discourse unit (normally a clause) is connected to at least one other discourse unit by means of a rhetorical relation, and via it, ultimately, to any other discourse unit. There exists a limited (although extensible) inventory of rhetorical relations, such as joint, sequence, cause, elaboration, etc. In terms of RST, each text can be represented as a tree graph consisting of nodes (discourse units) and connections (rhetorical relations). Rhetorical distance between nodes A and B is then the number of horizontal steps one needs to make to reach A from B along the graph (one example of a rhetorical graph is shown below in Section 2.4). Fox was correct in suggesting that rhetorical distance measurement is a much more powerful tool for

[2] There are certain reservations regarding this coincidence, but they are irrelevant for this paper.

modelling reference than linear distance. However, linear distance also plays its role, though a more modest one.

In a number of works, it was suggested that a crucial factor of referential choice is episodic structure, especially in narratives. Marslen-Wilson *et al.* (1982), Tomlin (1987), and Fox (1987b) have all demonstrated, though using very different methodologies, that an episode/paragraph boundary is a borderline after which speakers tend to use full NPs even if the referent was recently mentioned. Thus, one can posit the third type of distance measurement – *paragraph distance*, measured as the number of paragraph boundaries between the point in question and the antecedent.

One more factor was emphasized in (Grimes, 1978) – the centrality of a referent in discourse, which we call *protagonisthood* below. For a discussion of how to measure a referent's centrality, see (Givón, 1990: 907-909).

Several other factors have been suggested in the literature, including animacy, syntactic and semantic roles played by the NP/referent and by the antecedent, distance to the antecedent measured in full sentences, and the referential status of the antecedent (full/reduced NP). Some of these factors will be discussed in greater detail in Section 2.4 below, in connection with the English data.

From the maximal list of potentially significant activation factors we picked a subset of those that proved actually *significant* for Russian narrative prose. The criterion used is as follows. Each factor can be realized in a number of values. For example, a distance factor may have values 1, 2, etc. Each potentially significant factor has a "privileged" value that presumably correlates with the more reduced form of reference. For example, for the linear distance to the antecedent, it is the value of "1", while for the factor of the antecedent's syntactic role it is "subject". Only those potential factors whose privileged value demonstrated a high co-occurrence (in at least 2/3 of all cases) with the reduced form of reference have been considered significant activation factors. For example, the factor of rhetorical distance patterns vis-à-vis pronouns and full NPs in a nearly mirror image way: there is a high co-occurrence of the value of 1 with pronominal reference (91%), and a high co-occurrence of rhetorical distance greater than 1 (79%) with full NP reference.

On the other hand, other potential factors did not display any significant co-occurrence with referential choice. In particular, the parameter of referential type of the antecedent does not correlate at all with the referent's current pronominalizability: for instance, a 3rd person pronoun is the antecedent of 10% of all 3rd person pronouns and 13% of full NPs, which makes no significance difference.

Seven significant activation factors have been detected. This is their list with the indication [in brackets] of the privileged value co-occurring with pronominal reference: animacy [human], protagonisthood [yes], linear distance [1], rhetorical distance [1], paragraph distance [0], syntactic [subject] and semantic [Actor[3]] roles of the antecedent, and sloppy identity[4].

After the set of significant activation factors had been identified, certain numeric weights have been assigned to their values. Variation of referents' AS from 0 to 1 was postulated. The activation factor weights take discrete values measured in steps of size 0.1. In each particular case, the weights of all involved factors can be summed and the resulting activation score is supposed to predict referential choice.

Table 1 below lists a selection of activation factors. Each factor is presented with the values it can accept and the corresponding numeric weights.

Activation factor	*Value*	*Numeric activation weight*
Rhetorical distance to the antecedent	1	0.7
	2	0.4
	3	0
	4+	–0.3
Paragraph distance to the antecedent	0	0
	1	–0.2
	2+	–0.4
Protagonisthood	Yes, and the current mention is:	
	the 1st mention in a series	0.3
	the 2nd mention in a series	0.1
	otherwise	0
	No	0

Table 1: Examples of activation factors, their values, and numeric weights

Activation factors differ regarding their logical structure. Some factors are sources of activation. The strongest among these is the factor of rhetorical distance to the antecedent. The closer the rhetorical antecedent is, the higher is the activation.

[3] The term "Actor" is an abstract semantic macrorole; it designates the semantically central participant of a clause, with more-than-one-place verbs usually agent or experiencer; see e.g. (Van Valin, 1993:43ff).

[4] The factor of sloppy identity occurs when two expressions are referentially close, but not identical. In the following example from the story under investigation, given in a nearly literal English translation, the first expression is referentially specific, and the second (*it*) generic:

(i) He understood that the engine skipped, that probably the carburetor had gotten clogged (through *it* gas gets into an engine) <…>

Sloppy identity is relevant in far fewer cases than other factors, and for this reason it can be called a second-order, or "weak", factor. Sloppy identity slightly reduces activation of a referent that has an antecedent, but a sloppy one.

The factor of paragraph distance is never a source of activation; vice versa, it is, so to speak, a penalizing factor. In the default situation, when the antecedent is in the same paragraph (paragraph distance = 0), this factor does not contribute to AS at all. When the antecedent is separated from the current point in discourse by one or more paragraph boundaries, the activation is lowered.

The third factor illustrated in Table 1, protagonisthood, has a still different logical structure. It can be called a compensating factor. It can only add activation, but does that in very special situations. When a referent is not a protagonist, this factor does not affect activation. If a referent is a protagonist, this factor helps to regain activation at the beginning of a series,[5] that is, in the situation of lowered activation. If the activation is high anyway, this factor does not matter.

The numeric weights such as those in Table 1 were obtained through a heuristic procedure of trials and errors. After several dozen of successive adjusting trials the numeric system turned out to predict a subset of referential choices correctly: reduced referential forms were getting ASs close to 1, and full NPs were getting ASs much closer to 0. When this was finally achieved, it turned out that all other occurrences of referential devices are properly predicted by this set of numeric weights without any further adjustment. It is worth pointing out that such trial-and-error procedure, performed by hand, is extremely time- and labour-consuming, even provided that the dataset was relatively small. The difference of this approach from the prior approaches is that the full control of the dataset, whatever size it has, has been gained.

After the calculative model was completely adjusted to the data of the Zhitkov's story, it was tested on a different narrative – a fragment of Fazil' Iskander's story "Stalin and Vuchetich", about 100 discourse units long. The result was that the model predicted all referential choices in the test dataset, without further adjustment (with the exception of minor adjustment in the numerical weights of two activation factors). These facts can be taken as evidence suggesting that the developed system does model actual referential choice in written narratives closely enough.

One more crucial point needs to be made about this model. When one observes actual referential choices in actual discourse, one can only see the ready results of referential device selection by the author – full NPs, pronouns, or zeroes. However, the real variety of devices is somewhat greater. It is important to distinguish between the *categorical* and *potentially alternating*

[5] The notion of "series" means a sequence of consecutive discourse units, such that: (i) all of them mention the referent in question, and (ii) the sequence is preceded by at least three consecutive discourse units not mentioning the referent.

referential choices. For example, the pronoun *on* in a certain context may be the only available option, while in another context it could well be replaced by an equally good referential option, say a full NP. These are two different classes of situations, and they correspond to two different levels of referent activation. The referential strategies formulated in (Kibrik, 1996) for Russian narrative discourse are based on this observation. Those referential strategies shown in Table 2 below represent the mapping of different AS levels onto possible referential choices.

Referential device:	Full NP only	Full NP most likely, pronoun /zero unlikely	Either full NP or pronoun/zero	Pronoun/zero only
AS:	0–0.3	0.4–0.6	0.7–0.9	1

Table 2: Referential strategies in Russian narrative discourse

What governs the speaker's referential choice when the AS is within the interval of the activation scale that allows variable referential devices (especially 0.7 through 0.9)? We do not have a definitive answer to this question at present. The choice may depend on idiolect, on discourse type and genre, or perhaps even be random. On the other hand, there may be some additional, extra-weak, factors that come into play in such situations.

2.4 The English study

The model developed for Russian narrative discourse was subsequently applied to a sample of English narrative discourse, which required a fair amount of modification. This study was described in (Kibrik, 1999), and its main results are reported here, along with some additional details. The sample (or small corpus) was the children's story "The Maggie B." by Irene Haas. There are 117 discourse units in it. 76 different *referents* are mentioned in it, not counting 13 other mentioned in the quoted songs. There are 225 referent *mentions* in the discourse (not counting those in quoted text). There are 14 different referents mentioned in discourse that are *important* for this study. They are those mentioned at least once in a context where any degree of activation can be possibly expected. Among the important referents, there are three protagonist referents: "Margaret" (72 mentions altogether), "James" (28 mentions), and "the ship" (12 mentions). An excerpt from the sample discourse, namely lines 1401–2104, is given in the Appendix below.

Any referent, including an important referent, can be mentioned in different ways, some of which (for example, first person pronouns in quoted speech) are irrelevant for this study. Those that are *relevant* for this study fall into two large formal classes: references by full NPs and references by activation-based pronouns. "Activation-based pronouns" means the unmarked, general type of pronoun occurrences that cannot be accounted for by means of any kind of syntactic rules, in particular, for the simple reason that they often appear in a different sentence than their antecedents. In order to explain and predict this kind of pronoun occurrence, it is necessary to construct a system of the type described in Section 2.3, taking into account a variety of factors related to discourse context and referents' properties. Typical examples of activation-based pronouns are given in (2) below.[6]

(2) 1607 Lightning split the sky
1608 as *she* ran into the cabin
1609 and slammed the door against the wet wind.
1610 Now everything was safe and secure.
1701 When *she* lit the lamps,
1702 the cabin was bright and warm.

There are two occurrences of the activation-based pronoun *she* in (2), and the second one is even used across the paragraph boundary from its antecedent. Besides the activation-based 3rd person pronouns, there are two dozen occurrences of syntactic pronouns that can potentially be accounted for in terms of simpler syntactic rules. At the same time, the activation-based principles outlined here can easily account for syntactic pronouns, see (Kibrik, 1999).[7]

Thus, the focus of this study was restricted to 39 full NP references and 40 activation-based pronominal references. As it was pointed out in Section 2.3 above, within each of the referential types – full NPs and pronouns – there is a crucial difference: whether the referential form in question has *an alternative*. In (3) below, an illustration of a pronoun usage is given that can vary with a full NP: in unit 1601, the full NP Margaret could well be used (especially provided that there is a paragraph boundary in front of unit 1601).

(3) 1502 A storm was coming!
1503 Margaret must make the boat ready at once.
1601 *She* took in the sail
1602 and tied it tight.

Contrariwise, there are instances of categorical pronouns. Consider (4), which is a direct continuation of (3):

---

[6] In the examples, as well as in Appendix 1, each line represents one discourse unit. In line numbers the first two digits refer to the paragraph number in the story, and the last two digits to the number of the discourse unit within the current paragraph.

[7] For an example of a syntactic pronoun cf. one sentence from the story under investigation (see Appendix, lines 1601-1602):

(ii) She took in the sail and tied *it* tight.

Pronouns occurrences such as *it* in this example can be accounted for by means of syntactic rules that are lighter, in some sense, than the activation-based procedure of referential choice described here. For an example of a generalized treatment of activation-based and syntactic referential devices see Section 3 of this article.

(4) 1603 She dropped the anchor
1604 and stowed all the gear <...>

In 1603, it would be impossible to use the full NP Margaret; only a pronoun is appropriate.

For the English data, it was found that referential forms of each type (for example, pronouns) fall into three categories: those allowing no alternative (= categorical), those allowing a questionable alternative, and those allowing a clear alternative. Thus, there are six possible correspondences between the five potential types and two actual realizations; see Table 3.

Potential referential form	Full NP only	Full NP, $^{?}$pronoun	Full NP or pronoun		Pronoun, $^{?}$full NP	Pronoun only
Frequency	*15*	*17*	*7*	*15*	*18*	*7*
Actual referential form	Full NP (*39*)			Pronoun (*40*)		

Table 3: Actual and potential referential forms, and their frequencies in sample discourse

The information about referential alternatives is crucial for establishing referential strategies. Of course, attribution of particular cases to one of the categories is not straightforward. It must be noted that such attribution is the second extremely laborious procedure involved in this kind of study (along with the search for optimal numerical weights of activation factors). To do this attribution properly, a significant number of native speakers must be consulted. There were two sources of information on referential alternatives used in this study: (i) an expert who was a linguist and a native speaker of English and had a full understanding of the problem and the research method, and who supplied her intuitive judgments on all thinkable referential alternatives in all relevant points of discourse; (ii) a group of 12 students, native speakers of English, who judged the felicity of a wide variety of modifications of the original referential choices through a complicated experimental procedure. These two kinds of data were brought together and gave rise to an integral judgment for each referential alternative. The details of this part of the study are reported in (Kibrik, 1999). In the end, all referential alternatives were classified as either appropriate, questionable, or inappropriate – see Table 4 below. The attribution of referential alternatives to categories is an indispensable component of this study, since the two formal categories "pronoun" vs. "full NP" are far too rough to account for the actual fluidity of referential choice.

The six strongest activation factors that were found to be most important in modelling the data of the sample discourse are the following: rhetorical distance to the antecedent (RhD), linear distance to the antecedent (LinD), paragraph distance to the antecedent (ParaD), syntactic role of the linear antecedent,[8] animacy, and protagonisthood. The first three of these factors are different measurements of the distance from the point in question to the antecedent. By far, the most influential among the distance factors, and in fact among all activation factors, is the factor of *rhetorical distance*: it can add up to 0.7 to the activation score of a referent. *Linear* and *paragraph distances* can only penalize a referent for activation; this happens if the distance to the antecedent is too high. To see how rhetorical (hierarchical) structure of discourse can be distinct from its linear structure, consider the rhetorical graph in Figure 2.[9]

Rhetorical distance is counted as the number of horizontal steps required in order to reach the antecedent's discourse unit from the current discourse unit. For a simple example, consider the pronoun *him* in discourse unit 1802. It has its antecedent James in discourse unit 1801. There is one horizontal step from 1802 to the left to 1801, hence RhD = 1. The pronoun *they* in 2004 has its antecedent Margaret and James in 2001. In order to reach 2001 from 2004, one needs to make two horizontal steps along the tree leftwards: 2004 to 2002 and 2002 to 2001. To visualize this more clearly, it is useful to collapse the fragment of the tree onto one linear dimension, see Figure 3. Thus RhD = 2.

In narratives, the fundamental rhetorical relation is that of sequence. Three paragraphs of the four depicted in Figure 2 (#18, #20, and #21) are connected by this relation, and within each of these paragraphs there are sequenced discourse units, too. If there were no other rhetorical relations in narrative besides sequence, rhetorical distance would always equal linear distance. However, this is not the case. In the example analysed, one paragraph, namely #19, is off the main narrative line. It provides the background scene against which the mainline events take place. Likewise, discourse unit 1904 reports a result of what is reported in 1903. The difference between the linear and the rhetorical distance can best be shown by the example of discourse unit 2001. For the referents "Margaret" and "James", mentioned therein, the nearest antecedents are found in discourse unit 1802. It is easy to see that the linear distance from 2001 to 1802 is 6 (which is a very high distance) while the rhetorical distance is just 2 (first step: from 2001 to 1803, second step from 1803 to 1802). Perhaps the most conclusive examples of the power of rhetorical distance as a factor in referential choice are the cases of long quotations: it is

---

[8] Note that one referent mention often has two distinct closest antecedents: a rhetorical and a linear one.

[9] It is a commonplace in the research on Rhetorical Structure Theory that there is certain constrained variation in how a given text can be represented as a hierarchical graph by different annotators (Mann *et al.*, 1992; Carlson *et al.*, 2003). To be sure, the fact of variation is the inherent property of discourse interpretation, and there is no other way of getting "better" hierarchical trees than rely on judgment of trained experts.

often the case that in a clause following a long quotation one can use a pronoun, with the nearest antecedent occurring before the quotation. This is possible in spite of the very high linear distance, and due to the short rhetorical distance: the pronoun's clause and the antecedent's clause in such case can be directly connected in the rhetorical structure.

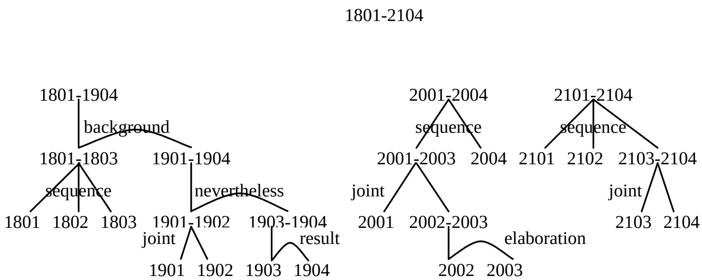


Figure 2: A rhetorical graph corresponding to lines 1801–2104 of the excerpt given in the Appendix

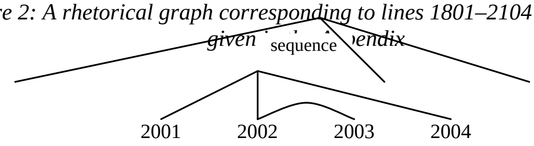


Figure 3. One-dimensional representation of a fragment of the rhetorical graph.

The following factor, indicated above, and the second most powerful source of activation, is the factor of *syntactic role of the linear antecedent*. This factor applies only when the linear distance is short enough: after about four discourse units it gets forgotten what the role of the antecedent was; only the fact of its presence may still be relevant. Also, this factor has a fairly diverse set of values. As it has long been known from studies of syntactic anaphora, subject is the best candidate for the pronoun's antecedent.[10] Different subtypes of subjects, however, make different contributions to referent activation, ranging from 0.4 to 0.2. Other relevant values of the factor include the direct object, the indirect (most frequently, agentive) object, the possessor, and the nominal part of the predicate. It is very typical of pronouns, especially for categorical pronouns (allowing no full NP alternative) to have subjects as their antecedents. For example, consider three pronouns in paragraph #16 (see Appendix): <u>*she*</u> (discourse unit 1603), <u>*her*</u> (1606), and <u>*she*</u> (1608). According to the results of the experimental study mentioned above, the first and the second pronouns are

[10] This observation is akin to the ranking of "forward-looking centres" in Centering theory, suggesting that the subject of the current utterance is the likeliest among other participants to recur in the next utterance with a privileged status; see e.g. (Walker & Prince, 1996: 297).

categorical (that is, Margaret could not be used instead) and they have subject antecedents. But the third one has a non-subject antecedent, and it immediately becomes a potentially alternating pronoun (Margaret would be perfectly appropriate here).[11]

The following two factors are related not to the previous discourse but to the relatively stable properties of the referent in question. *Animacy* specifies the permanent characterization of the referent on the scale "human – animal – inanimate". *Protagonisthood* specifies whether the referent is the main character of the discourse. Protagonisthood and animacy are rate-of-deactivation compensating factors (see discussion in Section 2.3). They capture the observation that important discourse referents and human referents deactivate slower than those referents that are neither important nor human. In addition, a group of second-order, or "weak", factors were identified, including the following ones. *Supercontiguity* comes into play when the antecedent and the discourse point in question are in some way extraordinarily close (e.g. being contiguous words or being in one clause). *Temporal or spatial shift* is similar to paragraph boundary but is a weaker episodic boundary; for example, occurrence of the clause-initial *then* frequently implies that the moments of time reported in two consecutive clauses are distinct, in some way separated from each other rather than flowing one from the other. *Weak referents* are those that are not likely to be maintained, they are mentioned only occasionally. Such referents often appear without articles (cf. NPs *rain, cinnamon and honey, supper* in the text excerpt given in the Appendix) or are parts of stable collocations designating stereotypical activities (*slam the door, light the lamps, give a bath*). Finally, *introductory antecedent* means that when a referent is first introduced into discourse, it takes no less than two mentions to fully activate it.

For details on the specific values of all activation factors, and the corresponding numeric weights, refer to (Kibrik, 1999). As in case of the Russian study, the numeric activation weights of each value were obtained through a long heuristic trial-and-error procedure. All referential facts contained in the original discourse and obtained through experimentation with alternative forms of reference are indeed predicted/explained by the combination of activation factors with their numeric weights, and the referential strategies.

---

[11] This demonstration of one factor operating in isolation is not intended to be conclusive, since the essence of the present approach is the idea that all factors operate *in conjunction*. It does, however, serve to illustrate the point.

The referential strategies formulated in this study are represented in Table 4. As in Section 2.3, the referential strategies indicate the mappings of different intervals on the AS scale onto possible referential devices.

Referential device:	Full NP only	Full NP, ?pronoun	Either full NP or pronoun	Pronoun, ?full NP	Pronoun only
AS:	0–0.2	0.3–0.5	0.6–0.7	0.8–1.0	1.1+

Table 4: Referential strategies in English narrative discourse

The quantitative system in this study was designed so that AS can sometimes exceed 1 and reach the value of 1.1 or even 1.2. This is interpreted as “extremely high activation” (it gives the speaker no full NP option to mention the referent, see the value in the rightmost column of Table 4 and below). The AS of 1 is then interpreted as “normal maximal” activation. Also, a low AS frequently turns out to be negative. Such values are simply rounded to 0.

According to the referential strategies represented in Table 4, the five categories of potential referential forms correspond to five different intervals on the activation scale. There are four thresholds on this scale. The thresholds of 0.2 and 1.0 are hard: when the AS is 0.2 or less, a pronoun cannot be used, and when it is over 1.0, a full NP cannot be used. There are also two soft thresholds: when the AS is 0.5 or less, a pronoun is unlikely, and when it is over 0.7, a full NP is unlikely.

To demonstrate how predictively the calculative system of activation factors works, several examples of actual calculations are presented in Table 5. All examples are taken from the text excerpt given in the Appendix. Examples are different in that they pertain to different referential options possible on the AS scale (see Table 4 above). There is one example for each of the following referential options: (a) full NP, ?pronoun; (b) either full NP or pronoun; (c) pronoun, ?full NP; (d) pronoun only.

The upper portion of Table 5 contains a characterization of each example: its location in the text, the actual referential form used by the author, the referent, the type of referential device and possible alternative devices, as obtained through the experimental study described above. Also, the AS interval corresponding to the referential option in question is indicated, in accordance with the referential strategies given in Table 4 above. The lower middle portion of Table 5 demonstrates the full procedure of calculating the ASs, in accordance with the values’ numeric weights. The last line of Table 5 indicates whether the calculated AS fits within the range predicted by the referential strategies.

Referential option		**(a) Full NP, $^{?}$pronoun**	**(b) Full NP or pronoun**	**(c) Pronoun, $^{?}$full NP**	**(d) Pronoun only**
Line number		1802	1701	1802	1603
Referential form		Margaret	*She*	*him*	*she*
Referent		"Margaret"	"Margaret"	"James"	"Margaret"
Actual referential device		full NP	pronoun	pronoun	pronoun
Alternative referential device		$^{?}$pronoun	full NP	$^{?}$full NP	—
Corresponding AS interval		*0.3–0.5*	*0.6–0.7*	*0.8–1.0*	*1+*
Relevant activation factors					
RhD	VALUE:	3	2	1	1
	NUM. WEIGHT:	*0*	*0.5*	*0.7*	*0.7*
LinD	VALUE:	3	2	1	1
	NUM. WEIGHT:	*–0.2*	*–0.1*	*0*	*0*
ParaD	VALUE:	1	1	0	0
	NUM. WEIGHT:	*–0.3*	*–0.3*	*0*	*0*
Lin. antec. role	VALUE:	S	S	passive S	S
	NUM. WEIGHT:	*0.4*	*0.4*	*0.2*	*0.4*
Animacy	VALUE:	Human, LinD $\geq$ 3	Human, LinD $\leq$ 2	Human, LinD $\leq$ 2	Human, LinD $\leq$ 2
	NUM. WEIGHT:	*0.2*	*0*	*0*	*0*
Protagonist-hood	VALUE:	Yes, RhD+ParaD $\geq$ 3	Yes, RhD+ParaD $\geq$ 3	Yes, RhD+ParaD $\leq$ 2	Yes, RhD+ParaD $\leq$ 2
	NUM. WEIGHT:	*0.2*	*0.2*	*0*	*0*
Calculated AS		*0.3*	*0.7*	*0.9*	*1.1*
Fit within the predicted AS interval		*Yes*	*Yes*	*Yes*	*Yes*

Table 5: Examples of calculations of the referents' ASs in comparison with the predictions of the referential strategies (for explanation of factors' values see Kibrik 1999)

2.5 Consequences for working memory

The studies outlined in Sections 2.3 and 2.4 rely on work in cognitive psychology, but they are still purely linguistic studies aiming at explanation of phenomena observed in natural discourse. However, it turns out that the results of those studies are significant for a broader field of cognitive science, specifically for research in *working memory*.

Working memory (WM; otherwise called short-term memory or primary memory) is a small and quickly updated storage of information. The study of WM is one of the most active fields in modern cognitive psychology (for reviews see (Baddeley, 1986; Anderson, 1990: Ch. 6); some more recent approaches are represented in (Gathercole, 1996; Miyake & Shah, 1999; Schroeger *et al.*, 2000). WM is also becoming an important issue in neuroscience: see (Smith & Jonides, 1997). There are a number of classical issues in the study of WM. Shah and Miyake (1999) list eight of major theoretical questions in WM. It appears that the results obtained in this linguistic study contribute or at least relate to the majority of these hot questions, including:

- *capacity*: how much information can there be in WM at one time?
- *forgetting*: what is the mechanism through which information quits WM?
- *control*: what is the mechanism through which information enters WM?
- *relatedness to attention*: how do WM and attention interact?
- *relatedness to general cognition*: how does WM participate in complex cognitive activities, such as language?
- *(non-)unitariness*: is WM a unitary mechanism or a complex of multiple subsystems?

Here, only some results related to the issues of capacity and attentional control will be mentioned. For more detail, refer to (Kibrik, 1999).

The system of activation factors and their numeric weights was developed in order to explain the observed and potential types of referent mentions in discourse. In the first place, only those referents that were actually mentioned in a given discourse unit were considered. But this system was discovered to have an additional advantage: it operates independently of whether a particular referent is actually mentioned at the present point in discourse. That is, the system can identify any referent's activation at any point in discourse no matter whether the author chose to mention it in that unit or not. If so, one can calculate the activation of *all referents* at a given point in discourse. Consider discourse unit 1608 (see Appendix). Only two referents are mentioned there: "Margaret" and "the cabin". However, the following other referents have AS greater than 0 at this point: "the anchor", "the gear", "rain", "the deck", "thunder", "lightning", and "the sky". The sum of ASs of all relevant referents gives rise to *grand activation* – the summed activation of all referents at the given point in discourse. Grand activation gives us an estimate of the *capacity* of the specific-referents portion of WM.

Figure 4 depicts the dynamics of activation processes in a portion of the English discourse (lines 1401 through 2104, see Appendix). There are three curves in Figure 4: two pertaining to the activation of the protagonists "Margaret" and "James", and the third representing the changes in grand activation. Observations of the data in Figure 4 make it possible to arrive at several important generalizations. Grand activation varies normally within the range between 1 and 3, only rarely going beyond this range and not exceeding 4. Thus the variation of grand activation is very moderate: maximally, it exceeds the maximal activation of an individual referent only about three to four times. This gives us an estimate of the maximal capacity of the portion of WM related to specific referents in discourse: three or four fully activated referents. Interestingly, this estimate coincides with the results recently obtained in totally independent psychological research looking at working memories specialized for specific kinds of information (Velichkovsky *et al.*, 1995; Cowan, 2000). Furthermore, there are strong shifts of grand activation at paragraph boundaries; even a visual examination of the graph in Figure 4 demonstrates that grand activation values at the beginnings of all paragraphs are local minima; almost all of them are below 2. On the other hand, in the middle or at the end of paragraphs, grand activation usually has local maxima. Apparently, one of the cognitive functions of a paragraph is a threshold of activation update.

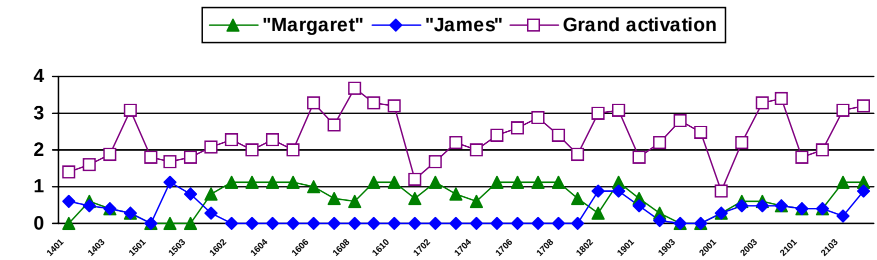


Figure 4: The dynamics of two protagonist referents' activation and of grand activation in an excerpt of English narrative (given in the Appendix)

The question of *control* of WM is the question of how information comes into WM. The current cognitive literature connects attention and WM (see e.g. (Miyake & Shah, 1999)). The issue of this connection is still debated, but the following claim seems compatible with most approaches:

➢ the mechanism controlling WM is what has long been known as attention

This claim is compatible with the already classical approaches of Baddeley (1990) and Cowan (1995), with the neurologically oriented research of Posner and Raichle (1994), and cutting-edge studies such as (McElree, 2001). According to Posner and Raichle (1994: 173), information flows from executive attention, based in the brain area known as anterior cingulate, into WM, based in the lateral frontal areas of the brain.

At the same time, as it has been convincingly demonstrated in the experimental study by Tomlin (1995), attention has a linguistic manifestation, namely grammatical roles. In many languages, including English, focally attended referents are consistently coded by speakers as the subjects of their clauses. As it has been demonstrated in the present paper, subjecthood and reduced forms of reference are causally related: antecedent subjecthood is among the most powerful factors leading to the selection of a reduced form of reference. In both English and Russian, antecedent subjecthood can add up to 0.4 to the overall activation of a referent. In both English and Russian sample discourses, 86% of pronouns allowing no referential alternative have subjects as their antecedent.

Considered together, these facts from cognitive psychology and linguistics lead one to a remarkably coherent picture of the interplay between attention and WM, both at the linguistic and at the cognitive level. Attention feeds WM, i.e. what is attended at moment t_n becomes activated in WM at moment t_{n+1}. Linguistic moments are discourse units. Focally attended referents are typically coded by subjects; at the next moment, they become activated (even if they were not before) and are coded by reduced NPs. The relationships between attention and WM, and between their linguistic manifestations, are presented in Table 6.[12]

Moments of time (discourse units)	**t_n**	**t_{n+1}**
Cognitive phenomenon	focal attention	high activation
Linguistic reflection	mention in the subject position	reduced NP reference
Examples	Margaret, she	*she, her*

Table 6: Attention and working memory in cognition and in discourse

2.6. Conclusions about the cognitive calculative approach

The approach outlined above aims at predicting and explaining all referential occurrences in the sample discourse. This is done through a rigorous calculative methodology aiming at maximally possible predictive power. For each referent at any point in discourse, the numeric weights of all involved activation factors are available. On the basis of these weights, the integral current AS of the referent can be calculated, and mapped onto an appropriate referential device in accordance with referential strategies. The objective fluidity of the process of referential choice is addressed through the distinction between the categorical and potentially alternating referential devices. This approach allows to overcome the traditional stumbling blocks of the studies of reference: circularity and multiplicity of involved factors. The linguistic study of referential choice in discourse was based on cognitive-psychological research,

[12] As it has been suggested by an anonymous reviewer, this account may resemble the claims of the Centering theory on dynamics of forward- and backward-looking centers. However, we would point out that the concept of “backward-looking center” is quite different from our idea of referent activation in the subsequent discourse unit. Centering theorists posit a single backward-looking center and claim that it is the referent that discourse unit is about (see e.g. (Walker & Prince, 1996: 294-5)). Therefore, backward-looking center must be more like topic or attention focus rather than activated referent. We do not know how such concept of backward-looking center could be incorporated in the cognitively inspired model of attention-memory interplay we propose.

and it proved, in turn, relevant for the study of cognitive phenomena in a more general perspective.

3 The neural network approach

3.1 Shortcomings of the calculative approach

There are some problems with the cognitive calculative approach, especially with its calculative, or quantitative, component that was mathematically quite unversed.

First, the list of relevant activation factors may not be exactly necessary and sufficient. Those factors were included in the list that showed a strong correlation with referential choice. However, only all factors in conjunction determine the activation score, and therefore the strength of correlation of individual factors may be misleading, and the contribution of individual factors is not so easy to identify. We would like to construct an "optimal" list of factors, i.e. a model that provides maximal descriptive power (all relevant factors identified and included) and at the same time has a minimal descriptional size (just the relevant factors contained and no others).

Second, numeric weights of individual factors' values were chosen by hand which not only was a laborious task, but also did not allow judging the quality or uniqueness of the set of calculated weights.

Third, the interaction between factors was mainly additive, ignoring possible non-linear interdependencies between the factors. Non-linear dependencies are particularly probable, given that some factors interact with others (cf. the discussion of the factor of syntactic role of the linear antecedent in Section 2.4 above, whose contribution to AS depends on the linear distance).[13] Other factors might be correlated, e.g. animacy and the syntactic role of subject (the distribution of animacy and subjecthood of the antecedent vis-à-vis full NPs vs. pronouns is very similar, indicating a possible intrinsic interrelationship between these).[14] Also, from the cognitive point of view, it is unlikely that such a simple procedure as addition can adequately describe processing of activation in the brain: the basic building blocks of the brain, the nerve cells or neurons, exhibit non-linear behaviour, for example due to saturation effects. It is well

[13] Indeed, the attribution of different weights to the syntactic role of the linear antecedent *depending* on the linear distance in the calculative approach can already be viewed as an element of non-linear interdependencies.

[14] As a mathematical consequence, the weights attributed to animacy and antecedent subjecthood are not "stable": The model would perform almost as well if the numeric weights for these two factors were interchanged or even modified so that their sum remained the same. Thus, the concrete single weights of correlated factors have no objective importance on their own, and it is important to single out correlated factors and describe their relationship in order to ascribe an objective meaning to a combination (most simply, the sum) of their weights.

known that purely linear learning schemes cannot even solve the simple exclusive-or problem, see e.g. (Ellis & Humphreys, 1999: Ch. 2.4). For an in-depth discussion of the usefulness of non-linearity in cognitive and developmental psychology, we refer to (Elman *et al.*, 1996).

Fourth, because of the additive character of factor interaction it was very hard to limit possible activation to a certain range. It would be intuitively natural to posit that minimal activation varies between zero and some maximum, which can, without loss of generality, be assumed to be one. However, because of penalizing factors such as paragraph distance that deduct activation, it often happens that activation score turns out negative (a consequence of the simple summing in the calculative approach), which makes cognitive interpretation difficult.

In order to solve these problems, the idea to develop a more sophisticated mathematical apparatus emerged, such that:

- identification of significant factors, numeric weights, and factor interaction would all be interconnected and would be a part of the same task
- the modeling of factors would be done computationally, by building an optimal model of factors and their interaction.

There are many well-known approaches that lend themselves naturally to the problems mentioned above (e.g. variants of decisions tree algorithms, multiple non-linear regression). Since what we have in mind is to develop a quantitative cognitive model of referential choice as a long-term goal, artificial neural network models had a strong appeal to us due to their inherent cognitive interpretation (Ellis & Humphreys, 1999), even though we cannot expect a concrete cognitive model or interpretation to derive from this pilot study based on just a small data set.[15] We note that the – at first sight – less transparent representation of knowledge in a neural network, as compared to classical statistical methods, is balanced by the fact that the type of regularities it can detect in the data is less constrained.

We would like to emphasize that the primary aim of this pilot study on a quite small data set is to evaluate whether neural networks are applicable to the problem of referential choice, and if so, to lay the ground for a larger-scale study. In order to keep the present study comparable to the calculative approach, we had to use the original data set and neglected from the outset factors that already had been judged secondary.

We dispense with a more sophisticated statistical analysis of the following computer simulations since – from the point of view of rigorous statistics – the

---

[15] With respect to the small data set, we would not be better off with any other of the above mentioned methods as all of them are quite data-intensive.

data set is too small to lead to reliable results. Our intention is to get a first taste of where neural networks might take us in the analysis of referential choice.

3.2 Proposed solution: a neural network approach

In the neural network approach, we lift the requirement of complete predictiveness: we posit that referential choice can predict/explain referential choice with a degree of certainty that can be less than 100%.[16] Also, at this time, the neural network approach does not make specific claims about cognitive adequacy and activation and there is no such thing as summary activation score in this approach at its present stage. Activation factors themselves are reinterpreted as mere parameters or variables in the data that are mapped onto referential choice. We expect that at a later stage – i.e. trained on bigger data sets – the neural network approach can embrace the quantitative cognitive component.

The term *artificial neural network* or *net* denotes a variety of different function approximators that are neuro-biologically inspired (Mitchell, 1997). Their common property is that they can, in a supervised or unsupervised way, learn to classify data. For this pilot study, we decided to employ a simple feed-forward network with the back-propagation learning algorithm.

A feed-forward network consists of *nodes* that are connected by *weights.* Every node integrates the activation it gets from its predecessor nodes in a non-linear way and sends it to its successors. The nodes are ordered in layers. Numeric data is presented to the nodes in the *input layer*, from where the activation is injected into one or more *hidden layers*, where the actual computation is done. From there activation spreads to the *output layer*, where the result of the computation is read off. This computed output can be compared to the expected target output, and subsequently the weights are adapted so as to minimize the difference between actual output and target (a so-called gradient descent algorithm, of which the backpropagation algorithm is an example, for details we refer to (Ellis & Humphreys, 1999)).

In this supervised learning task, the network must learn to predict from ten factors (Table 7), whether the given referent will be realized as a pronoun or a full noun phrase. In order to input the factors with symbolic values into the net, they have to be converted into numeric values. If the symbolic values denote some gradual property such as animacy, they are converted into one real variable with values between –1 and 1. The same holds true for binary variables. When there was no a priori obvious order in the symbolic values,[17]

[16] This might be a desirable feature, e.g. to account for alternating referential options.

[17] For example, the factor of syntactic role can take the values “subject”, “direct object”, “indirect object”, “possessive”, etc. One might speculate that a hierarchy of these values, similar to the hierarchy of NP

they were coded unary (e.g. Syntactic Role), i.e. to every value of that factor corresponds one input node, which is set to one if the factor assumes this value and to zero otherwise.

Thus, 24 input nodes and 1 output node are needed. The output node is trained to predict whether the referent in question is realized as a full noun phrase (numeric output below 0.4) or as a pronoun (numeric output above 0.6).[18] At this point, all numeric input values were normalized to have zero mean and unit variance. This normalization ensures that all data are a priori treated on equal footing and the impact of a factor can be directly read off from the strength of the weights connecting its input node to the hidden or output layer.

3.3 Simulation 1 – full data set

A network with 24 nodes in a single hidden layer was trained on the data set of 102 items[19] from (Kibrik, 1999) (see Section 2.4) for 1000 epochs.[20] As parts of the training are stochastic, that experiment was repeated several times. In all runs the net learned to predict the data correctly except for a small number (below six) cases. Typically, the misclassifications occurred for the same items in the data set, independently of the run. A closer analysis of a well-trained net with only four misclassifications revealed that three of them were due to referential conflict (which was not among the input factors), that is, in the situation when the full noun phrase is used only because a pronoun (otherwise expected) may turn out ambiguous.

accessibility (Keenan & Comrie, 1977), might operate in referential choice. But since this is not self-evident, we code such factors unarily so that the network can find its own order of the values as relevant for the task at hand.

[18] An output value between 0.4 and 0.6 is considered unclassified. However, this did not happen in the simulations presented here. Of course, the target values are 0 and 1 for pronouns and full NPs, respectively. Yet, for technical reasons it is preferable to admit a small deviation of the output value from the target values.

[19] As opposed to the study in Section 2.4, here the syntactic pronouns were included. Note that due to short linear distance all of them are easily predicted correctly.

[20] Technical details for NN experts: learning parameter is set to 0.2; no momentum; weights were jogged every epoch by maximally 0.1%; input patterns are shuffled. The simulations are run on the SNNS network simulator (http://www-ra.informatik.uni-tuebingen.de/SNNS).

Factor	Values	Coding	Input Nodes
Syntactic role	S, DO, IOag, Obl, Poss	Unary	1–5
Animacy	Human, animal, inanimate	Human: 1, animal: 0, inanimate: –1	6
Protagonisthood	Yes / no	Binary	7
Syntactic role of rhetorical antecedent*	S, DO, IOag, Obl, Poss, Pred	Unary	8–13
Type of rhetorical antecedent	Pro, FNP	Binary	14
Syntactic role of linear antecedent	S, Poss, Obl, Pred, DO, IOag	Unary	15–20
Type of linear antecedent	Pro, FNP	Binary	21
Linear distance to antecedent	Integer	Integer	22
Rhetorical distance to antecedent	Integer	Integer	23
Paragraph distance to antecedent	Integer	Integer	24

S, DO, IOag, Obl, Poss mean subject, direct object, agentive indirect object, oblique, and possessor. Pred means predicative use, Pro pronoun and FNP full noun phrase.

Table 7. Factors used in Simulation 1, their possible values and the corresponding input nodes.

3.4 Simulation 2 – pruning

Not only did we want our net to learn the data but also to make some statements about the importance of the input factors and their interdependency. To achieve this goal, we submitted the trained net from Simulation 1 to a *pruning procedure*, which eliminates nodes and weights from the net that contribute to the computation of the result only little or not at all. In such case, a node or weight is selected and eliminated. Then the net is retrained for 100 epochs. If net performance does not drop, the elimination is confirmed; otherwise, the deleted node or weight is restored. This procedure is repeated until no further reduction in the size of the net is possible without worsening the performance.[21]

This procedure leads to smaller nets that are easier to analyse and furthermore can reduce the dimensionality of the input data. They have a lower number of weights (i.e. a lower number of free parameters: in the case analysed here, the number of weights was reduced from 649 for the full net to 26 for the pruned net). The weights of a generic example of a pruned network trained on our data are shown in Table 8. There are no weights connecting the input nodes 3, 4, 5, 6, 11, 13, 18, 19, 20, 23 (see Table 8; the meanings of the nodes can be found in Table 7). This means that not all input factors or all their values are

[21] More precisely, first we apply the *non-contributing units* algorithm (Dow & Sietsma, 1991), and then pruning of the minimal weight.

relevant for computing the output. Also, all but two hidden nodes have been pruned. So the two remaining suffice to model the interaction between the input factors.

Some input nodes have a direct influence on the output node (27), e.g. the node indicating that the rhetorical antecedent was a possessor (node 9). Others influence the outcome only indirectly by interacting with other nodes, e.g. paragraph distance (node 24), while yet others influence the output both directly and indirectly. Some nodes enter in multiple ways that seem to cancel each other, e.g. node 14 (type of rhetorical antecedent).

Target node	**Source Nodes (Weights)**							
25	1 (-2.4)	2 (2.1)	8 (-1.7)	12 (1.9)	14 (-1.6)	16 (-2.4)	22 (-4.7)	24 (-4.9)
26	7 (1.7)	10 (-2.0)	12 (-5.0)	14(-1.9)	15 (2.8)	16 (-1.8)	21 (-4.2)	
27	2 (-3.7)	8 (3.9)	9 (2.0)	15 (2.7)	17 (1.8)	22 (-22.0)	25 (10.9)	26 (-10.0)

Nodes 1—24 denote the input nodes, 25 and 26 are the two remaining hidden nodes and 27 is the output node. The weights connecting a source and a target node are given in parentheses after the source node.

Table 8. Weights of a typical pruned net.

Pruning again is partly a stochastic procedure, as it for example depends ultimately on the random initialisation of the network, so we repeated the experiment until we got an impression of which factors are almost invariably included. It turned out that subject and possessor roles,[22] protagonisthood, subjecthood of the antecedent and type of antecedent are most important, and those nodes related to the rhetorical antecedent are more involved than those for the linear one. Accordingly, the most important distance is rhetorical distance. As expected, this list of factors and values coincides to a great extent with what was discovered through the trial-and-error procedure in the calculative approach. Therefore, at least qualitatively, the neural network approach is on the right track, and we can use the results of the pruning case study as a hint on how to reduce the dimensionality of the input data. This leads us to the next simulation.

3.5 Simulation 3 – reduced data set

In a third case study, we trained a similar net with 12 hidden nodes on a reduced set of only five input factors (corresponding to six input nodes): We included the values “subject” and “possessor” for syntactic role (nodes 1, 2),

[22] Interestingly, some hints on the difference in the usage of argumental and possessive pronouns were observed already during the original work on the calculative approach. The fact that the networks themselves frequently keep the input for the possessive role can be viewed as a corroboration of this thought, and also as a proof that neural networks can be used as an independent tool for discovering regularities in the data. Work focusing on this differentiation is underway.

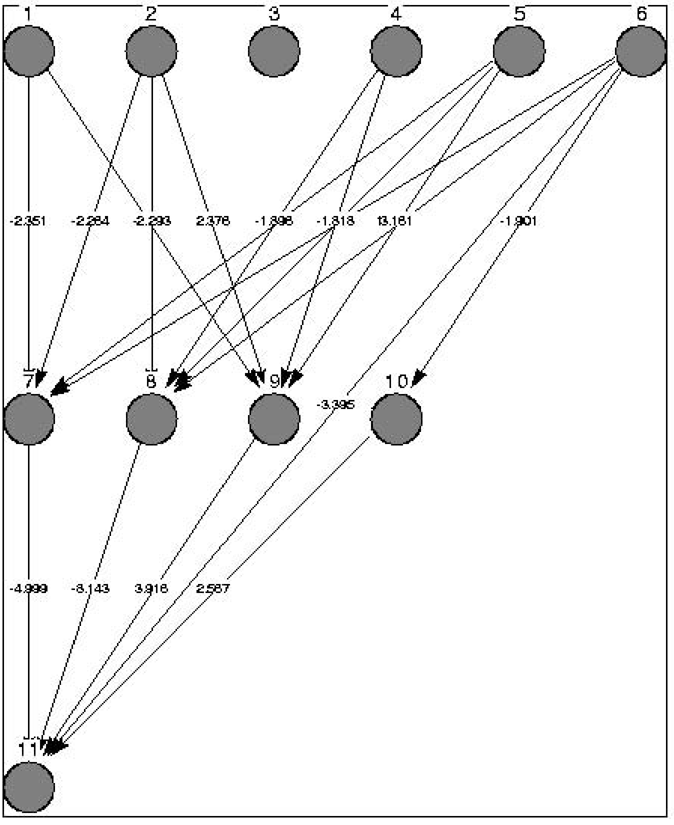
protagonisthood (node 3), both when the rhetorical antecedent was a subject (node 4), and when it was realized as a pronoun or full NP (node 5), and rhetorical distance (node 6). The new net had 12 hidden nodes, corresponding to 103 weights. On this reduced net, we executed the back-propagation learning algorithm for 500 epochs and then pruning (50 epochs retraining for each pruning step) with the same parameters as before. We ended up with a small net (23 parameters), shown in Figure 5, that classified only 8 out of 102 items wrongly. Note that all remaining factors interact strongly, except for protagonisthood (node 3), which has been pruned away.

3.6 Simulation 4 – cheap data set

Reliable automatic annotators for rhetorical distance and consequently for all factors related to the rhetorical antecedent, as well as for protagonisthood, are not available. Since these factors require comprehension of the contents of the text, they must be annotated by human experts and are therefore costly. So we decided to replace the rhetorical factors included in Simulation 3 by the corresponding linear ones and protagonisthood by animacy. Keeping the six input nodes as before, we added a seventh one to indicate that the linear antecedent was a possessor and an eighth one for paragraph distance to help the net to overcome the smaller amount of information that is contained in the linear antecedent factors. Training and pruning proceeded as before.

One typical resulting network in this case had 32 degrees of freedom. Again, animacy, which had been substituted for protagonisthood, is disconnected from the rest of the net. On the 102 data items, the net produced only six errors (three are due to referential conflict).

Thus, even though the logical structure of the factors and their values was considerably simplified, and none of the factors included relate to the rhetorical antecedent, the accuracy (six errors versus four with the full set of factors) did not deteriorate dramatically.



The circles denote the nodes, the arrows the weights connecting the nodes, to which the weight strength is added as a real number. Nodes 1–6 are input nodes, 7–10 the nodes in the hidden layer, and node 11 is the output.

Figure 5. Net from Simulation 3

3.7 Comparison to the calculative approach

In the calculative model discussed in Section 2.4 above, referential choice was modelled by 11 factors using 32 free parameters (counting the number of the different numeric weights for all factors and their values). The activation score allowed a prediction of the referential choice in five categories. In our study with neural networks, we modelled only a binary decision (full NP/pronoun) and lifted the requirement of cognitive adequacy. The smallest net in the study, in simulation 3, had only 23 free parameters (weights), 5 input factors, and the best net on the full set of input factors, in Simulations 1 and 2, misclassified only four items, having 26 free parameters.

Even though the accuracy dropped in the neural network approach (using a reduced set of input factors) as compared to the calculative approach (with the full set of input factors), the descriptional length (measured in the number of free model parameters) was reduced by approximately one third and thus yields in this sense a more compact description of the data.

These findings are important in the following respects. First, we can find a smaller set of factors that still allows a relatively good prediction of referential choice, but it is much less laborious to extract from a given corpus, thus making the intended large-scale study feasible. Second, we can reduce the descriptional length without a too severe drop in accuracy. This means that the networks were able to extract the essential aspects of referential choice since about 100 instances can be described by only 23 parameters. Compare this to the worst case in which a learning algorithm needs about 100 free parameters to describe 100 instances. In such case, the algorithm would not have learnt anything essential about referential choice, because it would be merely the list of the 100 instances. The ratio of the number of parameters to the size of data set has a long tradition of being used for judging a model's quality. A high value of this ratio is an indicator for overfitting[23] (see any standard textbook on statistics).

In large-scale studies, which are due to follow this pilot study, we expect to construct models with an even better ratio of descriptional length to the size of data set.

[23] Overfitting means sticking too closely to the peculiarities of a given training set and not finding the underlying general regularities. Overfitting is roughly the opposite of good generalization of unknown data.

3.8 Comparison to (Strube and Wolters, 2000)

As it has been pointed out above, there are relatively few studies of referential choice – most authors are interested in resolution of anaphoric devices. Furthermore, there are almost no studies that would attempt to integrate multiple factors affecting reference. However, we are familiar with one study that is remarkably close in spirit to ours, namely (Strube & Wolters, 2000). Strube and Wolters use a similar list of factors as the calculative approach discussed above, except that the costly factors related to the rhetorical antecedent are missing. They analyse a large corpus with several thousand of referring expressions for the categorical decision (full NP/pronoun) using logistic regression. The logistic regression is a form of linear regression adapted for a binary decision.

Factor interaction and non-linear relations are thus not accounted for in their model, and they present no cognitive interpretation of their model either. Still, the gist and intention of their and our studies – independently developed – largely agree, which provides evidence for the usefulness and appropriateness of quantitative approaches towards referential choice.

4 Conclusion and outlook

In Section 3, we reported a pilot study testing whether artificial neural networks are suitable to process our data. We trained feed-forward networks on a small set of data. The results show that the nets are able to classify the data almost correctly with respect to the choice of referential device. A pruning procedure enabled us to single out five factors that still allowed for a relatively good prediction of referential choice. Furthermore, we demonstrated that costly input factors such as rhetorical distance to the antecedent could be replaced by those related to the linear antecedent, which can be more easily collected from a large corpus.

Because of the small amount of data for this pilot study, the result must be taken with due care. But these results encourage us to further develop this approach.

Future work will include a study of a larger data set. This is necessary since neural networks as well as classical statistics need a large amount of data to produce reliable results that are free of artefacts. In our corpus, some situations (i.e. an antecedent that is an indirect object) appear only once, so that no generalization can be made. In a larger study, the advantages of the neural network approach can be used fully.

We also aim at reintroducing a cognitive interpretation at a later stage, and we want to work with different network methods, that not only allow dimensional reduction and data learning, but also an easy way to explicitly extract the knowledge from the net in terms of more transparent symbolic rules (see e.g. (Kolen & Kremer, 2001)).

Furthermore, we feel the need not only to model a binary decision (full NP/pronoun), but also to have a more fine-grained analysis. The calculative approach of Section 2.4 has done the first steps in this direction, allowing for five different categories that not only state that a pronoun or a full NP is expected, but also to what degree a full NP in a particular situation can be replaced by a pronoun and vice versa.

A statistical interpretation of referential choice can be suggested: if a human expert judges that a particular full NP could be replaced by a pronoun, s/he must have experienced that in a very similar situation where the writer did indeed realize the other alternative. The expert will be more certain that substitution is suitable if s/he has often experienced the alternative situation. Thus we think it is promising to replace the five categories discussed in Section 2.4 by a continuous resulting variable that ranges from 0 to 1 and is interpreted as the probability that referential choice realizes a pronoun in the actual situation: 1 means a pronoun with certainty, 0 means a full NP with certainty, and 0.7 means that in 70% instances a pronoun is realized and a full NP in the remaining 30% instances.

As an anonymous reviewer pointed out to us, there is an interesting potential application of neural network-based models of referential choice to anaphor resolution. Consider a knowledge-poor anaphor resolution algorithm as a quick-and-dirty first pass that suggests several potential referents for a pronominal mention. Counterchecking the referent mentions in a second pass, a suggested referent could be ruled out if the network does not predict a pronominal mention for it at the point in question. The advantage over anaphor resolution algorithms based purely on classical methods would be that computations in a neural network are really fast compared to algorithmic and symbolic computing once the training of the network is finished.

Acknowledgements

Andrej Kibrik's research underlying this article has been supported by grant #03-06-80241 from the Russian Foundation for Basic Research. Andrej Kibrik expresses his gratitude to the Alexander von Humboldt Foundation and Max-Planck-Institute for Evolutionary Anthropology that made his research in Germany (2000–2001) possible. The assistance of Russ Tomlin and Gwen Frishkoff in 1996 was crucial in the research reported in Section 2.4. We are as well indebted to the valuable comments of three anonymous referees.

References

Anderson, J. R. 1990. *Cognitive Psychology and its Implications*. 3rd ed. New York: W. H. Freeman & Co.
Baddeley, A. 1986. *Working Memory*. Oxford: Clarendon Press.
---------- 1990. *Human Memory: Theory and Practice*. Needham Heights: Allyn and Bacon.
Botley, S. and A. M. McEnery (eds.) 2000. *Corpus-based and Computational Approaches to Discourse Anaphora*. Amsterdam: John Benjamins.
Carlson, L., D. Marcu and M. E. Okurowski 2003. “Building a Discourse-tagged Corpus in the Framework of Rhetorical Structure Theory.” Jan van Kuppevelt and Ronnie Smith (eds.) *Current Directions in Discourse and Dialogue*. Dordrecht: Kluwer. To appear.
Chafe, W. 1994. *Discourse, Consciousness, and Time. The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.
Clifton, C. Jr. and F. Ferreira 1987. “Discourse Structure and Anaphora: Some Experimental Results.” M. Coltheart (ed.) *Attention and Performance XII*. Hove: Erlbaum.
Cornish, F. 1999. *Anaphora, Discourse, and Understanding. Evidence from English and French*. Cambridge: Cambridge University Press.
Cowan, N. 1995. *Attention and Memory: An Integrated Framework*. New York, Oxford: Oxford University Press.
---------- 2000. “Childhood Development of some Basic Parameters of Working Memory.” Schroeger, Mecklinger, and Friederici 2000. 139-160.
Dale, R. 1992. *Generating Referring Expressions*. Cambridge, MA: MIT Press
Dow, R. and J. Sietsma. 1992. “Creating Artificial Neural Networks that Generalize.” *Neural Networks* 4(1):67-79.
Ellis, R. and G. W. Humphreys 1999. *Connectionist Psychology*. East Sussex: Psychology Press.
Elman, J., E. Bates, M. Johnson, A. Karmiloff-Smith, D. Parisi and K. Plunkett 1996. *Rethinking Innateness – A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.
Fox, B. 1987a. *Discourse Structure and Anaphora*. Cambridge: Cambridge University Press.
---------- 1987b. “Anaphora in Popular Written English Narratives.” R. Tomlin (ed.) *Coherence and Grounding in Discourse*. Amsterdam: John Benjamins. 157-174.
Gathercole, S. E. (ed.) 1996. *Models of Short-term Memory*. East Sussex: Psychology Press.
Gernsbacher, M. A. 1990. *Language Comprehension as Structure Building*. Hillsdale: Erlbaum.
Givón, T. (ed.) 1983. *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam: John Benjamins.
---------- 1990. *Syntax: A Functional-typological Introduction*. Vol. 2. Amsterdam: John Benjamins.
---------- 1995. *Functionalism and Grammar*. Amsterdam: John Benjamins.
Gordon, P., B. Grosz and L. Gilliom 1993. “Pronouns, Names, and the Centering of Attention in Discourse.” *Cognitive Science* 17.311-347.
Grimes, J. 1978. *Papers in Discourse*. Arlington: SIL.
Gundel, J., N. Hedberg and R. Zacharski 1993. “Cognitive Status and the Form of Referring Expressions in Discourse.” *Language* 69.274-307.
Keenan, E. and B. Comrie 1977. “Noun Phrase Accessibility and Universal Grammar.” *Linguistic Inquiry* 8(1).63-98.

Kibrik, A. A. 1991. “Maintenance of Reference in Sentence and Discourse.” Winfred P. Lehmann and Helen J. Hewitt (eds.) *Language Typology*. Amsterdam: John Benjamins. 57-84.

---------- 1996. “Anaphora in Russian Narrative Prose: A Cognitive Account.” Barbara Fox (ed.) *Studies in Anaphora*. Amsterdam: John Benjamins.255-304.

---------- 1999. “Reference and Working Memory: Cognitive Inferences from Discourse Observation.” K. van Hoek, A. A. Kibrik and L. Noordman (eds.) *Discourse Studies in Cognitive Linguistics*. Amsterdam: John Benjamins.29-52.

---------- 2000. “A Cognitive Calculative Approach Towards Discourse Anaphora.” P. Baker, A. Hardie, T. McEnery and A. Siewierska (eds.) *Proc. of the Discourse Anaphora and Anaphor Resolution Conference (DAARC2000)*. Lancaster: University Centre for Computer Corpus Research on Language.72-82.

Kolen, J. F. and S. C. Kremer (eds.) 2001. *A Field Guide to Dynamical Recurrent Networks.* New York: IEEE Press.

Mann, W., C. Matthiessen and S. A. Thompson 1992. “Rhetorical Structure Theory and Text Analysis.” W. Mann and S. Thompson (eds.) *Discourse Description. Diverse Linguistic Analyses of a Fund-raising Text*. Amsterdam: John Benjamins.39-78.

Marslen-Wilson, W., E. Levy and L. K. Tyler 1982. “Producing interpretable discourse: The Establishment and Maintenance of Reference.” *Speech, Place, and Action. Studies in Deixis and Related Topics*, ed. R.J. Jarvella and W. Klein, 339-378. Chichester: Wiley.

McElree, B. 2001. “Working Memory and Focal Attention.” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27(3).817-835.

Mitchell, T. M. 1997. *Machine Learning*. Boston: McGraw-Hill.

Miyake, A. and P. Shah (eds.) 1999. *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. Cambridge: Cambridge University Press.

Posner, M. I. and M. E. Raichle. 1994. *Images of Mind.* New York: Scientific American Library.

Schroeger, E., A. Mecklinger and A. D. Friederici (eds.) 2000. *Working on Working Memory*. Leipzig: Leipziger Universitätsverlag.

Shah, P. and A. Miyake 1999. “Models of Working Memory: An Introduction.” Miyake and Shah (eds.). 1–26.

Smith, E. and J. Jonides. 1997. Working Memory: A View from Neuroimaging. *Cognitive Psychology* 33.5-42.

Strube, M. and M. Wolters 2000. “A Probabilistic Genre-Independent Model of Pronominalization.” *Proc. of the 1st Meeting of the North American Chapter of the ACL. Seattle, WA, USA, April 29 - May 4, 2000,* 18–25.

Tomlin, R. 1987. “Linguistic Reflections of Cognitive Events.” R. Tomlin (ed.) *Coherence and Grounding in Discourse*. Amsterdam: John Benjamins. 455-479.

---------- 1995. “Focal Attention, Voice and Word Order: An Experimental Cross-linguistic Study.” P. Downing and M. Noonan (eds.) *Word Order in Discourse*. Amsterdam: John Benjamins. 517-554.

---------- and M.-M. Pu 1991. “The Management of Reference in Mandarin Discourse.” *Cognitive Linguistics* 2: 65–93.

Toole, J. 1996. “The Effect of Genre on Referential Choice.” T. Fretheim and J. Gundel (eds.) *Reference and Referent Accessibility*. Amsterdam: John Benjamins. 263-290.

van Hoek, K. 1997. *Anaphora and Conceptual Structure*. Chicago: University of Chicago Press.

Van Valin, R. D. Jr. 1993. “A Synopsis of Role and Reference Grammar.” R. D. Van Valin, Jr. (ed.) *Advances in Role and Reference Grammar*. Amsterdam: John Benjamins.

Velichkovsky, B., B. Challis and M. Pomplun 1995. “Arbeitsgedächtnis und Arbeit mit dem Gedächtnis: Visuell-räumliche und weitere Komponenten der Verarbeitung.“ *Zeitschrift für experimentelle Psychologie* 42(4).672-701.

Vonk, W., L. Hustinx and W. Simons 1992. “The Use of Referential Expressions in Structuring Discourse.” *Language and Cognitive Processes* 7.301-333.

Walker, M. and E. Prince 1996. “A Bilateral Approach to Givenness: A Hearer-status Algorithm and a Centering Algorithm.” T. Fretheim and J. Gundel (eds.) *Reference and Referent Accessibility*. Amsterdam: John Benjamins. 291-306.

Appendix: An Excerpt from an English Narrative
("The Maggie B." by Irene Haas)

1401 Margaret and James were cold.
1402 The sky grew darker.
1403 The goat and chickens fled into their little shelter,
1404 the toucan flew screeching into the cabin.

1501 James started to cry.
1502 A storm was coming!
1503 Margaret must make the boat ready at once.

1601 She took in the sail
1602 and tied it tight.
1603 She dropped the anchor
1604 and stowed all the gear,
1605 while rain drummed on the deck
1606 and thunder rumbled above her.
1607 Lightning split the sky
1608 as she ran into the cabin
1609 and slammed the door against the wet wind.
1610 Now everything was safe and secure.

1701 When she lit the lamps,
1702 the cabin was bright and warm.
1703 It was nearly suppertime
1704 so Margaret mixed up a batch of muffins
1705 and slid them into the oven.
1706 She sliced some peaches
1707 and put cinnamon and honey on top,
1708 and they went into the oven, too.

1801 James was given a splashy bath in the sink.
1802 Margaret dried him in a big, warm towel,
1803 and then supper was ready.

1901 Outside, the wind howled like a pack of hungry wolves.
1902 Rain lashed the windowpanes.
1903 But the sturdy little Maggie B. kept her balance
1904 and only rocked the nicest little bit.

2001 Margaret and James ate the beautiful sea stew
2002 and dunked their muffins in the broth,
2003 which tasted of all the good things that had cooked in it.
2004 For dessert they had the peaches with cinnamon and honey, and glasses of warm goat's milk.

2101 When supper was over,
2102 Margaret played old tunes on her fiddle.
2103 Then she rocked James in his cradle
2104 and sang him his favorite song.

Degrees of Indirectness: Two Types of Implicit Referents and their Retrieval via Unaccented Pronouns[1]

Francis Cornish

CNRS ERSS (UMR 5610) and Université de Toulouse-Le Mirail

In this chapter, I aim to show that so-called "indirect" anaphora, when realized via unaccented pronouns, is less of a marked discourse phenomenon than previously claimed. After a definition of indirect anaphora, which is distinguished from "exophora", in particular, the chapter tries to delimit the threshold of discourse-cognitive activation or saliency beyond which the retrieval of an intended "indirect" referent by a token of this indexical form type is not possible without incurring a processing cost. One condition for such a retrieval is claimed to be the degree of centrality of the referent (central argument of the predicate concerned, or peripheral instrument) within the semantic-pragmatic structure in terms of which the antecedent-trigger is represented in the discourse already established at the point of retrieval. Another is the nature of the referent itself (specific though indefinite, on the one hand, or non-specific frame-bound entity, on the other). Finally, I will present the format for an experimental verification of the hypothesis outlined above which has recently been carried out, in both a French and an English version, and will summarize its main results.

1 Introduction

In this chapter, I aim to show that so-called "indirect" anaphora, when realized via unaccented pronouns, is less of a marked discourse phenomenon than has previously been claimed.[2] Pronouns (as well as zero forms) are particularly sensitive to the "in-focus" – that is, topical – status of the discourse representation of the referent which they are intended to retrieve. Taking indirect anaphora into account requires a redefinition of this psychological status. This is what I shall try to do in this chapter.

As will be evident from what follows, I take what might be termed a "discourse-cognitive" view of anaphoric reference, rather than a syntactic-

---

[1] This chapter is a revised and extended version of a paper entitled "Indirect Anaphora: The Discourse-referential Scope of Unaccented Pronouns and Zero Anaphors", which was presented at the 4th international colloquium on *Discourse Anaphora and Anaphor Resolution* (DAARC2002), held in Estoril, Portugal (18-20 Sept. 2002). I would like to thank the audience of this paper for the useful discussion which followed its presentation, Marion Fossard, Monika Schwarz, Knud Lambrecht, Daniel García Velasco, Lachlan Mackenzie and three anonymous reviewers from the book's Scientific Committee for commenting on various revised versions, and Jean-Pierre Koenig both for his comments on an earlier draft and for discussion of issues relating to implicit arguments and their potential status as discourse referents.

[2] E.g. by Dik, 1978; Sanford & Garrod, 1981; Sanford *et al*, 1983; Erkü & Gundel, 1987; Gundel *et al*, 2000.

textual one. The use and interpretation of (non-bound) anaphors requires not only a relevant co-text as well as context, but also, crucially, a psychologically salient representation of the discourse evoked via what I call the "antecedent-trigger" (an utterance token, gesture or percept). Anaphora is a means of managing the memory representation of the discourse being constructed by the speech participants on the basis of a co-text as well as a relevant context (for further details of this view, see (Cornish, 1999, 2002a, 2003)). The central thrust of this chapter is that, given that certain referents retrieved via given anaphors in a text will not have been introduced into the corresponding discourse via an explicit textual antecedent, but evoked "obliquely" via an association or a (stereotypical) inference of some kind, then there is more than one type of "indirect" or oblique, non-explicit referent which is potentially accessible via unaccented pronouns as anaphors: namely what I call "nuclear" and "peripheral" indirect referents. It is important, for a proper understanding of pronominal anaphoric reference, to separate these two types of indirect referent. By not doing so, certain accounts of the phenomenon, it will be argued, have posited erroneous constraints and principles regarding indirect anaphora.

I will start, then, by defining a view of the three-way distinction between "anaphora", "deixis" and "textual/discourse deixis", and go on to give a definition of indirect anaphora, distinguishing it from exophora (with which it may be confused) (Section 2). Section 3 presents a selection of existing hypotheses claiming that unaccented pronouns cannot retrieve "indirect" referents, and puts forward the chapter's central argument, namely that this is possible and natural when such referents are "nuclear", though it is only so with difficulty when they are "peripheral". Section 3.2 attempts to motivate the "nuclear" vs. "peripheral" distinction in terms of the argument structure and lexical-semantics of given verbal and adjectival antecedent-trigger predicates. Section 3.3 further distinguishes between specific indirect referents, which are subject to the nuclear/peripheral distinction, and non-specific, frame-bound referents, which are not. To end the chapter (Section 4), I briefly present the format for an empirical verification of the hypothesis put forward here (in both an English and a French version), together with a summary of the results of its implementation.

2 *Some useful concepts and distinctions in the study of indexical reference*

2.1 "Anaphora", "deixis" and "textual/discourse deixis"

To my mind, "anaphora" constitutes a procedure (realized via the text) for the recall of some item of information previously placed in discourse memory and already bearing a minimal level of attention activation. It is essentially a procedure for the orientation of the interlocutor's attention, which has as essential function the *maintenance* of the high level of activation which characterizes a discourse representation already assumed to be the subject of an attention focus by the interlocutor at the point of utterance. It is not only the anaphoric expression which is used (typically, a third person pronoun) which realizes anaphora, but also the clause in which it occurs as a whole. This predicational context acts as a kind of "pointer", orienting the addressee towards the part of the discourse representation already cognitively activated, and which will make it possible to extend in terms of an appropriate coherence relation (cf. Kleiber, 1994, Ch. 3). Here are some examples involving different possible continuations of the antecedent predication in terms of distinct anaphoric predications:

(1) a. Sean Penn attacked a photographer. *The man* was badly hurt.
b. Sean Penn attacked a photographer. *The man* must be deranged.
(Examples from Wilson, D. 1992. Relevance and reference, *UCL Working Papers in Linguistics* 4:167-192)

(2) Joanne$_i$ saw a foreign movie$_j$ yesterday$_k$ at the local cinema$_l$.
She$_i$ had a really good time/*It*$_j$ was sub-titled/#*It*$_k$ was bright and sunny/?#*It*$_l$/*The place*$_l$ had just been refurbished.

In (2), the first two argument referents introduced ('Joanne' and 'the foreign movie Joanne saw the day before the utterance of (2)') may be naturally continued via unaccented pronouns – but not the scenic referent 'the day before utterance time', nor (or at least, not as easily as with the first two entity referents evoked) 'the local cinema', which is expressed by an adjunct, and which serves as a locative frame of reference for the situation evoked as a whole. The slashes here are intended to indicate alternative continuations of the initial sentence. The crosshatch preceding an example is intended to signal that, as a potential utterance, it is infelicitous in the context at hand. Example (2) is intended to be discourse-initial, and not part of an earlier, ongoing discourse.

"Deixis", on the other hand, is a procedure which relies on the utterance context to re-direct the interlocutor's attention towards something associated with this context (hence which is potentially familiar to him or her), but to which s/he is assumed not already to be attending.[3] As Kleiber and other pragma-semanticists have observed, deixis causes a *break* in the continuity of the discourse at the point where the deictic procedure is used, so that the

[3] I am confining my remarks here to what Kleiber (1994) calls "opaque indexical symbols" (essentially, demonstrative pronouns and NPs), setting aside what he terms "transparent or complete indexical symbols", for example, first and second person pronouns. These latter "primary deictics" carry with them, by virtue of their use by a speaker, an automatic means for the identification of their referent, whereas of course the "opaque indexical symbols" I am dealing with here do not.

interlocutor is invited to "step out" of this discourse context to grasp a new referent in terms of the current situation of utterance – or alternatively an aspect of a same referent, which has already been focussed upon. Example (3) illustrates:

(3) [*Context: it is getting dark, and John and Mary are returning from a shopping trip. As John is parking the car, Mary exclaims*:]
Good God! Look at *that* incredibly bright light. [*Mary gestures towards a point in the evening sky*] What on earth do you think it could be?

Now, "textual" as well as "discourse deixis" provide a transition between the notions of anaphora and deixis, since they consist in using the deictic procedure to point to part of a pre-existing memory representation, but which is not necessarily highly activated. The interlocutor will therefore need to exert a certain cognitive effort in order to retrieve it.[4] (4) is an illustration of textual deixis (see also example (20) further on, illustrating discourse deixis):

(4) [*Denis Philps, introducing R. Langacker's paper at the conference on Linguistics and the English Language, Université de Toulouse-Le Mirail, 8 July 2000*]
I'd like to introduce Professor Ronald Langacker (… everyone says *that*!)…

2.2 "Direct" vs. "indirect" anaphora

As far as I am aware, the first use in print of the term "indirect anaphora" was made by Erkü and Gundel (1987) in the very title of their article (The pragmatics of indirect anaphors). Indirect anaphora is any use of the anaphoric procedure which does not consist in straightforwardly retrieving the referent of a prior linguistic mention from within the co-text (as is the case with example (5a)) – or of a subsequent one, in the case of cataphora; nor of a referent which is visible and salient within the situation of utterance (as is the case with example (5b)). (5a,b), then, are examples of "direct", not "indirect" anaphora.

(5) a. A young goat suddenly entered the open front door; but no-one could guess exactly what *it* was looking for.
b. [*Context: a young goat suddenly enters the open front door. A to B, observing the scene in fascination*:]
What do you think *it*'s looking for, exactly?

These two types of reference retrieval characterize, in the first case (textual retrieval) "anaphora", under the traditional view, and in the second (situational retrieval), "exophora". However, a memory conception such as the one I am

---

[4] At the same time, this interpretative effort will involve *constructing* an 'entity', on the basis of the discourse representation in question, in order for it to be the subject of a predication, an anchor for the introduction of new information. In (4) above it is arguably a 'stereotypical introduction of a public speaker' which is constructed from the antecedent-trigger predication via the use of the distal demonstrative.

adopting here does not retain this ‘geographical’ criterion as the defining condition of each sub-type: that is, referent located in the co-text or in the situational context. For what unites these two types of use is the speaker’s assumption that his or her interlocutor already has access to a discourse representation of the intended referent within his/her memory model of the discourse at issue (or that s/he can easily instantiate one via the context without undue processing cost), and that that representation is highly activated at the point of utterance. In any case, the expressions used to realize one or the other of these two types of use are the same in each instance.[5] The deictic procedure would not be appropriate in the case of exophora, just because the referent exists ‘outside the text’, in the situational context – unless it is a question of specifically directing the addressee’s attention towards a referent which is not yet in his or her attention focus.

Defined in this way, anaphora in the strict sense of the term (so-called “endophora”) and exophora (which I group together under the heading of “anaphora” *per se* – see (Cornish, 1999: Ch.4)) would correspond to “direct anaphora”: the intended referent is – in principle – immediately retrievable via its co-textual mention or via its physical presence in the utterance situation. We have to do with an instance of “indirect” anaphora in the following types of circumstance: when the anaphor does not retrieve the ‘basic’ referent directly evoked via a co-textual mention or via the interlocutors’ prior focussing their attention on an object or a scene in the situation surrounding them, but a different one which may be associated with it in virtue of a relation of the type part-whole, token-type, instance-class, or in terms of a metonymic relation of some kind.[6] However, I shall mainly be studying here the indirect pronominal anaphora linked with implicit internal arguments of predicates as “antecedent-triggers”. These arguments are what are responsible for introducing into the discourse the referents retrieved via the relevant pronoun (see (Cornish, forth.) for a study of implicit internal arguments in English and French). Here are some attested examples from French (6a-c), English (6d), German (6e) and Argentine Spanish (6f), all involving unaccented third person pronouns.

(6)	a.	Ah	dis	donc	maman	tu	t’	souviens	Cinéma Paradiso,
		Ah	say	then	mother	you	REFL	remember	Cinéma Paradiso,
		ben	**il**	a	fait	un	nouveau	film	
		well	he	has	made	a	new	film	

[5] Unaccented third person pronouns, definite or demonstrative nominal expressions, and so on.

[6] See (Reichler-Béguelin, 1993) and (Béguelin, 1998) for a large number of attested (spoken and written) examples of indirect pronominal anaphora in French, (Gundel *et al.*, 2000) for English, (Consten, 2003) for English and German, and (Ziv, 1996) for a certain number in both English and Hebrew. I take it that “associative” anaphora (see the title of Reichler-Béguelin’s article) is a sub-class of “indirect” anaphora.

’Hey, mum, you remember Cinéma Paradiso, well he’s made a new film’ (Spoken utterance, 26.10.90, ex. (65) cited in (Reichler-Béguelin, 1993))

b. [*Article about the disappearance on 17 March 1992 of Christiane, a woman of 62 suffering from Alzheimer’s disease*:]
… Christiane aurait été vue dans les Alpes de
… C. would have been seen in the Alpes de
Haute Provence. Même si vous n’ êtes pas sûr de vous,
Haute Provence. Even if you NEG are NEG sure of you(rself)
signalez-**le**…
report it
’C. is reported to have been sighted in the Alpes de Haute Provence department. Even if you’re not sure, report it (nonetheless)’ (*Carnets de Provence*, 1st August 1992, p.17)

c. [*Sticker on glass door of an Optician’s shop, Vence, France*:]
Si vous trouvez moins cher, je **les** casse!
If you find less expensive, I them break/undercut
’If you find (any) less expensive, I will break/undercut them!’

d. [*Fragment of dialogue in film*:]
Woman: Why didn’t you write to me?
Man: I did…, started to, but I always tore ’*em* up.
(Extract from the film *Summer Holiday*).

e. Ich ware wunschlos glücklich, wenn Sie nicht immer auf
I would-be completely happy if you not always on
dem Lehrerparkplatz parken würden. Das nächste mal lasse
the teachers’ parking lot park would. The next time have
ich **ihn** abschleppen.
I it[M.SG. ACC.] tow away
’I would be very pleased if you wouldn’t always park on the teachers’ parking lot. The next time it happens, I will have it towed away’
(Example taken from the TV soap *Beverly Hills 90210*, German RTL 31.8.95, cited in (Consten, 2001))

f. [*Vampire film on Argentine TV: a young boy is telling his friend what to do when face to face with the vampire woman*. BsAS, September, 2002:]
- Tenes el rosario?
You have.2SG the.M.SG rosary?
‘Do you have the rosary?’
- Si.
‘Yes’
- Bueno, cuando la veas, se **la** pones
Well, when her see.2SG REFL.3SG it.F.SG place.2SG
de una.
of one.F.SG
’Well, when you see her, you put it in front of her’
(Thanks to Erica Walz for this example)

In (6a), the referent of *il* (he) is intended to be accessible via a metonymic relation (film → director of film); that of *le* (it) in (6b) via the illocutionary point of the discourse as well as via the ellipsis in the antecedent-trigger predication, and similarly for the deliberately ambiguous pronoun *les* in (6c) (spectacles/prices of spectacles: the latter implicit referent is made available in this instance via the idiomatic phrase *casser les prix* (break (i.e. slash) prices (of commodities)). In (6d) it is again the illocutionary point of the woman's initial question, which bears on the non-existence of (a) letter(s) which she had expected the man to write to her, together with the lexical-semantic structure of the verbal predicate *write* (in the sense "engage in correspondence"), which provide an interpretation for the clitic pronoun *'em* in the third conjunct of the man's reply. The illocutionary point of the injunction in (6e) together with the lexical-semantic structure of the verbal predicate *parken* (to park), make available the referent of the pronoun *ihn* (it). As for example (6f), the pronoun *la* refers to *la cruz* (the cross) which the two boys (as well as the viewers) may be expected to be familiar with as an infallible weapon against the vampire. Its role in the situation evoked here is thus absolutely central. Erica Walz confirms that no cross was visible in the scene at issue, nor had one been explicitly mentioned in the co-text preceding this reference. It is thus an instance of indirect, and not direct anaphora (exophora).

3 The functioning of unaccented pronouns as indirect anaphors

3.1 The conception according to which "indirect" pronominal anaphora is ipso facto "marked" in relation to canonical anaphora

A number of linguists (e.g. Dik (1978), Sanford and Garrod (1981), Sanford *et al.* (1983), Gundel *et al.* (2000)) claim that indirect anaphora, especially that realized by unaccented pronouns (or *a fortiori* zero forms), forms which are specialized in retrieving extremely activated referents in psychological terms, is marked, marginal even, in usage in relation to direct or unmarked anaphora (Gundel *et al.*, 2000: 88). Dik (1978: 20), for example, states that: "Anaphoric reference to implicit antecedents is the exception rather than the rule", and claims that it is always more difficult to retrieve sub-lexical antecedents than explicit ones. This then leads us directly to the theoretical issue of "anaphoric islands"[7] (an "island", of course, is a place which is inaccessible by land).

(7) #Peter recently became an orphan, and he misses *them* terribly.

Note in this connection that this is not an absolute prohibition (i.e. the possibility in (7) of a retrieval via an unaccented pronoun (here *them*) of the implicit referent 'Peter's deceased parents'), since the pragma-semantic context may render this referent more accessible, more highly topic-worthy. A small adjustment to (7) would involve converting the anaphoric predication *and he*

[7] See Postal (1969) for the origin of this term.

misses them terribly, which is a conjunct, hence tightly connected grammatically to the antecedent clause, into a new conversational turn, uttered by a different speaker, as in (8):[8]

(8) A: Did you know that little Peter recently became an orphan?
B: Oh really? He must miss *them* terribly.

This change of turn, by making the two utterance acts partially independent of one another, enables a re-structuring of the information conveyed by the first utterance. In making this alteration, we have moved from the realm of "micro-syntax" (the syntacticized relation of coordination holding between the two clauses) in (7) to that of "macro-syntax" (two independent utterances, each corresponding to a distinct conversational turn) in (8) – to borrow the terms put forward by Berrendonner (1990); see also (Béguelin, 1998). These changes thus make available the conceptual space for an inference of the existence of Peter's deceased parents. In the discourse context in which the pronoun *them* occurs in (8), there does not seem to be any particular unnaturalness, as there is in (7). I am not claiming here that (8B) is perfect, simply that it is a good deal more natural than (7) – the level of accessibility of the intended pronoun's referent 'Peter's deceased parents' being that much higher.

As for Erkü and Gundel (1987), these authors claim explicitly that indirect anaphora (at least, of the types they discuss in their article) cannot be realized via pronouns. Witness their property III, p. 539:

(9) Neither type of indirect anaphora may be pronominal.

In their study, the authors distinguish three sub-types of indirect anaphora: (a) "inclusive anaphora"; (b) "exclusive anaphora"; and (c) "created anaphora".[9] Inclusive anaphora seems to correspond to what is commonly known as "associative anaphora",[10] since the introducing element (the antecedent trigger) always evokes a frame within which the anaphor will find

[8] Cf. also (Sproat & Ward, 1987) and (Ward *et al.*, 1991) for development of the same point here. Ward *et al.* (1991: 467, ex.(34)) cite an extract from a novel, where the concept of being an orphan is explicitly evoked, though the child in question's (now dead) parents are not (apart from his mother). Here, the latter are referred to via a subject pronoun, without a trace of unnaturalness: …***They*** *died when I was three.* (Elswyth Thane, *Ever After,* New York: Hawthorn Books, 1945: 155).

[9] Not examined here. Briefly, this subtype involves reference back to the event evoked by an antecedent-trigger clause (as in *Mary went from Paris to Istanbul by train and coach.* ***It/The trip*** *took a whole week.* Here, the subject pronoun *it* seems acceptable in the anaphoric clause, though Erkü and Gundel reject the pronoun in a similar example, claiming that only a full definite NP such as *the trip* is capable of assuming the required reference in their example).

[10] See Hannay's (1985) notion of 'sub-topics' within Functional Grammar in the case of English, as well as (Kleiber, 2001) for a recent account of associative anaphora in French.

its interpretation and reference. In (10), the anaphor clearly presupposes a whole-part relation between its antecedent's referent and its own.

(10) I couldn't use the box you gave me. The bottom/#it fell out.
(Erkü & Gundel, 1987: ex. (1))

"Exclusive" anaphora, on the other hand, introduces a partition within a more comprehensive set of entities[11] of which the referent of the trigger is a part, and another sub-set which includes the anaphor's referent:

(11) The ant daubs part of her burden onto a cocoon and passes the *rest*/#*it* to a thirsty lava.
(Erkü & Gundel, 1987: ex. (6))

Now, it's clear that no third person pronoun could possibly realize these two examples of indirect anaphora: for given that the use of this type of indexical expression is reserved for the anaphoric retrieval of highly activated referents, the "indirect" referents involved in (10) and (11) could never have this status. For by definition, those entities which "form part of a set", or "are associated with that set" in some way, will *ipso facto* not be in attention focus at the point when the set in question is evoked (the case of "the box" in (10)). Similarly, the entities which form the residue of a set or a mass of which only a part has been evoked previously (the case of "part of the burden borne by the ant" in (11)) will thereby not be in attention focus in the same way as the latter at the point of use. An unaccented pronoun or a zero form could not therefore retrieve them.

In the same vein, Sanford and Garrod (1981: 154, 161) state that what they call "explicit focus" within working memory[12] contains only representations of extremely active referents which have been explicitly introduced (via a linguistic mention, then) in terms of the co-text; whereas "implicit focus" houses only the representations of less salient entities which have been evoked implicitly inasmuch as they form part of the scenario retrieved from long-term memory ("semantic memory") in order to facilitate the understanding of a given text.[13] Initially, Sanford and Garrod (1981: 104) distinguished pairs of examples such as (12a,b) and (13a,b):

(12)a. "<u>Stated antecedent</u>"
Mary put the baby's clothes on.
The clothes were made of pink wool.

b. "<u>Implied antecedent</u>"
Mary dressed the baby.
The clothes were made of pink wool.

---

[11] In Erkü and Gundel's example presented as (11), this is a mass, in fact.
[12] The highly activated part of the workspace where incoming utterances are temporarily held and processed, in other words, short-term memory.
[13] This excessively powerful constraint is weakened somewhat in (Sanford *et al.*, 1983: 314), who extend the definition of explicit focus in characterizing it simply as "the current focus of attention", without mentioning the mode of entry into this space.

(13)a. Mary put the baby's clothes on. *They* were made of pink wool.

b. Mary dressed the baby. *#They* were made of pink wool.

An analogous pair of texts which these authors present is (14a) in contrast to (14b):

(14)a. Harry drove to London. *The car* broke down half-way.

b. Harry drove to London. #?*It* broke down half-way.

The conclusion reached by Sanford and Garrod is predictable: the referents 'the clothes which Mary put on the baby' in (13b) in the "implied antecedent" condition, and 'the car which Harry drove to London' in (14b) reside only in what they call "implicit focus", not having been explicitly introduced into the discourse via a textual antecedent. And it is for this reason, according to the authors, that only a semantically more explicit anaphor, such as the definite NP *the clothes* in the "implied antecedent" condition in (12b) and *the car* in (14a), would be capable of retrieving these "implicit" referents.

3.2 Distinction between two different cognitive-semantic statuses of candidate indirect referents for retrieval via an unaccented pronoun

However, just like the constraint on the realization of indirect anaphora postulated by Erkü and Gundel (1987), this restriction is too simplistic and thus insufficiently fine-grained to be able to grasp the true behaviour of indirect anaphors. For if we examine more closely the pragmatically deviant examples where the retrieval of an implicit referent is realized via a pronoun, it is clear that the reason for such exclusions lies elsewhere: as in Erkü and Gundel's examples (10) and (11), the "associative" referents 'the clothes' in (13b) and 'the car' in (14b) are not highly activated at the point where the verbs *dressed* and *drove* are processed by the reader. Indeed, the meaning definitions of the verbs *to dress* and *to drive* are, respectively, "to cover someone/oneself with clothes" (CAUSE x,p:(BE-COVERED x/y (BY-MEANS-OF z): (CLOTHES z))), and "to travel somewhere by motor-powered vehicle – by default, a car" (CAUSE x, p:(GO x (TO y) (BY-MEANS-OF z): (MOTOR-VEHICLE z))).

Now, it's clear that the elements "with clothes" and "by motor-powered vehicle" are non-nuclear arguments ("y-satellites", in Mackenzie's (1986) terminology – see below) in relation to the predicate in Dik's (1997) Functional Grammar terminology, and not "nuclear" arguments of the general verbs acting as functors within the lexical-semantic structure of these verbal predicates. According to Mackenzie (1986: 16-21), there is a "scale of intimacy" in terms of connectedness to the predicate: 1^{st} argument > 2^{nd} argument > 3^{rd} argument > y-satellites > z-satellites (1986:18). "Satellites" in FG are optional adverbial adjuncts; however, depending on the nature of the predication at issue, they may fall into different sub-categories. Thus, “y-satellites” (e.g. those of Manner) are claimed by Mackenzie to be implied with Action, Position (the expression of a controlled state) and Process predications; and Time and Location satellites are also said to be implied with Action predications, as are Duration satellites in durative states of affairs, and Frequency ones in momentaneous situations. However, satellites bearing the semantic functions Beneficiary, Instrument, Comitative, Cause, Circumstance, Result and Concession, etc. “are non-implied with Action predications” (Mackenzie, 1986: 18). This latter group would thus fall into the “z-satellite” subcategory.

Now, Mackenzie clearly takes the “y-satellite” subcategory to be a semantically-determined, syntactically-realizable aspect of predications as a whole, and not a potential part of the meaning definition of a given predicate. In standard FG, predicate-frame structures (specifying *inter alia* the number and type of arguments selected by a given predicate, as well as potential y-satellites) are subject to expansions eventually leading to a clause realization. As such, they are indicated as needing to be kept strictly separate from the meaning definition associated with each predicate frame in the Lexicon – in particular, no underlying-clause-structure expanding rule may operate upon units of the meaning definitions, even though predicate frames and meaning definitions are formulated using the same types of unit. Recently, however, a number of criticisms have been made of this strict stipulation, in the interests of economy of representation and derivation, of predictive power and of the transparent mapping between lexical semantics and syntax (see e.g. (García Velasco & Hengeveld, 2002; Mairal Usón & Faber, 2002; Cornish, 2002b)). Mairal Usón and Faber (2002) propose a semantically-transparent substitute for the orthodox FG predicate-frame in the shape of what they call “lexical templates”, where a partial semantic decomposition of a given predicate is formalized, along the lines of Van Valin and LaPolla’s (1997) approach.

Lexical templates include alongside the representations of the argument variables, what the authors call “internal variables”: these are the “semantic parameters which characterize an entire [lexical] class” of predicates, and can be seen to correspond lexical-semantically to Mackenzie’s (1986) “y-satellites”. They are encoded in lexical templates as Greek letters (in (15) below, a representation of the hyperonymic verb *cut*, these are α symbolising the cutting instrument, and β representing the particular manner of cutting). Lexical templates are intended to capture the specific lexical-semantic structure of given predicates, but insofar as they form part of a particular lexical class. Thus, basing their format for lexical templates on the lexical-semantic formalism used by Van Valin and LaPolla (1997), their representation of the hyperonymic verb *cut* is as follows (Mairal Usón & Faber, 2002: 55, ex. (14)):

(15)[[**do'** (w, [**use.sharp-edged.tool**(α)**in**(β)**manner'** (w,x)])] & [BECOME **be-at'** (y,x)]]
CAUSE [[**do'(**x, **[make.cut.on'** (x, y)])] CAUSE [BECOME **pred'** (y, (z))]], α = x.

This representation is of an effector *w*, first argument of a generalized activity verb (**do'**) who "uses a sharp-edged tool *x* in such a way that the tool becomes in contact with a patient *y*, causing an event such that *x* makes a cut on *y*, and this, in turn, causes *y* to become *cut*" (Mairal Usón & Faber, 2002: 55). The variable *z* is present in order to cater for situations where the final result is "further specified" (e.g. *into pieces* (as in (16a) below), *in strips*, or *open*). Clearly, then, the two "means" referents in (13b) and (14b) above, which are evidently part of the lexical-semantic representations of the verbal predicates *dress* and *drive*, respectively, would be represented in lexical template format as internal variables. They are thus presuppositional, background and not foreground components of the lexical meaning of these verbs.

Verbs of *cutting*, then, imply the presence of a knife (or other cutting instrument), and those of *painting* entail the presence of a paintbrush. As these instrumental objects act only as the means by which the activity described by the verb is manifested, they are not highly activated psychologically when the verb in question occurs in a text.[14] We thus find the following distribution of pronouns and definite, lexically-headed NPs in anaphoric clauses:

(16)a. Susan tried to cut the broiling chickens into pieces, but #*it*/*the knife* wasn't sharp enough.
b. George had difficulty in painting the vases: #*it*/*the brush* was too worn.

One possible test of internal argumenthood, versus "y-satellite" status, is the *do it* test. This shows that 'the broiling chickens' in (16a) and 'the vases' in (16b) are central participants in the acts of cutting and painting, respectively, but that 'the knife' and 'the paintbrush' are more peripheral ones: *Susan cut the broiling chickens into pieces: she also did it *the turkeys/she did it ?with a knife/with a kitchen knife; George painted the vases: he also did it *the chairs/ he did it ?with a paintbrush/with a decorator's paintbrush.* The question marks prefixing the continuations with the unmodified instrument phrases indicate that these referents are implied by the verb in each case.

If however we choose a verb or adjective whose lexical-semantic structure includes an implicit entity which is *nuclear*, a direct argument of the general functor defining its lexical-semantic structure, then we see that its retrieval via an unaccented pronoun is possible:

---

[14] See the experimental results in this direction obtained by Lucas *et al.* (1990).

(17)a. Joan is six months' pregnant with a new baby,
b. … and she has already knitted a bonnet and gloves for *it*.
(18)a. Joan is six months' pregnant,
b. … and she has already knitted a bonnet and gloves for *it*.

What differentiates the discourse fragments in (17) and (18) is that, in the first case, the baby with which Joan is pregnant is specified as being a "new" (i.e. subsequent) one, whereas in the second, the reference is non-determinate (non-definite, non-identifiable) though specific ('the baby which Joan is expecting'). What enables the unaccented pronoun *it* in (18b) to retrieve the argument 'Joan's baby' evoked via the antecedent-trigger utterance (18a) is the fact that the adjective *pregnant* means "to have conceived a baby", where 'a baby' is a nuclear argument in relation to the predicate *conceived* (CONCEIVED x, y: (BABY y)).[15] This does not constitute presupposed information within this lexeme, but foreground, essential information. As such, the entity 'the baby with which Joan is pregnant' would reside in the central attention focus space at the point where the second conjunct is processed (Sanford and Garrod's "explicit focus", and not their "implicit focus"). This status, as predicted, therefore enables its retrieval via an unaccented pronoun. What distinguishes cases like *pregnant* in (18) on the one hand, and cases like *cut* in (16a) and *paint*[16] in (16b) on the other, is that, unlike the former, the latter two predicates cannot occur with a null complement designating a specific, though unidentified referent (even though these referents may be contextually highly salient). After all, one can cut or paint all manner of physical objects, though when one (inevitably a woman) is "pregnant", it is necessarily with a human baby, a much more specific kind of entity.[17]

Here are one or two other examples of a similar type to (18a-b):

(19)a. John got married last week… *She*'s Swedish, if you want to know.
b. Professor Parker has been marking all morning. He's got *them* all piled up on his desk.
c. Paul has started smoking again. He seems to prefer *them* without filters.

One condition regulating this kind of retrieval is thus the degree of centrality of the referent at issue (nuclear argument of the general functor representing the lexical-semantic structure of the predicate concerned, and not a more peripheral "y-satellite"), within the pragma-semantic structure in terms of which the antecedent-trigger is represented in the discourse already constructed at the point of retrieval. Where there is too great a conceptual or referential

[15] See (Cote, 1998) for a similar 'lexical-conceptual' approach to implicit reference, exploiting Jackendoff's (1990) lexical-conceptual structure representations.
[16] Where these two verbal predicates also have available a nuclear internal argument.
[17] See (Cornish, forth.) for further discussion of this point.

"distance", or where there is a difference in topic-worthiness between the representation introduced by a trigger and the intended referent, the discourse-deictic procedure must be used, as in (20), an attested utterance (see also example (4)):

(20) [*End of the words of welcome uttered by the director of the Language Centre, at the start of a conference, University of Edinburgh, 19 September 1991*]
…We intend to record the guest speakers, so *these* will be available to participants at the end of the Conference…

In order to access the referent targeted via the proximal demonstrative pronoun *these* (namely, 'the recordings of the guest speakers' papers'), the hearer will have to draw an inference of the type: If the guest speakers' papers are recorded at time t_0, then at time t_n ($t_n > t_0$), there will be recordings of these papers. The existence of a morpho-lexical relation between the verb *to record* and the noun *recording* is not sufficient to enable an indirect retrieval via a pronoun – though such a regular relationship does act to speed up the inference leading to the existence of 'recordings of the guest speakers' papers'. Unlike the indirect referents in (6a-f), (18b) and (19a-c), here the implicit referent has not attained the status of a potential topic by the time the initial clause is processed, 'the guest speakers' enjoying this status at this point. It is thus predictable that the elaborative *so*-clause which immediately follows will continue to be about these entities. Like *that* in example (4), *these* in (20) orients the hearer's attention towards a referent which s/he must create on the basis of the representation introduced via the initial conjunct, as well as in terms of his or her knowledge of the world. It is thus an instance of discourse deixis rather than of anaphora. The personal pronoun *they* in its place would have maintained the situation evoked via the initial conjunct, resulting in the retrieval of the only salient topic-worthy entity within it, 'the guest speakers' – an interpretation leading to quite severe incoherence.

The zero internal-argument of *pregnant* in (18a) and of *married* in (19a) would not appear to correspond to Koenig and Mauner's (2000) notion "a-definite" (since a potential discourse referent IS in fact evoked when the relevant predicate is encountered). An "a-definite" is an implicit argument which does not evoke a discourse referent, since it is neither definite, nor indefinite, but partakes of both values simultaneously; all that such implicit arguments do is satisfy the argument position of the predicate which they "fill".[18] All non-definite implicit arguments are taken to correspond to "a-

[18] The authors' key example is the short passive, where the implicit internal argument is interpreted as the unspecified agent of the result of the action denoted: (1a) *A ship was sunk ø*. See Koenig and Mauner (2000) for further details of their notion of "a-definite".

definites", under this view (Jean-Pierre Koenig, p.c.). In principle, then, according to the authors, those anaphors which are meant to "retrieve" this argument would only do so via "accommodation".[19] Since definite NPs, which by that token carry a substantial descriptive component, may effect such an accommodation in this way, the "retrieval" is acceptable (see the "implied antecedent" condition of (12b) and (14a)); whereas with definite pronouns, this is more problematic, since their very meagre descriptive content is not sufficient to allow this: see (13b), (14b) and the "pronoun" condition of (16a,b). However, this hypothesis could not predict the acceptability of the pronominal retrievals in (6a-f), (18b) and (19a-c).[20]

3.3 The distinction, in terms of the possibilities of retrieval established in Section 3.2, between specific, and non-specific frame-bound or stereotypical "indirect" referents

One other relevant factor here is the nature of the "indirect" referent intended by the speaker itself: that is, whether it is specific, or non-specific frame-bound or a stereotypical accompaniment to a given state of affairs (see also (Gundel *et al.*, 2000: 94-6)). Given the non-specific character of indirect frame-bound indefinite or stereotypical referents, these may easily be targeted via an unaccented pronoun, even when these referents do not correspond to a nuclear argument of the antecedent-trigger verb. In the case of (6a, d, e and f), (8), (18b) and (19a), the indirect referents retrieved via the definite pronouns *il, 'em, ihn, la, them, it* and *she*, respectively, were specific, though not necessarily determinate (definite, identifiable).

On the other hand, those non-specific referents which form part of a stereotypical frame do not need to enjoy nuclear argument status, and thus to be in the foreground of the situation evoked via the trigger utterance. The "nuclear" vs. "peripheral" distinction is thus irrelevant in the case of referents of this type. Such a referent type is close to what Koenig and Mauner (2000) call "a-definites", where the implicit "a-definite" agent evoked via their key examples of short passives may be referred to via the "indefinite" pronoun *they*, as also in examples (21a, b and d) below; but it is not an *anaphoric* "retrieval", unlike the instances of specific or non-specific implicit arguments illustrated in the previous sections. It was precisely this type of referent which Sanford and

[19] In other words, via the introduction of the presupposition of the existence of such a referent.

[20] Notwithstanding, however, J-P. Koenig (p.c.) claims that it can, and that the hypothesis put forward in (Koenig & Mauner, 2000) would predict that the "accommodation" of the existence of a relevant discourse referent in such cases will have a processing cost, relative to the situation where the antecedent-trigger is lexically explicit. This is exactly what the experiment described in Section 4 of this chapter was designed to test.

Garrod (1981) had in mind when they postulated the existence of an implicit focus space, which according to them would contain the stereotypical referents associated with the scenario evoked via an utterance – e.g. the waiters in a restaurant, the clerks in a bank, etc. Several examples follow:

(21)a. The house on the corner of Edward Street was broken into last night, but *they* didn't take anything precious.
b. Mary was operated for cancer of the thyroid this morning. *They* conducted the operation masterfully.
c. #We went to a new restaurant in our area last night, but *she* was most uncooperative.
d. We went to a new restaurant in our area last night, but *they* were most uncooperative.

4 The form of an experiment designed to test this hypothesis, and its results

A self-paced reading experiment designed to test the psychological reality of the existence of two types of "indirect" or implicit referents, as argued for in this chapter, has recently been carried out.[21] The rationale behind this experiment is as follows: as we have seen, the two types of indirect referent at issue are (1) those which correspond to a central, nuclear ingredient of the discourse representation targeted by the (pronominal) anaphor, and (2) those which form part of it only in a more peripheral sense – corresponding to the *means* by which the situation is set up via the predicate itself, or to an *expected accompaniment* to it. I have mainly illustrated this distinction in terms of the lexical-semantic structure of given predicators (verbs and adjectives), in similar fashion to (Cote, 1998) – so clearly, as Cote also points out, implicit arguments of the predicates corresponding to given, potentially transitive verbs and adjectives must be taken into account in establishing potential discourse referents.[22]

To test the psychological reality of this distinction, then, we proposed to measure the reading times of anaphoric predications oriented towards referents which have not been explicitly introduced into the discourse, but which are nevertheless not inferable on the basis of a morphological connection between the antecedent-trigger and the expression in terms of which the anaphor will receive an interpretation (cf. the pair *a guitarist…the guitar/it*). Only

[21] The two versions of the experiment were designed, prepared and conducted in collaboration with François Rigalleau (Université de Poitiers) and Marion Fossard (Université de Toulouse-Le Mirail) for the French one, and with Alan Garnham, Wind Cowles (both at the University of Sussex) and Marion Fossard for the English one. See (Cornish *et al.*,submitted) for further details of this two-fold experiment.

[22] Cote argued that these should be counted as part of the forward-looking centre (or Cf) list within a given utterance for the Centering Theory algorithm (for an introduction to Centering Theory, see (Walker *et al.*, 1998)). Similarly, Mauner *et al.* (2002) showed experimentally that subjects access participant information as soon as a given verbal predicate is encountered in an incoming utterance, whether the participants in question are syntactically realized or not.

unaccented (clitic, in French) non-subject pronouns used anaphorically were tested here. The central implicit referents were introduced in virtue of the lexical meaning (predicate-argument structure) of the trigger involved – for example, *be pregnant*, *write* (in the "correspond" sense), or *get married* –, or as a function of relevant world knowledge (e.g. 'a burglary', 'St. Valentine's Day', etc.).

Twenty four experimental texts consisting of two conversational turns, in a familiar genre characteristic of spontaneous conversation (dialogues) were constructed and divided into four Conditions. The subjects tested were made aware in advance that the texts they were about to read fell within the genre of spontaneous spoken discourse, so that they would expect to encounter a type of unplanned language which does not correspond to normative written prose. The texts were constructed by crossing the two variables chosen for the experiment: *type of referent* (nuclear vs. peripheral) and *type of antecedent-trigger* (implicit vs. explicit "trigger"). The crossing of these variables produced the four experimental conditions used here.

The first Condition, then, consisted of an initial utterance in which a referent is evoked implicitly as a central participant in the situation denoted. This initial utterance was followed immediately by a target utterance belonging to a different conversational turn, and including a non-subject pronoun which retrieved the implicit nuclear entity evoked via the initial utterance.

The second Condition consisted of a set of nearly parallel dialogues in relation to those of Condition 1 (same initial utterance, same target utterance with an identical pronoun); however, the first turn contained two utterances, the first of which was identical to the initial utterance of Condition 1, and the second consisted of an explicit lexical evocation of the referent which was to be retrieved. Moreover, the antecedent-trigger was always introduced in these utterances in subject position. The distance in terms of number of words was held constant between the mention of the trigger in the second utterance of the first turn, and the pronoun in the second turn. These four factors (two utterances for the first turn, the referent to be retrieved being introduced explicitly via the subject function as well as in the second of the two initial utterances, and the distance between occurrence of trigger and resumptive pronoun being held constant) were identical in the two "explicit" conditions (2 and 4).

Condition 3 consisted once again of a near parallel set of dialogues in relation to those of Condition 1, but this time the pronoun in the second turn was oriented towards a non-central, peripheral participant which might be evoked implicitly via the antecedent-trigger used. This entity was either a non-nuclear argument, or an instrument, or a stereotypically expected accompaniment to the situation denoted.

Finally, as with Condition 2, Condition 4 acted as a Control, where the referent of the indirect pronoun of Condition 3 was introduced explicitly in the antecedent-trigger utterance. As in the case of the target utterance of Condition 2 in relation to that of Condition 1, the target utterance of Condition 4 was identical to that of Condition 3.

See the Appendix below for a sample of these materials taken from the English version of the experiment, in the four Conditions selected, as well as Cornish et al. (submitted) for further details of the experiment.

The predictions then were that, although the Reading times of the target utterances (those containing the pronoun at issue) would be different in each of the four conditions – lower in the case of the explicit textual evocations of Conditions 2 and 4 than in those of the implicit evocations of 1 and 3 -, the differences between Conditions 1 and 2 would not be significant. By contrast, this difference was predicted to be more noticeable, and significant, in the case of Conditions 3 and 4. Moreover, we predicted that the Reading times would be higher (significantly so) in Condition 3 than in Condition 1 – the two "implicit" conditions.

As will be evident, the main results of this experiment (the mean Reading times of the target (pronominal) utterances) fully bore out these predictions. They are given in Table 1 below.

Condition	**French**	**English**
Nuclear-Implicit	2953	2375
Nuclear-Explicit	2743	2218
Peripheral-Implicit	3548	3057
Peripheral-Explicit	2831	2250

Table 1: Reading Times in msecs of the pronominal (target) utterances in French and English

The French materials were submitted to 20 native-speaker subjects from the University community at the University of Poitiers, in self-paced, auto-segmented reading mode on a micro-computer. Similarly, the English materials were administered to 20 native-speakers from the University community at the University of Sussex under identical conditions. The results as between the French and English data are remarkably similar. In both sets of results, there was a statistically significant interaction between the two factors of referent-type and antecedent-trigger, with faster reading times for utterances containing references to implicit triggers when the referent was nuclear than when it was peripheral. Conversely, there was no difference in reading times for nuclear and peripheral referents in the case of explicit antecedent-triggers. This shows a clear effect of the "nuclear" vs. "peripheral" status of indirect referents, as predicted. Further, reading times for utterances containing peripheral referents were faster when the antecedent-trigger was explicit than when it was implicit, while no such difference was found for the utterances with nuclear referents.

5 Conclusion

In conclusion, then, it is clear that "explicit focus" within working memory is not limited to representations of entities which have been explicitly introduced by lexical means into the discourse. This criticism can be levelled not only at Sanford and Garrod's initial hypothesis, but also at the standard Centering Theory approach to establishing the forward-looking centre rank-list for a given utterance within a discourse segment, which is in effect based solely upon the explicit mention of given referents within a co-text. We can retain the partition proposed by Sanford and Garrod between "explicit focus" and "implicit focus",[23] but place the dividing line elsewhere: central focus, equivalent to the cognitive status "in-focus" of Gundel *et al.* (1993, 2000), will include referents and denotata – and the situations in which they are involved – which are introduced linguistically via *nuclear* NPs and PPs (subject, direct and indirect object functions), or via predicative phrases. Nuclear arguments within the lexical-semantic structure of adjectives and verbs which are non-realized linguistically, of the kind we have seen in this chapter (see examples (6a-f)), as well as referents introduced perceptually via the interlocutors' focussing their attention on an object or an event within the situation of utterance (see (5b)), will also figure here.

The referents or denotata associated with embedded PPs and NPs will not be in central focus, even if they are introduced linguistically: as an illustration, see the scenic adverb *yesterday* and the PP adjunct *at the local cinema* in example (2) above. Nor will the referents or denotata associated with modifying phrases (e.g. bearing an epithet function). These last referents/denotata will reside in peripheral focus (the cognitive status "activated" in the Gundel *et al.* (1993, 2000) model): they are clearly not easily retrievable or accessible via an unaccented overt pronoun or a zero form. Moreover, those referents peripherally associated with a given referent which has been introduced explicitly or which is the target of perceptual attention-focussing on the part of

---

[23] Though these somewhat inappropriate terms might now more accurately be replaced by the terms "central focus" and "peripheral focus", respectively.

the interlocutors, will also reside in this less central storage space within working memory. See in this respect examples (4) and (20), which both involve a demonstrative pronoun. Their instantiation will necessarily be the outcome of a (semi-)conscious, and not automatic, inference, as in the case of the central arguments – this inference expressing itself in terms of a processing cost, since the representation targeted will have actually been created as a potential discourse referent *via* this discourse-deictic reference.

References

Béguelin, M.-J. 1998. “Anaphores non Catégorisantes, Structures Lexicales, et Cognition.” P. Valentin and M. Fruyt (eds.) *Lexique et Cognition*, Paris: Presses de l’Université de Paris. 9-32.

Berrendonner, A. 1990. “Pour une Macro-syntaxe.” *Travaux de Linguistique* 21.25-36.

Consten, M. 2001. “Indirect Anaphora and Deixis in Texts: Domain-bound Reference and Coherence.” *Int. Workshop on Reference and Coherence*, University of Utrecht, Poster.

---------- 2003. “Towards a Unified Model of Domain-bound Reference.” Friedrich Lenz (ed.) *Deictic Conceptualisation of Space, Time and Person*, Amsterdam: John Benjamins. 223-243.

Cornish, F. 1999. *Anaphora, Discourse and Understanding. Evidence from English and French.* Oxford: Clarendon Press.

---------- 2002a. “Anaphora: Lexico-textual Structure, or Means for Utterance Integration Within a Discourse? A critique of the Functional Grammar Account.” *Linguistics* 40.469-493.

---------- 2002b. “‘Downstream’ Effects on the Predicate in Functional Grammar Clause Derivations.” *Journal of Linguistics* 38.247-278.

---------- 2003. “The Roles of (Written) Text and Anaphor-type Distribution in the Construction of Discourse.” *Text* 23.1-26.

---------- (forth.). “Implicit Internal Arguments, Event Structure, Predication and Anaphoric Reference (in preparation).” N. Hedberg and R. Zacharski (eds.) *Topics on the Grammar-Pragmatics Interface: Papers in Honor of Jeanette K. Gundel*. Amsterdam: John Benjamins.

----------, A. Garnham, W. Cowles, M. Fossard, and V. André(submitted). “Indirect Anaphora in English and French: A Cross-linguistic Study of Pronoun Resolution.” Submitted to *Journal of Memory and Language.*

Cote, S. 1998. “Ranking Forward-looking Centers.” Marilyn. A. Walker, Aravind. K. Joshi and Ellen. F. Prince (eds.) *Centering Theory in Discourse,* Oxford: Clarendon Press. 55-69.

Dik, S. 1978. *Stepwise Lexical Decomposition*. Lisse: Peter de Ridder Press.

---------- 1997. *The Theory of Functional Grammar Part I: the Structure of the Clause* (2nd edition). Berlin: Mouton-de Gruyter.

Erkü, F. and J. Gundel 1987. The Pragmatics of Indirect Anaphors. J. Verschueren and M. Bertuccelli-Papi (eds.) *The Pragmatic Perspective: Selected Papers from the 1985 Int. Pragmatics Conference*, *Amsterdam 1987*. Amsterdam: John Benjamins. 533-545.

García Velasco, D. and K. Hengeveld 2002. “Do We Need Predicate Frames?” Mairal Usón and M. J. Pérez Quintero. 95-123.

Gundel, J., N. Hedberg and R. Zacharski 1993. “Cognitive Status and the Form of Referring Expressions in Discourse.” *Language* 69.274-307.

----------, ---------- and ---------- 2000. “Statut Cognitif et Forme des Anaphoriques Indirects.” *Verbum* XXII.79-102.

Hannay, M. 1985 “Inferrability, Discourse-boundness, and Sub-topics.” A. Machtelt Bolkestein, Casper de Groot and J. Lachlan Mackenzie (eds.) *Syntax and Pragmatics in Functional Grammar*. Dordrecht: Foris. 49-63.

Jackendoff, R. 1990. *Semantic Structures*. Cambridge, MA: MIT Press.

Kleiber, G. 1994. *Anaphores et Pronoms*. Louvain-la-Neuve: Duculot.

---------- 2001. *L’Anaphore associative*. Paris: Presses Universitaires de France.

Koenig, J.-P. and G. Mauner 2000. “A-definites and the Discourse Status of Implicit Arguments.” *Journal of Semantics* 16.207-236.

Lucas, M., M. Tanenhaus and G. Carlson 1990. “Levels of Representation in the Interpretation of Anaphoric Reference and Instrument Inference”. *Memory & Cognition* 18.611-631.

Mackenzie, J. 1986. “Aspects of Nominalization in English and Dutch”. *Working Papers in Functional Grammar* 15.1-32.

Mairal Usón, R. and P. Faber 2002. “Functional Grammar and Lexical Templates.” Mairal Usón and Pérez Quintero. 39-94.

----------, P. Quintero and M. Jesús (eds.) 2002. *New Perspectives on Argument Structure in Functional Grammar*. Berlin: Mouton-de Gruyter.

Mauner, G. and M. Tanenhaus 1995. “Implicit Arguments in Sentence Processing.” *Journal of Memory and Language* 34.357-382.

----------, A. Melinger, J.-P. Koenig and B. Bienvenue 2002. “When is Participant Information Encoded? Evidence from Eye-monitoring.” *Journal of Memory and Language* 47, 386-406.

Postal, P. 1969. “Anaphoric Islands.” *Papers from the 5th Regional Meeting of the Chicago Linguistic Society*. 205-236.

Reichler-Béguelin, M.-J. 1993. “Anaphores Associatives Non-lexicales: Incomplétude Macrosyntaxique?” S. Karolak and T. Muryn (eds.) *Complétude et Incomplétude dans les Langues Romanes et Slaves. Actes du 6è Colloque International de Linguistique Romane et Slave*. Crakow. 327-379.

Sanford, A. and S. Garrod 1981. *Understanding Written Language. Explorations in Comprehension Beyond the Sentence*. Chichester: John Wiley.

----------, ----------, A. Lucas and R. Henderson 1983. “Pronouns Without Explicit Antecedents?” *Journal of Semantics* 2.303-318.

Sproat, R. and G. Ward 1987. “Pragmatic Considerations in Anaphoric Island Phenomena.” E. Need, E. Schiller and A. Bosch (eds.) *Papers from the 23rd Annual Regional Meeting of the Chicago Linguistic Society*. Chicago: CLS 23. 321-335.

Van Valin, R. and R. LaPolla 1997. *Syntax: Structure, Meaning and Function*. Cambridge: Cambridge University Press.

Walker, M., A. Joshi and E. Prince 1998. “Centering in Naturally Occurring Discourse: An Overview.” M. Walker, A. Joshi and E. Prince (eds.) *Centering Theory in Discourse*. Oxford: Clarendon Press. 1-28.

Ward, G., R. Sproat and G. McKoon 1991. “A Pragmatic Analysis of So-called Anaphoric Islands.” *Language* 67.439-473.

Ziv, Y. 1996. “Pronominal Reference to Inferred Antecedents.” *Belgian Journal of Linguistics* 10.55-67.

Appendix

Sample of test materials used in the English version of an experiment designed to study the processing of two types of implicit referents

<u>Condition 1</u> (Implicit antecedent + Nuclear referent)

A: Have you noticed that Mark isn't shaving?

B: *Target sentence*: Yes, in fact he's really allowing it to grow now.

Statement: Mark does seem to be growing a beard. (TRUE)

<u>Condition 2</u> (Explicit antecedent + Nuclear referent)

A: Have you noticed that Mark isn't shaving? His straggly beard makes him look like a tramp.

B: *Target sentence*: Yes, in fact he's really allowing it to grow now.

Statement: Mark is sharply reducing the length of his beard. (FALSE)

<u>Condition 3</u> (Implicit antecedent + Peripheral referent)

A: Have you noticed that Mark isn't shaving?

B: *Target sentence*: Yes, he tells everyone he's thrown them all away.

Statement: Mark has clearly decided to get rid of his razors. (TRUE)

<u>Condition 4</u> (Explicit antecedent + Peripheral referent)

A: "Have you noticed that Mark isn't shaving? His disposable razors have all completely disappeared."

B: *Target sentence*: "Yes, he tells everyone he's thrown them all away."

Statement: Mark always uses an electric razor for shaving. (FALSE)

Pronominal Interpretation and the Syntax-Discourse Interface: Real-time Comprehension and Neurological Properties[1]

Maria Mercedes Piñango and Petra Burkhardt
Department of Linguistics, Yale University

We report on the dynamic implementation and neurological distribution (through the study of Broca's aphasia) of certain properties of pronominal interpretation. Three pronominals – reflexives, logophors and pronouns – are examined within a model of interpretation that capitalizes on the interaction between two levels of representation – syntax and discourse – relative to one factor: coargumenthood. Our model hypothesizes that coargumenthood dissociates reflexives from logophors and pronouns by determining the level of representation where the antecedent is identified: presence of coargumenthood between the pronominal and the antecedent allows syntax to determine pronominal-antecedent dependency. Absence of coargumenthood forces the system to search for the antecedent at the level of discourse. The latter is predicted to require more computational resources. Results show that at least for the logophor, antecedent identification increases the amount of computational resources used. The same distinction is revealed in Broca's patients' comprehension thus supporting the view that the processing of these pronominals is only partially dependent on syntactic processing.

1 Introduction

Pronominals such as pronouns (e.g. *him* in 'Bill$_i$ blamed him$_{*i}$') and reflexives (e.g. *himself* in 'Bill$_i$ blamed himself$_i$') are fully formed morphosyntactic elements which nonetheless lack full referential content and as a result demand access to information either in the local sentential context or in the general context of the utterance for the purposes of interpretation. The research presented here is concerned with how this process of interpretation takes place in terms of both dynamic implementation and cortical distribution. This is done through the observation of pronominal interpretation as sentence comprehension unfolds in both normal and lesioned systems (i.e. Broca's aphasia).

Following Reinhart and Reuland (1993) and Reuland's (2001a) approach to pronominal interpretation (i.e. Reflexivity theory), we take the presence or absence of *coargumenthood* between the pronominal and a potential antecedent (i.e. where a potential antecedent is a nominal matching the pronominal in

[1] The research reported here was supported by a grant to MMP from the Social Science Research Fund (Yale University). We thank Sergey Avrutin, Eric Reuland, the audiences at UMass Amherst-Linguistics, Euresco 2002 and DAARC 2002 for insightful discussion on the ideas presented here and to four anonymous reviewers for their comments and recommendations.

gender, number and person) as the fundamental factor that distinguishes among the various processes involved in pronominal interpretation. This factor alone creates a categorical distinction not only between reflexives and pronouns, but also between reflexives and a different type of SELF anaphor, namely, logophors. Logophors are defined here as pronominal elements (pro-forms with the syntax of noun phrases) that contrary to reflexives are not in complementary distribution with pronouns and do not hold a coargument relation with their antecedent. Examples (1a) and (1b) below illustrate the distribution of these three pronominals as a function of coargumenthood:[2]

(1) a. The lawyer$_i$ who was young defended himself$_i$/*him$_i$ [refl/*pron]
    b. The daughter$_i$ hid a present behind herself$_i$ /her$_i$…[logophor/pron]

In (1a) *himself* and lawyer are coarguments of *defended*, and the reflexive is in complementary distribution with the pronoun. By contrast in (1b), HERSELF is not in a coargument relation with its antecedent (HERSELF is an argument of BEHIND) and it is in non-complementary distribution with the pronoun.

As can be seen, coargumenthood allows us to dissociate reflexives on the one hand from logophors and pronouns on the other. It is our claim that such a dissociation carries over in the way these elements are processed in real-time, and in the way the underlying processes are cortically distributed. In what follows, we explain the model of pronominal interpretation that predicts this dissociation. This approach thus allows us on the one hand, to render supporting evidence to an approach to anaphora resolution, and on the other hand, it enables the beginning of the research line that seeks to connect the language system from all possible perspectives.

1.1 Coargumenthood in a model of pronominal interpretation

Our model takes the language system to include a set of correspondence rules that connect syntactic and discourse representations. 'Interpretation' from this perspective is the result of a specific syntactic configuration connected to discourse and semantic representations via a specific set of independent correspondences. According to this model, syntactic representation is connected to discourse via at least two functional projections, Determiner Phrase and Tense Phrase. These two projections "introduce" or have correlates at the level

[2] This use of the term "logophor" does not exclude those better known instances of logophoricity where both lack of coargumenthood in addition to distance play a part e.g. *Max boasted that the queen invited Mary and himself for a drink* (Reinhart & Reuland, 1993:670). Moreover, an additional discourse property on which we are not directly capitalizing here is that logophoric elements are also characterized by having a PIVOT role in the sentence (Sells, 1987) and by being coreferential with the 'internal protagonist of the sentence or discourse' (Huang, 2000:172-173). Yet, in all sentences considered here, the antecedent is a PIVOT/internal protagonist of the sentence.

of discourse called, respectively, individual file cards and event file cards (e.g. Heim, 1982; Avrutin, 1999, 2002; Piñango *et al.*, 2001). An *individual* file card represents the discourse correlate of a DP and an *event* file card the discourse correlate of a TP. All file cards have two types of information: a *frame*, a placeholder of discourse information (i.e. it tells the system that it is a constituent with a corresponding unit in syntax), and a *heading*, which tells the system the referential content of that unit. This content includes all the information necessary to identify the participant or event among other file cards in discourse representation (see also (Kamp & Reyle, 1993) for a different implementation of the same insight).

Pronominals have the syntax of DP's and they are therefore represented as individual file cards in discourse representation. As mentioned, however, the file card that is triggered by the $DP_{pronominal}$ is by definition referentially deficient. That is, it lacks the necessary information that enables the semantic system to connect that file card with an entity in the mental model. In terms of our model, this means that the individual file card that is triggered by the pronominal, contains a *frame* (i.e. it is identifiable as corresponding to an individual) but no *heading* (i.e. has no referent, no entity in semantic representation to which it unambiguously corresponds). That is the case of a pronoun like *her*. *Her* has a frame, as it tells the system that there is an individual in the discourse, but it has no heading, as its referent remains unknown. We encode this 'link' in the syntax-discourse correspondence system by having the frame generate in D^o and the heading generate in the complement NP.

It is also possible that a file card will have a frame and a heading, but that the heading be incomplete. That is the case of SELF-anaphors like *herself*, where *her* introduces the frame and *self* the heading. But this heading behaves like a variable as it is itself referentially deficient (SELF = *stand for* x, where x = antecedent) (e.g. (Reuland, 2001a; Avrutin, 2002)).

This situation, where the heading is empty or partially filled, presents a problem for the comprehension system, since a file card must have both a frame and a heading in order to be interpreted. A heading-less file card is therefore untenable. Such a conflict must be resolved by "filling in" the heading of the file card. This means that a connection must be established between the pronominal file card and a potential antecedent in order to obtain the necessary referentiality information. Two alternative mechanisms are available to this end: 1) fill the heading guided by syntactic mechanisms alone, 2) fill the heading guided by syntactic and discourse mechanisms (e.g. Reuland, 2001a).[3] Based on our model, the former can only occur when the antecedent and the pronominal are in a coargument relation whereas the latter *must* occur when there is no coargument relation between the potential antecedent and the pronominal (see also (Reinhart & Reuland, 1993; Reuland, 2001a; Avrutin, 2002) for independent linguistic and psycholinguistic evidence for this notion).

The syntax-alone mechanism leads to a bound variable interpretation (i.e. *John criticized himself*[4]). Here, both the antecedent and the anaphor trigger their corresponding file card. The antecedent's file card is fully referential, but the anaphor's is not. It lacks a fully filled heading. In this case the situation is resolved via a dependency which is established at the level of syntax. (e.g. (Farmer & Harnish, 1987; Reinhart & Reuland, 1993:661)). This dependency is possible due to the coargument relation that exists between the antecedent and the anaphor. This dependency allows the filling in of the heading to occur: *John* likes *himself*; SELF=*stands for* John.

The second mechanism underlies both pronoun and logophoric interpretation as in the sentence: *John put a blanket around himself/him*. In cases like this, there is no syntactic mechanism that can be used as a guide in finding the antecedent because there is no coargumenthood between the pronominal and the antecedent. Consequently, the necessary dependency between the antecedent and the pronominal can only be formed at the level of discourse. The mechanism is as follows: both antecedent and pronominal trigger their corresponding file cards; the pronominal's file card is deficient so a dependency with an antecedent must be established so that the sentence can be interpreted. Absence of coargumenthood makes it impossible for syntax to guide in the identification of the antecedent. As a result, such identification must be made at the level of discourse. At this level, an antecedent is identified using the frame information, and other discourse relevant factors such as whether the potential antecedent is the PIVOT of the sentence as in the case of logophors (see footnote 2) or whether or not the antecedent belongs to the same event representation, as in the case of pronouns (Reuland, 2001b; Avrutin, 2002).[5]

In sum, coargumenthood, established over syntactic representation imposes a constraint on pronominal distribution which informs a model of the syntax-

[3] A fully formed syntactic representation is assumed for the formation of a discourse representation. In this case, syntactic representation serves to determine that the pronominal is not in a coargument relation with a potential antecedent. So, the distinction being drawn here is between the interpretation that requires syntax alone and the one requiring discourse *in addition to* syntactic representation.

[4] In these examples, underlining indicates coreference.

[5] In our model, it is this the only factor that prevents a pronoun to be coreferential with a coargument as in **John likes him*.

discourse interface by determining the level of representation where the various interpretive dependencies are formed. In particular, absence of this property forces the system to establish an antecedent-pronominal dependency at the level of discourse for interpretation of the pronominal to take place. By contrast, presence of coargumenthood allows a bound-variable interpretation which is fully determined within syntactic representation. That is, the connection to the antecedent takes place within syntax, Discourse only allows the transfer of the necessary information.

Here, we explore the hypothesis that these two different interpretive mechanisms -syntax-based dependency and discourse-based dependency – are observable as sentence comprehension unfolds, thus making the representational considerations outlined above directly relevant to our notions of the architecture of the language system from a dynamic perspective, and vice versa. In what follows, we present experimental evidence to support both this hypothesis and the model in which it is couched.

1.2 Pronominal interpretation and real-time comprehension

Previous evidence has shown that the processing of elements that are independently claimed to require access to discourse representation – in order for full interpretation to be achieved – engages more 'computational resources' than interpretation that is constrained by syntactic principles alone (i.e. bound-variable interpretation).[6]

As mentioned, the notion of "cost" on which we capitalize here is observed in the form of a significantly higher reaction time (relative to the interpretation of a non-discourse demanding element) or in related work in a delay in the priming effect for the interpretation of discourse-demanding elements. That is the case of the so-called discourse-linked pronouns ("Wh-NP" pronouns) relative to non discourse-linked pronouns ("who" pronouns) (De Vincenzi, 1996; Shapiro, 2000). Discourse-linked pronouns have been shown to elicit a priming effect just like the "who" counterparts, but at a comparatively delayed time. This line of research has further revealed that the interpretation of pronouns with non-quantified antecedents (e.g. the boy) also elicits a higher reaction time as compared to the interpretation of pronouns with quantified

[6] The notion of 'computational resources' is intended to convey the widely held assumption that the process of comprehending or producing language requires mental work or computations. Moreover, the resources from which the computations are drawn are taken to be finite. This predicts that the system will carry 'more economical' computations before 'costlier' ones (e.g. (Reuland, 2001b:353)). The model presented here capitalizes on this distinction by proposing that the "use" of computational resources is observable behaviorally through latency measures. In terms of comprehension, these measures can come in the form of reaction time from, say, reading times or a lexical decision task (as in the work presented here).

antecedents (e.g. everyone), which are hypothesized to trigger a bound-variable interpretation (Piñango *et al.*, 2001).

This distinction between syntactic and discourse-based interpretation has also received support from the acquisition and aphasia literature. This research has shown that both children and Broca's patients exhibit significantly poorer comprehension for elements postulated to require access to discourse representation (e.g. pronouns with non-quantified antecedents, logophors) over those whose interpretation is obtainable through syntactic mechanisms alone (e.g. pronouns with quantified antecedents and reflexives) (e.g. (Avrutin, 1999; Grodzinsky & Reinhart, 1993; Grodzinsky *et al.*, 1993; Piñango, 2001)). Finally, discourse-based interpretation has also been involved in neurological models of pronominal interpretation. In this respect, Harris *et al.* (2000) report that processing of logophors (i.e. which require a discourse-based dependency) elicits electrical patterns of activation akin to those associated with problems with semantic interpretation (i.e. reanalysis). This suggests that even at the level of electrical patterns, discourse-based processing behaves in a manner that can be at least partly dissociated from syntactically-based interpretation.

All this evidence taken together motivates a basic architecture of the language system where dependencies are distinguishable by the level of representation in which they are formed. These levels of representation are nonetheless connected, and as we claim here, observable as comprehension unfolds.

Consequently, and having established a distinction between syntax-based interpretive mechanisms and discourse-based ones, we begin by investigating the interpretive properties of pronouns as compared to reflexives. Moreover, we ask how pronominal elements which allow, in addition to intra-sentential antecedents, the possibility of extra-sentential antecedents (i.e. "some person") is dealt with within this system. This question is investigated in Contrast 1.

The second contrast represents evidence where the connection between coargumenthood and pronominal interpretation is examined by contrasting pronominals which are lexically alike but which have different distributional properties. That is the case of logophors vs. coargument reflexives. These two experiments are discussed directly below.

2 Dynamic Implementation of Pronominal Interpretation

2.1 Contrast 1: Coargument Reflexive vs. Pronouns

This first contrast is examined using the cross-modal lexical decision interference paradigm (e.g. (Shapiro *et al.*, 1989; Piñango *et al.*, 1999)). In this methodology, two tasks are performed: listening and understanding a sentence (primary task), and performing a lexical decision (secondary task). Sentences are presented auditorily over headphones. At some point during the presentation of the sentence (a point independently determined to be of theoretical relevance), a letter string is presented visually on a computer screen. The subject is asked to indicate whether or not the letter string presented is a word of English by pressing a "yes" button on a response box. The dependent variable for this task is the reaction time (RT) it takes the subject to make her/his lexical decision. As we are working within an interference paradigm, the meaning of the letter string (probe) presented with each of the experimental sentences is completely unrelated to the content of the sentence. Finally, in order to ensure that the subject understands the sentences presented (primary task), the tape is stopped at random intervals during the session and a question about the sentence just heard is asked. In Experiment 1, the tape was stopped about 20 times during a session.

The logic of this paradigm is as follows. Linguistic activity demands computational resources. When one compares the comprehension of two conditions hypothesized to require the same amount of resources (e.g. computationally equivalent sentences), and a secondary task is simultaneously carried out, reaction times for the secondary task should not differ between the two conditions. However, if one of the conditions requires more resources (in the present case as a result of establishing a dependency at the level of discourse representation between the pronominal and the antecedent), this will be reflected in a higher RT for the corresponding secondary task as compared to the RT for the secondary task of the hypothesized "less costly" counterpart. That is, both conditions are still competing for the same amount of resources, but the one requiring more of these resources will in turn decrease the resources available for the secondary task; hence a higher RT. Within the context presented, a higher RT for a condition is interpreted as an instance of "cost".

For each condition, 25 sentences were created with the following constraints: 1) there had to be the same distance in terms of number of words and number of syllables between the pronominal and the potential antecedent; 2) both verbs used had to have similar frequency, and if there was going to be a difference, the highest frequency verb would be the one belonging to the coargument reflexive condition. These experimental sentences were inserted quasi-randomly in a script containing 150 filler sentences, for a total of 200 sentences.

Together with the experimental sentence pairs, fifty experimental probes were created. Probes for each sentence pair were matched in terms of number of syllables and letters, as well as frequency (Francis & Kucera, 1982). Probes were created so as not to bear any meaning relation with their corresponding sentence. 200 probes were created and distributed in the following manner: 25 probes for each condition, 50 filler word probes, and 100 filler nonword probes.

Each sentence had two probes assigned to it. One half of the subjects was presented with one probe-sentence pairing, and the other half was presented with the remaining probe-sentence pairing. This was done to control for any unforeseen meaning connection between probe and sentence. If such a relation were to exist it would be detected as a significant interaction (probe type x condition) in the statistical analysis.

Our "theoretical window" for measuring the processing cost of interpretation is the moment right after the pronominal has been heard since it is at this moment that other effects of pronominal interpretation such as priming have been reported (e.g. (Love *et al.*, 1998)). Since interpretation has to occur through the formation of a dependency we then infer that it is at this moment that the dependency at the level required for the interpretation of the pronominal is being formed. Nevertheless, since the sentences being contrasted differ also along syntactic lines, we had to be sure that if a difference were to be obtained it would not be due to this syntactic distinction. Consequently, a second position, the "control" position, was also probed at 400 msecs before the pronominal. In this manner, any unforeseen difference in resource demands between conditions, that would have accrued due to syntactic differences alone, would be detected at this early stage. And, by the same token, if a difference were to be obtained between sentences only right after the pronominal, we would be able to attribute it with confidence to the factor being hypothesized.

For each of the conditions, two positions are examined: "#" signals the control position and "^" signals the experimental position. Examples (2.a) and (2.b) below illustrate the conditions (Coargument reflexive and Pronoun conditions respectively), the probe positions and the corresponding probe:

(2) a. The driver$_i$ who caused a crash $^{\#}$blamed himself$_{i/*k}$ $^{\wedge}$…
b. The therapist$_i$ rolled a ball $^{\#}$around him$_{i/*k}$ $^{\wedge}$… Probe: GINGER

As mentioned, the control position is probed to ensure that up to the point of the pronominal both sentences are equally costly.[7] It is at the experimental position, right after the presentation of the pronominal, that the interpretive effects described above are expected to take place. Consequently, it is predicted that RT's will be significantly higher for the pronoun over the reflexive only at

[7] A methodological point must be made here. Given that both logophor and pronoun conditions are identical up to the control position, this position was tested at the beginning of the experimental series and only once. They were only tested with the logophors sentences because chronologically, the logophor condition was tested before the pronoun condition. The results are nevertheless assumed to be valid for both the logophor and the pronoun conditions, since as we mentioned up to this point in the sentences subjects could not predict whether they would hear a pronoun or a logophor.

the experimental position, as the pronoun, lacking a coargumenthood relation with its antecedent, must access discourse representation in order to find its antecedent. And establishing a dependency at the level of discourse is costlier to the system than establishing it at the level of syntax.

In this experiment, a total of 18 subjects were recruited to be tested for the control position and a total of 40 subjects were recruited for the experimental position. All subjects were native speakers of English and were recruited out of the Yale undergraduate student body. Ages ranged from 17 to 24. All subjects had normal or corrected-to-normal vision and hearing, and were neurologically intact by their own report.

All subjects performed at ceiling levels in the primary task as measured by the accuracy level in the comprehension task, thus indicating that subjects correctly understood the sentences presented in the script. There was no main effect of probe-sentence pairings – nor was there any interaction between Group and Condition, Control position: F(1,32)=0.37, p=.54., Experimental position (pronoun): F(1,80)=0.10, p=.74. This shows that there were no inexplicable sentence-probe effects, meaning, in turn, that any effect found could be safely attributed to the occurrence of the discourse operation under investigation.

There were two criteria for data inclusion: (1) The lexical decision had to be "yes" as only real words were used as probes for the experimental sentences; (2) The reaction time for any one trial had to fall within 3 standard deviations from the mean (this resulted in the exclusion of only 11% of the total data set from statistical analysis).

Regarding the control position, results reveal no interference effect, signalling that up to this point both conditions exacted the same computational cost onto the system (see Table 1 below).

	Control Position
Reflexive	729.04 (100.98)
Logophor	722.72 (98.02)
	t(17)=-0.46, p=.32

Table 1. Reflexives vs. Pronouns/Logophors. Mean (Standard Dev) and Significance Value

Regarding the experimental position, Table 2 below shows that there was no significant difference between the two conditions – reflexive and pronoun.

	Experimental Position
Reflexive	788.26 (188.78)
Pronoun	782.79 (193.08)
	t(39)=-0.68, p=.24

Table 2. Reflexives vs. Pronouns. Means (Standard Dev) and Significance Value

Two reasons could explain the absence of interference observed: 1) pronouns allow an extra-sentential antecedent, and this is the first option the system takes. As a result, no difference in interpretation can be observed right after the pronoun. The reason for this would be that in the absence of a possible non-sentential antecedent, the system is forced to wait. This period of "non action" is observed as lack of interference; 2) even though pronouns differ lexically from reflexives, morphologically they are very close (e.g. the morphology of *her* is found within that of *herself*). In the reflexive condition, this coincidence may cause activation of the pronoun to occur alongside that of the reflexive thus eliminating any potential difference that may accrue immediately afterwards. Contrast 2 below first reported in Piñango *et al.* (2001) controls for these two potential confounds by investigating the interpretation of another pronominal which, like pronouns, requires a discourse–based dependency for interpretation; but which like reflexives require an intra-sentential antecedent. That is the case of logophors.

2.2 Contrast 2: Coargument reflexive vs. logophoric reflexive

Piñango *et al.* (2001) investigate this contrast based on a script similar to that for Contrast 1 containing fifty experimental sentences and one-hundred and fifty filler sentences. The logophor condition was created by only substituting the pronoun in Experiment 1 for the corresponding logophor.

Examples (3a) coargument reflexive and (3b) logophor below illustrate the two conditions and the probes used:

(3) a. The driver$_i$ who caused a crash blamed $^{\#}$ himself$_{i/*k}$ $^{\wedge}$…
b. The therapist$_i$ rolled a ball around # himself$_{i/*k}$ $^{\wedge}$…
Probe: GINGER

Based on our model of pronominal interpretation, it was hypothesised that logophor interpretation would be costlier than coargument reflexive interpretation. Coargument reflexives find their antecedent through syntactic means only, whereas logophors additionally require the formation of a discourse-based dependency – as they are not in a coargument relation with their antecedent. Consequently, this predicts that RT for logophors will be significantly greater than RT for coargument reflexives.[8]

[8] A reviewer suggests that the difference in reaction time could be attributed to locality: Reflexives are more "local" than logophors/pronouns. But this distinction does not apply in the cases examined here since no long-distance reflexives are examined. As can be seen in the sentences in (3), except for coargumentbood, locality as defined in traditional binding theory (e.g. (Chomsky, 1981)) is the same in both instances (i.e. both pronominals are part of the same clausal unit, with the same accessible SUBJECT). And as (3) also shows, any potential difference to be found cannot be attributed to "raw" distance either (i.e. number of words between the pronominal and the antecedent). By contrast, another reviewer observes that non predicate-centered views of binding (e.g. (Chomsky, 1986)) also capitalize on

Similar to the previous contrast, for this contrast, subjects performed flawlessly in the primary task, comprehension of the sentences as measured by the paraphrase task. Following the criteria for data exclusion in Experiment 1, only 7.85% of all data points were excluded from statistical analysis. Moreover, there was no main effect of probe-sentence pairings – nor was there any interaction between Group and Condition $F(1,52)=0.21$, $p=.64$.

As reported for contrast 1, the control position revealed no significant difference between the two conditions. In contrast to Experiment 1, in this experiment, our prediction for the experimental position is borne out by the results: the RT for the logophoric reflexive *is* statistically significantly higher than that for the coargument reflexive (see Table 3 below). This suggests, on the one hand, that the arguments presented to explain the lack of interference in the contrast reflexive vs. pronoun may be on the right track. That is so because in this contrast the morphological differences between the pronominals have been controlled. On the other hand, the results suggest that the formation of a discourse dependency results in increased computation over the formation of a syntactic dependency.

	Experimental Position
Reflexive	675.84 (70.75)
Logophor	701.52 (94.76)
	$t(26)=2.71$, $p=.005$

Table 3. Reflexives vs. Logophors. Means (Standard Dev.) and Significance Value

2.3 Preliminary conclusions

Results from this second contrast serve to support the model of pronominal interpretation developed here where coargumenthood between the pronominal and its antecedent functions as a determining factor in the way (and level of representation) the necessary interpretive dependency between these two elements is ultimately formed. As this contrast shows, this dependency appears to demand additional computational resources, as measured by increased reaction times for the secondary task, whenever establishing such a dependency must be guided by discourse representation in addition to syntax.

We have also presented another piece of evidence – the contrast between interpretation of reflexives vs. interpretation of pronouns – which did not exhibit the expected effect. Nevertheless, when put in the context of the findings for logophors, we suggest this is a combination of a) morphosyntactic

locality, and that from this perspective the distribution of logophors examined is equally local as that for reflexives. Indeed, contrasting these two kinds of constructions reveals that unless one appeals to coargumenthood as the determining factor for pronominal interpretation, the results discussed here cannot be explained.

differences, and b) the potential availability of an extra-sentential antecedent for the pronoun that allows the system to delay the computation of the necessary coreferential dependency. This hypothesis clearly impels further experimentation.[9]

One complementary way in which the evidence presented here can be further supported is through the study of pronominal interpretation in systems that have impaired syntax. As we have been arguing so far, certain kinds of pronominals (e.g. logophors) require a discourse-based dependency, and rely only minimally on syntactic representation. This would predict that systems with impaired syntax should be able to establish the discourse dependency without a problem. Such a population is found in subjects diagnosed with Broca's aphasia; a syndrome caused by a lesion to the anterior left hemisphere which produces a language deficit characterized as syntactic in nature. The existence of this population thus allows us 1) to confirm the non-syntactic nature of logophor interpretation, and 2) to connect this processing deficit to cortical distribution distinctions necessary to build the bridge between mental processes and their neurological substrate. Directly below, we present experimental evidence illustrating this approach.

3 Pronominal Interpretation in Connection to the Brain

In order to examine the syntax-discourse correspondence within the context of brain-language relations, we look at the comprehension behaviour of patients with Broca's aphasia. As mentioned, Broca's aphasia is a language deficit that results from a focal lesion to the anterior left cortex. This syndrome is relevant to the study of pronominal interpretation because both in terms of production and of comprehension, the deficit observed has been characterized as syntactic in nature (e.g. (Friedmann & Grodzinsky, 1997; Piñango, 2000; Burkhardt & Piñango, 2002)). Consequently, the Broca's deficit constitutes an opportunity to observe the hypothesized dissociation in pronominal interpretation between syntactic and discourse representations. Given our focus of online comprehension, we adopt one generalization that seeks to describe the Broca's comprehension system: the slow syntax hypothesis (e.g. (Piñango, 2000, 2001; Burkhardt & Piñango, 2002)). It is stated as follows: Broca's patients are unable to carry out the construction (e.g. merge operation) of syntactic structure *on*

[9] We are currently exploring this possibility through a study investigating pronominal interpretation, this time using event-related response potentials (ERP). We hypothesize that, barring morphological differences, if the pronominal allows it, as in the case of pronouns, the system will opt for the least costly option, namely, to postpone the formation of the necessary discourse-based dependency. However, this would be done at the expense of partial interpretation, as the corresponding discourse unit remains necessarily partly unfilled. Preliminary results support this hypothesis (Burkhardt & Piñango, 2003).

time. However, there is a delayed point in time during comprehension at which syntactic structure is fully formed. At this point, Broca's and normal syntactic representations are indistinguishable from each other.

Thus stated, this characterization allows us to predict that the Broca's system will show normal-like comprehension of pronominals that do not require syntax for interpretation (e.g. logophors), and perform deficiently in those that do (e.g. reflexives). In what follows, we review the existing evidence on pronominal interpretation by Broca's patients within the context of the slow syntax hypothesis, and present an experiment that evaluates our prediction.

3.1 Pronominals in Broca's comprehension

Previous research shows a distinction in the behaviour of reflexives and pronominals in Broca's comprehension: In tasks that measure overall sentence comprehension (offline tasks), Broca's patients exhibit good performance in their interpretation of reflexives but perform no different from *chance* in their interpretation of pronouns (Grodzinsky *et al.*, 1993; Piñango, 2000). In tasks that tap the comprehension system as it unfolds (online tasks), which measure facilitation of an antecedent in the presence of a pronoun (e.g. cross-modal lexical priming), Broca's patients show no priming for reflexives, and only aberrant priming for pronouns; i.e. pronouns prime, but for the wrong antecedent (Love *et al.,* 1998).

This evidence suggests to us that the Broca's deficit, which we describe here as characterizable in terms of slow syntax, interacts with the comprehension of pronominals. Moreover, it does so only to the extent that syntactic representation is necessary for the determination of coargumenthood. This allows us to explain the seemingly contradictory performance by Broca's patients in the comprehension of reflexives as an artefact of the task. In the offline task, the comprehension system is able to wait until syntactic structure is fully formed to guide the identification of the antecedent. This is so because in this task, subject response is elicited *after* the sentence has been presented. This allows the system to complete the formation of the syntactic tree and then properly identify the antecedent through coargumenthood. In the online task, by contrast, the system is tapped *as comprehension unfolds*. Antecedent reactivation (in the form of priming) is elicited right after the reflexive, but syntactic structure, being slow to form, cannot enable the dependency to be established at the time the system is tapped. No priming effect is observed right after the reflexive because no dependency has formed that can license the reactivation of the antecedent. Yet, as the slow syntax hypothesis states, the priming effect is expected to fully emerge some time after the pronominal, once the syntactic tree is allowed to form.[10]

A similar situation presents itself with logophor interpretation which is observed here in the form of interference. That is, given our model of pronominal interpretation together with the description of the deficit, we have the following prediction: If (a) processing of logophors due to higher interpretive demands exerts a cost onto the system, and (b) Broca's patients are unable to construct syntactic structure on time, a contrast should obtain such that Broca's patients will exhibit the "normal" pattern (i.e. RT-logophors > RT-reflexives), but *only at a delayed point in the time course*.

3.2 Contrast 3: Reflexives vs. Logophors in Broca's comprehension

For this study, two Broca's patients JB and RD, clinically diagnosed with Broca's aphasia, with lesions consistent with the diagnosis (i.e. lesions involving the anterior frontal region, both cortex and underlying white matter, while sparing the anterior superior temporal cortex), and 13 control subjects matching the Broca's patients in age and level of education. The procedure was the same as in the previous two experiments discussed.

The materials used were the same as in the previous contrast. In this case, however, *three* positions are tested: 1) 400 msecs before the reflexive (control position), 2) 100ms after reflexive (experimental position), and 3) 600ms after reflexive (experimental position).[11] Contrast (4) below illustrates the conditions (reflexive vs. logophor respectively) and the positions tested:

(4) a. The *driver*$_i$ who caused a crash blamed $^{(1)}$ *himself*$_i$ $^{(2)}$ when the man $^{(3)}$ with…
b. The *therapist*$_i$ rolled a ball around $^{(1)}$ *himself*$_i$ $^{(2)}$ when the man $^{(3)}$ with . …

To recapitulate, given our descriptive generalization of a syntactic delay, we hypothesize that in contrast to unimpaired controls, Broca's aphasic patients will not *show an interference effect 100msecs after the reflexive*, since the interpretation of logophors and non-logophors requires a fully formed syntactic representation, which would not be formed at the normal early stage. The interference effect should instead emerge at position (3) a point within the time window where we hypothesize the syntactic representation supporting the pronominal and the antecedent would have been formed. So, whereas in the normal system the formation of syntactic structure can be said to overlap with the process of lexical retrieval (i.e. as soon as a morphological unit is detected all pertinent syntactic information is retrieved and integrated into the sentential

[10] To our knowledge, this hypothesis is yet to be tested.

[11] The calculation of this temporal point is based on independent on-line measures regarding (a) lexical retrieval (Prather *et al.*, 1997) and (b) implementation of long-distance dependencies (e.g. (Burkhardt, Piñango & Wong, 2003)).

structure), in the Broca's system syntactic information begins to be implemented later. Moreover, given our model of logophor interpretation, we hypothesize that Broca's patients will show the normal interference effect even at this later point, as this interpretation depends on a discourse-level dependency; a type of dependency that is predicted to be intact in the Broca's system.

Data were selected using constraints similar to those described for the previous two experiments. As a result, the percentage of data points excluded was 22%. Regarding the control position, no significant difference was found for both patients and normal controls: elderly control subjects: $t(47)=.40$, $p=.34$. Similarly, our two Broca's patients showed no significant effect at this position: JB: $t(16)= -.14$, $p=.44$; RD: $t(13)=-.20$, $p=.41$.

Regarding position (2), 100 msecs after the reflexive, elderly control subjects show an effect of logophoricity at 100msecs after the reflexive $t(24)=2.55$, $p=.008$. The Broca's patients, on the other hand, show no such interference: JB: $t(19)=1.04$, $p=.15$; RD: $t(23)=1.24$, $p=.11$.

Regarding position (3), 600 msecs after the reflexive, our prediction is confirmed by the results. A significant interference effect was observed for both Broca's patients [JB: $t(18)=1.97$, $p=.03$; RD: $t(22)=1.97$, $p=.03$]. By contrast, and as predicted control subjects showed no such effect [$t(24)=-.37$, $p=.35$].

The pattern of performance in the control and experimental positions reveal that older subjects pattern after young controls in logophor interpretation; Broca's patients, however, do not. Crucially though they fail to show interference in a predictable manner. The pattern these patients exhibit suggests that the interpretation of logophors is, as our model states, only minimally supported by syntactic representation. For these pronominals, the bulk of the interpretive processes takes place at the level of discourse.

4 General Conclusions

In this paper, we have investigated a model of pronominal interpretation that takes coargumenthood to be a determining factor in the manner in which syntax and discourse carry out 'labor distribution'. Presence of coargumenthood leads to a bound-variable interpretation (a syntactic process). Absence of-coargumenthood forces the system to achieve interpretation through means other than syntactic. Specifically, it forces the creation of a dependency at the level of discourse where referentiality is encoded. We have shown that the formation of this kind of dependency is observable in the form of computational cost.

Overall, the results presented here support a view of pronominal interpretation where interpretive mechanisms are the result of the interaction of levels of representation rather than the product of one level of representation. In particular, they show that, contrary to what would be predicted by non-predicate centred approaches to binding, coargumenthood forces the system to resort to discourse-based dependencies to achieve interpretation. By showing that these mechanisms are observable during the course of comprehension the results further allow a ready connection between abstract representation and processing which did not exist before (but one which is intended in, say, Reuland (2001), and Avrutin, (2002)).

From a neurological perspective, the localization value of Broca's aphasia together with the evidence from the online comprehension of the Broca's system support the hypothesis that discourse-based dependencies, and by extension, discourse representation is not supported by Broca's area (anterior left hemisphere), and by the same token, that only syntactic representation depends on the integrity of that region. Even though, the syntax-Broca's area connection had so far been circumscribed to the formation of long-distance dependencies, we show here that such a restriction is not necessary. Instead, our results suggest that the left anterior cortex is necessary for the implementation of even more fundamental syntactic mechanisms such as timely tree structure formation, without which interpretive relations that depend on coargument relations could not be formed.

Finally, the model proposed here supports a view of the language system where sentence comprehension (observed in this case through pronominal interpretation) is the end result of mutually constraining levels of representation, syntax and discourse, which are as we have shown, potentially observable at all levels of implementation: representational, dynamic and neurological.

References

Avrutin, S. 1999. *Development of the Syntax-Discourse Interface*. Boston: Kluwer.

---------- 2002. *Beyond narrow syntax.* Utrecht University Manuscript.

Burkhardt, P. and M. Piñango 2003. "Implications of Antecedent Availability During Pronominal Interpretation: An ERP Study". *The 10th Cognitive Neuroscience Annual Meeting*; New York City, NY.

----------, M. Piñango and K. Wong 2003. "The Role of the Left Anterior Hemisphere in Real-time Sentence Comprehension: Evidence from Split Intransitivity". *Brain and Language* 86.9-22.

Chomsky, N. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

De Vincenzi, M. 1996. "Syntactic Analysis in Sentence Comprehension: Effects of Dependency Types and Grammatical Constraints". *Journal of Psycholinguistic Research* 25.117-133.

Farmer, A. and M. Harnish 1987. "Communicative Reference with Pronouns". *The Pragmatic Perspective*. M. Papi and J. Verschueren (eds.). Amsterdam: John Benjamins.

Francis, W. and H. Kucera 1982. *Frequency Analysis of English usage: Lexicon and Grammar*. Boston: Houghton Mifflin.

Grodzinsky, Y., K. Wexler, Y. Chien and S. Marakovitz 1993. "The Breakdown of Binding Relations". *Brain and Language* 45.396-422.

---------- and T. Reinhart 1993. "The Innateness of Binding and Coreference". *Linguistic Inquiry* 24.69-101.

Harris, T., K. Wexler and P. Holcomb 2000. "An Electrophysiological Investigation of Binding and Coreference". *Brain and Language* 75.313-346.

Heim, I. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. PhD Dissertation, Univ. of Massachusetts, Amherst.

Huang, Y. 2000. "Anaphora: A Cross-linguistic Study." *Oxford Studies in Typology and Linguistic Theory*. Oxford: Oxford University Press.

Kamp, H. and U. Reyle 1993. *From Discourse to Logic*. Dordrecht:Kluwer.

Love, T., J. Nicol, D. Swinney, G. Hickok and E. Zurif 1998. "The Nature of Aberrant Understanding and Processing of pro-forms by Brain-damaged Populations." *Brain and Language* 65(1).59-62.

Motta, G., F. Rizzo, D. Swinney and M. M. Piñango 2000. *TEMPO: A Software for Real-time Comprehension and Neuroimaging Testing*. Brandeis University, UC San Diego & Yale University.

Piñango, M. 2000. "Canonicity in Broca's Sentence Comprehension: The Case of Psychological Verbs". Y. Grodzinsky, L. Shapiro and D. Swinney (eds.). *Language and the Brain*. San Diego: Academic Press.

--------- 2001. "Cortical Reflections of Two Pronominal Relations." H. Simon and H. Wiese (eds.). *Pronouns: Representation and Processing*. Amesterdam: John Benjamins.

---------, E. Zurif and R. Jackendoff 1999. "Real-time Processing Implications of Enriched Composition at the Syntax-semantics Interface." *Journal of Psycholinguistic Research* 28(4).395-414.

----------, P. Burkhardt, D. Brun and S. Avrutin 2001. "The Psychological Reality of the Syntax-Discourse Interface: The Case of Pronominals". *Workshop from Sentence Processing to Discourse Interpretation: Crossing the Borders*. UIL-OTS, Utrecht, June, 15-16.

Prather, P., E. Zurif, T. Love and H. Brownell 1997. "Speed of Lexical Activation in Nonfluent Broca's Aphasia and Fluent Wernicke's Aphasia". *Brain and Language* 59.391-411.

Top-down and Bottom-up Effects on the Interpretation of Weak Object Pronouns in Greek

Stavroula-Thaleia Kousta
University of Cambridge

A fragment-continuation experiment investigated the interaction of subjecthood, implicit causality, and syntactic parallelism for the interpretation of weak object pronouns (clitics) in Greek. Participants wrote continuations for fragments consisting of a subject NP followed by an accusative or genitive clitic. The clitic could be interpreted as co-referring with one of two NPs used in the previous sentence as arguments of implicit causality verbs, such as *frighten* (subject-biasing verb) and *hate* (object-biasing verb). While there was no subject antecedent preference for clitics, syntactic parallelism and implicit causality jointly determined preferred assignments. The results obtained challenge the underlying assumptions of computational models of focus, such as Centering Theory and Local Focusing, as well as recent claims about the role of semantic factors in anaphora resolution.

1 Introduction

Pronominals, due to their reduced semantic content, are especially dependent on aspects of the preceding discourse representation for their interpretation. The preceding discourse exerts a top-down influence on resolution, in the sense that this influence is independent of the presence of a pronoun that needs to be resolved. Nevertheless, when pronominals are not null, they do contain semantic/syntactic information (such as gender, number, case) that constrains their interpretation. This information exerts a bottom-up influence on the processing of pronominals, in that it is driven by the input (the pronoun) that needs to be integrated in the developing discourse representation.

There is a substantial body of research devoted to the identification of individual top-down/bottom-up factors influencing the interpretation of pronouns – even though the exact nature of most of these factors is in fact controversial. On the other hand, there is comparatively little work addressing the manner in which these factors interact in pronoun resolution. The experiment reported here investigated the nature of three factors proposed to influence pronoun resolution, parallelism, subjecthood and implicit causality, as well as the relative strength of each factor when they are in competition. It was conducted in Greek because, due to the particular facts about the linear position of weak object pronouns (unlike English, where object pronouns canonically follow the verb, in Greek weak object pronouns precede it), it was possible to investigate the effects of the three factors independently from other confounding factors, such as verb-semantics in the clause/sentence containing the pronoun.

2 Background

Bottom-up effects of interest in this paper are any effects that would not be registered unless there was a pronoun that needed to be resolved, as well as effects due to the morphosyntactic features of the pronoun and whether they match with those of a potential antecedent. Syntactic parallelism biases in pronoun resolution are bottom-up effects, since they depend on whether a pronoun shares the same grammatical role with a potential antecedent. For instance, the pronoun '*him*' below is biased to co-refer with 'George' rather than 'John':

(1) John kicked George on the leg and then Mary slapped *him* in the face.

With regards to top-down effects on anaphora resolution, the degree of salience/accessibility of an entity in the discourse representation has been assumed to constitute one of the main influences on pronoun interpretation. This notion of salience/accessibility is encapsulated in the concept of 'focusing': certain elements in the discourse representation of a text are more prominent than others and are hence more likely candidate referents for an upcoming pronoun. There are various syntactic, semantic, and pragmatic factors affecting the degree of prominence of an entity in the discourse representation. In the experiment reported below we investigated the role of two of these factors, subjecthood and implicit causality.

In the rest of this section, we will first provide some descriptive facts about word order and the pronominal paradigms in Greek, and then outline past research on the effects of parallelism, subjecthood, and implicit verb causality, relating it to the objectives of the experiment reported below.

2.1 Greek

Modern Greek has two paradigms of personal pronouns (Holton *et al.*, 1997), strong (emphatic) and weak (clitic), marked for case, number, and gender (masculine, feminine, neuter). Strong pronouns are stressed, are normally used for emphasis/contrast, and can function as the subject or the direct/indirect object of a verb or as the object of a preposition. Weak pronouns (clitics) are normally unstressed and cliticise to the left of the verb, thus preceding it. Clitics can function as the direct (case-marked accusative) or indirect (case-marked genitive) object of verbs. The accusative (masc: 'ton', fem: 'ti(n)', neuter: 'to') and genitive forms (masc: 'tu', fem: 'tis', neuter: 'tu') of the clitics are phonologically identical with the accusative and genitive forms of the definite article.[1] Greek also permits clitic doubling (example (2) below) and clitic left dislocation (example (3) below), where both the full noun phrase object and its corresponding clitic appear in the verb phrase. In cases of clitic doubling, the clitic appears before the verb and the full noun phrase object follows the verb, while in instances of left dislocation both the clitic and the noun phrase object precede the verb, the noun phrase object appearing before the clitic.

(2)	Tin	teliosa	tin	ergasia	htes.
	clitic-fem-acc	finished-I	the-fem-acc	essay-fem-acc	yesterday.
	"I finished the essay yesterday."				
(3)	Tin	ergasia	tin	teliosa	htes.
	the-fem-acc	essay-fem-acc	clitic-fem-acc	finished-I	yesterday.
	"I finished the essay yesterday."				

2.2 Parallelism

Early recognition of parallelism effects was made with regards to the resolution of subject pronouns (Garvey *et al.*, 1976; Grober *et al.*, 1978). However, it is not clear whether subject pronouns are resolved to a preceding subject because of syntactic parallelism or because entities realised in subject position are inherently more salient (see Section 2.3 below). More recently, parallelism effects have been demonstrated with non-subject pronouns (Chambers & Smyth, 1998; Smyth, 1992, 1994; Stevenson *et al.*, 1995; Stevenson & Urbanowicz in Stevenson, 1996; Pearson *et al.*, 2000), for which the two hypotheses yield distinct predictions: the subjecthood-salience hypothesis predicts an object pronoun will preferentially be resolved to the grammatical subject, whereas the parallelism hypothesis predicts that the pronoun will be co-referential with an antecedent having the same grammatical role.

It is important to note that strong preferences for an antecedent having the same grammatical role have been obtained only when there is both semantic and syntactic parallelism between the clause/sentence containing the pronoun and the preceding clause/sentence (Smyth, 1992; Chambers & Smyth, 1998) and when the two clauses/sentences share the same overall constituent structure (Smyth, 1994; Stevenson *et al.*, 1995). For example, in sentence (1) above, repeated below as (4) for convenience, both verbs are semantically related and assign the same θ-roles to their arguments, and the two clauses have the same constituent structure (NP V NP PP):

(4) John *kicked* George on the leg and then Mary *slapped* him in the face.

[1] The only difference between clitics and the definite article is that for the masculine accusative form of the clitic, 'ton', the final 'n' is obligatory, while for the masculine accusative form of the definite article, 'to(n)', the final 'n' is obligatory only in certain phonological environments.

When there is no semantic parallelism and the two clauses/sentences do not have the same global constituent structure, parallel assignments for non-subject pronouns are reduced and a preference for the grammatical subject is attested instead.

As was mentioned above, in Greek clitics precede the verb. In our experiment, participants were asked to complete fragments which ended with a clitic and from which the verb was missing. This paradigm makes it possible to tease apart syntactic parallelism, on the one hand, and semantic parallelism/common constituent structure, on the other. Whether the sentences would eventually be both semantically and syntactically parallel and share the same global constituent structure depended on the continuations participants produced.

With semantic parallelism and congruence of constituent structure not playing a role in the experiment, it was also possible to assess the proposal that parallelism biases operate in such general terms as: subject pronouns are resolved to subjects, while non-subject pronouns are resolved to non-subjects. This proposal has been made within the framework of Local Focusing[2] (Suri & McCoy, 1993), a computational model that attempts to account for parallelism effects in the interpretation of anaphors. The model assumes the existence of two foci per utterance, dubbed Subject Focus and Current (or Local) Focus. Candidates for the Subject Focus are primarily entities realised in subject position, and subject pronouns are preferentially resolved to that focus; candidates for the Current Focus are primarily entities realised in non-subject position and non-subject pronouns are preferentially resolved to that focus. Local Focusing assumes a very loose notion of parallelism, which seems not to be supported by experimental evidence. Smyth (1992, 1994) and Stevenson *et al.* (1995) have shown that the degree of semantic/syntactic parallelism and common global constituent structure between utterances is crucial in determining preferred interpretations. We were interested in seeing whether the same pattern of results would obtain when semantic parallelism and equivalence of global constituent structure do not apply.

2.3 Subjecthood

There is a large body of research demonstrating that pronouns preferentially select antecedents realised in subject position in the preceding clause/sentence, on the assumption that subjects are more salient than entities realised in other grammatical positions (Brennan *et al.*, 1987) and that pronouns select salient

---

[2] The framework proposed by Suri and her collaborators is in fact a modification of an earlier model by Sidner (1979). It was later renamed as Revised Algorithms for Focus Tracking and Revised Algorithms for Pronoun Resolution (RAFT/RAPR; Suri & McCoy, 1994; Suri *et al.*, 1999).

antecedents (Gundel *et al.*, 1993). This subject-preference has been demonstrated in a variety of languages (see (Arnold, 1998) for English, Spanish, and Mapudungun; (Miltsakaki, 2001) for Greek). The primacy of subjecthood has been most forcefully argued in work based on Centering Theory (Grosz *et al.*, 1995), a computational model that attempts to relate focus of attention, the form of anaphoric expressions and inferential complexity in a psychologically motivated account of local coherence. Although various proposals have been made regarding the determinants of salience, most experimental research on Centering has adopted the view that salience ordering is solely determined by surface structural considerations and serial linear position (Gordon *et al.*, 1993; Gordon & Scearce, 1995; Hudson-D'Zmura & Tanenhaus, 1998; Gordon & Chan, 1995; Kennison & Gordon, 1997). Some of this research has been used to support the view that the grammatical \hierarchy as proposed by Brennan *et al.* (subject > object > object2 > other subcategorised functions > adjuncts) is such a strong determinant of salience that it overrides any influences by semantics (Gordon & Scearce, 1995) or the thematic structure of utterances (Hudson-D'Zmura & Tanenhaus, 1998).

Nevertheless, studies on parallelism effects in comprehension cited earlier have demonstrated that the primacy of subjecthood is subject to modification by bottom-up influences due to the grammatical role of the pronoun. Under conditions of 'strict' syntactic parallelism, non-subject pronouns are preferentially resolved to a structurally parallel antecedent rather than the subject, as Centering would claim. Here we were interested in determining whether the grammatical role of the clitic alone would be enough to override any subjecthood effects.

2.4 Implicit Verb Causality

With certain interpersonal verbs, the cause of the state/event denoted by the verb is attributed to one of its arguments (Brown & Fish, 1983; Au, 1986; Corrigan, 1988; Fiedler & Semin, 1988; *inter alia*). Verbs exhibiting such causal structure can be broadly divided in two classes, on the basis of the surface grammatical realisation of the argument that is perceived as the causal initiator in each class. For instance, the verbs *irritate* and *criticise* both subcategorise for two arguments, a subject and a direct object, but with verbs of the *irritate* type (subject-biasing (SB) verbs), it is the entity realised in subject position that is perceived as the cause of the event/state, while with verbs of the *criticise* type (object-biasing (OB) verbs), it is the grammatical object that is assumed to be the cause of the state/event denoted by the verb. Implicit verb causality has been suggested to affect preferred assignments for pronouns. For instance, in (5) below, the ambiguous pronoun is normally interpreted as co-referring with ‘Jim’, although a situation can be conceived where the alternative assignment is more plausible.

(5) John criticised Jim because he failed to meet the deadline.

It is not entirely clear, however, if implicit causality exerts a top-down influence on the interpretation of pronouns. In the online processing literature, it is an open question whether implicit causality functions as a focusing mechanism (Greene & McKoon, 1995; McDonald & McWhinney, 1995; Koornneef *et al.*, 2002; Long & DeLey, 2000); or if, when utilised in comprehension, it is registered as a late effect occurring during the integration of propositions (Garnham *et al.*, 1996; Stewart *et al.*, 2000). Even within focusing accounts, it is not always clear whether the effect is due to the fact that implicit causality makes the argument perceived as the causal initiator more accessible/salient; whether it is due to the presence of a causal connective such as ‘because’, which directs attention to the causes of the event/state denoted by the verb; or whether it is due to the presence of an anaphor that needs to be resolved (in which case implicit causality effects would be more appropriately viewed as the outcome of bottom-up processing). When implicit causality is examined independently of connectives/pronoun resolution, the evidence from different methodologies is mixed. Using a cross-modal probe recognition paradigm, McDonald and MacWhinney (1995) failed to find any differential activation of either of the verb’s arguments as a function of implicit causality and attributed implicit causality effects to resolution processes. Holtgraves and Raymond (1995), on the other hand, demonstrated with a series of cued recall experiments that there is significantly better recall for the causal argument of implicit causality verbs (at least in active sentences).

More recently, Stevenson *et al.* (2000) have claimed that interpersonal state verbs, which form a large sub-class of implicit causality verbs, lack an independent focusing structure. Any preferences for either of the arguments of these verbs is determined by the semantics of the connective linking the clause/sentence containing the interpersonal verb and the clause/sentence containing the pronoun, the full stop being assumed to function as an implicit causal connective. Modifying Stevenson *et al.*’s proposal, Miltsakaki (2002, 2003) suggested that verb biases influence antecedent assignments only within sentences, while anaphora resolution across sentence boundaries is mainly determined by structural considerations.

In the experiment reported here, we were interested in clarifying the nature of verb biases in pronoun resolution. Do verb biases determine co-reference patterns for spontaneously produced and already present pronouns across sentence boundaries and in the absence of an explicit/implicit causal link between sentences? If in the absence of an anaphor and a causal link between the sentence containing the biasing verb and the fragment in our experiment, spontaneously produced pronominals co-refer with the argument perceived as the causal initiator, that would be taken as evidence for a top-down (focusing) implicit causality effect. If, on the other hand, verb biases are evident only when a pronoun that needs to be resolved is present, then implicit causality effects can be attributed to resolution processes only.

3 Experiment

A fragment-completion task was employed in order to investigate the interaction of subjecthood, implicit causality and the syntactic role of the anaphor in the interpretation of weak object pronouns in Greek. Subjects wrote continuations for fragments consisting of a subject NP alone or a subject NP followed by a clitic. The clitic could be interpreted as co-referring with one of two NPs used in the previous sentence as arguments of implicit causality verbs and was case-marked either accusative or genitive. A potential antecedent having the same syntactic role was available for accusative but not for genitive clitics. In short, the texts used had the following structure (only the English translation is given here; a detailed description of the materials used and some examples can be found in Section 3.2 under *Materials and Design*):

> [neutral context sentence]
> John-NOM *irritates/hates* Jim-ACC.
> Mary-NOM *clitic-ACC… / clitic-GEN… / no clitic*

Anaphora resolution was chosen to be studied in *inter*- rather than *intra*-sentential contexts for two reasons. First, it seemed to be the case that a variety of coherence relations were possible if a sentence boundary intervened. Hence, any preferences obtained could not be attributed to the semantics of the connective used. Second, we could address the suggestion (Miltsakaki, 2002, 2003) that verb semantics has no influence on pronoun resolution inter-sententially.

Centering Theory would predict that both types of clitic with both types of verb would be resolved to the subject. Following Local Focusing claims, one would expect both types of clitic, regardless of case-marking and verb type, to be resolved to the grammatical object. More interactive approaches would expect individual effects of subjecthood, parallelism, and implicit causality to interact to give the preferred interpretation. If this is the case, further questions arise: what is the strength of each of these factors and in what manner do they interact? Two scenarios were *a priori* considered to be plausible, given the research outlined above and theoretical considerations:

A: Genitive clitics are resolved to the grammatical subject, but subjecthood effects are stronger for genitive clitics following subject-biasing than object-biasing verbs. Similarly, accusative clitics are resolved to the grammatical object, but preferences for the object with accusative clitics are stronger in the context of object-biasing than subject-biasing verbs. This pattern is suggested by earlier research on parallelism, but also allows a role for implicit causality, albeit weak.

B: All clitics are resolved to the grammatical subject apart from accusative clitics after sentences with object-biasing verbs. In this version, implicit causality and syntactic parallelism have the same 'weight', and only when these two factors point to the same antecedent do they override the stronger subjecthood effect. Subjecthood in the preceding sentence may turn out to be the strongest determinant of antecedent assignments in the Greek experiment because parallel interpretations are supported only by congruence between the syntactic role of the anaphor and that of a potential antecedent, rather than congruence of overall constituent structure and simultaneous semantic parallelism as was the case in all the English experiments that have registered a pronounced parallelism effect.

The conditions without clitics were included for two reasons. First, we were interested in seeing whether weak object pronouns would be produced spontaneously under circumstances where gender is not a disambiguating factor. Second, if implicit causality exerts a top-down influence, more spontaneously produced clitics should co-refer with the grammatical subject in the context of subject-biasing verbs than in the context of object-biasing verbs. That is, if an implicit causality effect is obtained, it is expected to enhance/reduce an overall subjecthood effect.

3.1 Free listing of causes task

Since implicit causality is one of the factors manipulated, it was essential to ensure that the verbs used in the materials exhibit implicit causality effects**,** and that the effects are in the predicted direction. Consequently, a set of forty verbs (20 subject-biasing and 20 object-biasing) was selected from English studies on implicit causality (Au, 1986; Greene & McKoon, 1995; Brown & Fish, 1983), and their Greek equivalents in meaning were used in a free listing of causes task. The majority of the verbs used were psychological predicates.[3] The task

[3] In Greek there are three classes of psychological predicates (Anagnostopoulou, 1996): experiencer (subject)-stimulus (object) verbs like 'thavmazo' (admire), with which the stimulus object has morphological accusative; stimulus (subject)-experiencer (object) verbs such as 'eknevrizo' (irritate), where the experiencer object is case-marked accusative; and stimulus (subject)-experiencer (object) verbs, such as 'areso' (like), with a dative experiencer (morphologically marked genitive or a prepositional

was modelled after a study by Brown and Fish (1983) and was conducted orally. Participants were told that they were going to hear short sentences involving interpersonal verbs such as ‘approach’, ‘A approached B’, followed by the question ‘Why?’. They were instructed to reply to the question with a short sentence beginning with ‘Because…’. They were also told that it was essential that their replies be about one of the two entities, and not about an event/state in which both entities were involved jointly.

A total of twenty-five subjects participated in this preliminary study, and the results appear in Table 1. The results for all verbs were in the predicted direction, significantly so for 34 out of 40 verbs. The items in boldface are the ones actually used in order to generate the materials for the experiment.

3.2 Method

Subjects. A total of 30 subjects (27 female, 3 male; mean age 26.1 ± 3.7 years) participated in the experiment as volunteers. Four subjects were replaced, either because they had become aware of the experimental manipulations (3) or because they consistently treated accusative clitics as definite articles (1).

Materials and design. The materials consisted of a context sentence in which 3 individuals, A, B, and C, were introduced by proper name. Two of the individuals, A and B, were of the same gender and always appeared in a co-ordinate NP, while the third entity, C, was always the first-mentioned entity in the context sentence. The second sentence used A and B from the context sentence with one of the 30 verbs from the preliminary study. Although Greek is a free word order language, SVO order was always used for this sentence. The final sentence introduced C in subject position (i.e., marked nominative), and was followed by a clitic marked accusative or genitive, or by no clitic at all.

An equal number of masculine and feminine clitics were used. The accusative clitic shared the same case-marking as the object of the previous sentence, while the genitive clitic did not. In short, the structure of the materials was as follows:

> C, A and B…/C…A and B
> or C, B and A…/C…B and A[4]
> A-nom verb B-acc.
> C-nom clitic-acc…/clitic-gen…/(no clitic)…

---

phrase). Only verbs from the first two classes were used in the free listing of causes task and the subsequent experiment.

[4] This extra manipulation of the order in which the two individuals were mentioned within the co-ordinate NP or the list was included in order to control for any ‘second order’ primacy effects; that is, effects due to the introduction of one of the individuals before the other, regardless of the fact that neither of them was the first-mentioned entity in the context sentence.

Subject-biasing Verbs	S	O	Object-biasing Verbs	S	O
eksorgizo (infuriate)	22***	3	**thavmazo** (admire)	3	22***
ediposiazo (impress)	21***	4	**katigoro** (accuse)	5	20**
sokaro (shock)	22***	3	**miso** (hate)	3	22***
eknevrizo (irritate)	22***	3	**zilevo** (be jealous of)	1	24***
tromazo (frighten)	18*	7	**fovame** (be afraid of)	3	22***
djaskedazo (amuse)	18*	7	**parigoro** (console)	0	25***
tromokrato (terrify)	20**	5	**ektimo** (esteem)	0	25***
eksapato (deceive)	21***	4	**ebistevome** (trust)	3	22***
goitevo (charm)	24***	1	**apehthanome** (detest)	4	21***
aidiazo (disgust)	24***	1	**antipatho** (dislike)	4	21***
enohlo (annoy)	20**	5	**tremo** (dread)	2	23***
anisiho (worry)	21***	4	**latrevo** (adore)	1	24***
stenohoro (distress)	21***	4	**sinhero** (congratulate)	0	25***
ekpliso (surprise)	21***	4	**epeno** (praise)	3	22***
anastatono (upset)	21***	4	**agapo** (love)	2	23***
disaresto (displease)	17	8	lipame (pity)	5	20**
apotho (repel)	15	10	sibatho (like)	4	21***
epireazo (influence)	15	10	efharisto (please/thank)	3	22***
ekseftelizo (humiliate)	13	12	epiplito (reprimand)	4	21***
apogoitevo (disappoint)	17	8	simvulevo (advise)	9	14

Table 1: Free listing of causes for subject-biasing and object-biasing verbs. S=Subject preference O=Object Preference. p≤.05 p≤.01** p≤.001*** (one-tailed test)*

Example stimuli for each verb type appear in Tables 2&3. Examples of continuations supplied by the participants exhibiting some of the attested coherence relations are also given in these Tables.

It was essential to ensure that, with both types of clitic, continuations in which the clitic co-specified with a subject or object antecedent were equally acceptable. Consequently, two continuations were generated for each clitic in each experimental text, one with a subject intended antecedent for the clitic and the other with an object intended antecedent. This yielded a total of four continuations for each text. The texts along with their continuations were given to four Greek native speakers who were asked to identify the antecedent of the clitic and judge the whole discourse for coherence. Each subject read a single continuation for each discourse and four presentation lists were constructed. If a discourse was not judged 100% coherent, subjects were asked to locate the source of the incoherence. Discourses that were not judged to be 100% coherent were modified and subjected to further pre-tests, until the 100% criterion was met.

Subject-Biasing Verb

[1] (context): "Kostas, Ioana, and Eftihia found each other by coincidence in the same song contest."

[2]	I	Eftihia	ediposiase	tin	Ioana.	
	The-fem-nom	Eftihia-nom	impressed	the-fem-acc	Ioana-acc.	
	"Eftihia impressed Ioana."					
[3]	O	Kostas	ti(n)…/	tis…/	…	
	The-masc-nom	Kostas-nom	cl-fem-acc…/	cl-fem-gen…/	…	
	"Kostas clitic-fem-acc…/clitic-fem-gen…/no clitic…"					

Example Continuations

O	Kostas	**tin**	enthusiase.[5]
The-masc-nom	Kostas	cl-fem-acc	delighted.

"Konstantinos delighted her." (resemblance: parallel)

O	Kostas	**ti**	zilepse	gia	ta	thetika	sxolia	pu	apespase.
The-masc-nom	Kostas	cl-fem-acc	was-jealous-of	for	the	positive	comments	that	received-she.

"Kostas was jealous of her for the positive comments she received." (cause-effect: result)

O	Kostas	**tis**	edose	ta	sigharitiria	tu.
The-masc-nom	Kostas	cl-fem-gen	gave	the	congratulations	his.

"Konstantinos congratulated her." (narration)

Object-Biasing Verb

[1] (context): "Unexpected circumstances made Anthi, Haris and Stelios share the same house."

[2]	O	Haris	apehthanete	to	Stelio.	
	The-masc-nom	Haris-nom	detests	the-masc-acc	Stelios-acc.	
	"Haris detests Stelios."					
[3]	I	Anthi	ton…/	tu…/	…	
	The-fem-nom	Anthi-nom	cl-masc-acc…/	cl-masc-gen…/	…	
	"Anthi clitic-masc-acc…/clitic-masc-gen…/no clitic…"					

Example continuations

I	Anthi	**ton**	sibathi.
The-fem-nom	Anthi	cl-masc-acc	likes.

"Anthi likes him." (resemblance: contrast)

I	Anthi	**tu**	ipe	na	min	to	dixni.
The-fem-nom	Anthi	cl-masc-gen	told	to	not	it	show.

"Anthi told him not to show it." (narration)

I	Anthi	**ton**	ixe	prokatavali	arnitika	apenadi	sto	Stelio.
The-fem-nom	Anthi	cl-masc-acc	had	prejudiced	negatively	towards	to-the	Stelios.

"Anthi had prejudiced him against Stelios." (cause-effect: explanation)

Tables 2&3: Example Stimuli and example continuations provided by participants, annotated for the type of coherence relation exhibited between the second and the third sentences.

[5] Coherence relations were identified following Kehler (2000).

Ten filler discourses were also constructed. These were similar in form to the experimental passages, but the text-final fragment always consisted of a proper name in subject position (i.e., marked nominative) alone. In total, participants completed twenty texts in which the fragment ended in a clitic and twenty texts in which the fragment consisted of a proper name alone.

Procedure. Participants were tested individually. Each was given a booklet containing the forty passages on separate pages. The fragments were followed by dots to indicate that a continuation was required. They were instructed to produce coherent continuations to complete the final sentence. Even though there were no time limits, subjects were advised not to spend too much time on finding a continuation, but to write what first came to mind.

After participants had provided continuations to all the passages, they were asked a number of questions to ensure that they had not become aware of the experimental manipulations, and hence, had not adopted special strategies in completing the texts. Finally, they were asked to go back to all texts that contained a clitic and underline the intended antecedent according to the continuation they had provided.

Results. In the two conditions without clitics, 22% of the observations contained a singular pronoun either immediately after the subject NP (17.3%) or further downstream (4.7%). Notable is the low occurrence (2.7%) of genitive clitics immediately after the subject NP, that is, in the configuration used in the conditions containing a genitive clitic in the experiment. The significance of this observation will be taken up again in the discussion of the results. Tables 4&5 below show for each verb type the number of observations (out of 300 possible contexts) that contained a singular accusative/genitive pronoun resolved either to he subject or the object of the preceding sentence.

	antecedent: **SUBJECT**			
	accusative		genitive	
	beginning	downstream	beginning	downstream
SB verb	3 (1%)	2 (0.6%)	1 (0.3%)	2 (0.6%)
OB verb	1 (0.3%)	1 (0.3%)	0	1 (0.3%)

	antecedent: **OBJECT**			
	accusative		genitive	
	beginning	downstream	beginning	downstream
SB verb	13 (4.3%)	2 (0.6%)	4 (1.3%)	1 (0.3%)
OB verb	27 (9%)	1 (0.3%)	3 (1%)	4 (1.3%)

Tables 4&5: Antecedent selection, case-marking and linear position of pronouns for each verb type in the two no-clitic conditions. (The first number in each cell represents the occurrences out of 300 possible contexts; the number in parentheses indicates percent occurrence.)

The data from these two conditions were analysed separately from the other four conditions because of the low number of observations. In order to perform statistical analyses, the following adjustment procedure was used: for each subject/item, the number of times a clitic with an object antecedent was used was divided by the total number of times a singular clitic was used. For all analyses reported below, subjects/items without any relevant observations (i.e. singular clitics) were ignored.

One-sample t-tests (assuming subjects (*t1*) and items (*t2*) as random variables) confirmed that, in the object-biasing verb condition, the proportion of spontaneously produced clitics with object instead of subject intended antecedents differed significantly from chance ($t1(18)=4.916$, $p<.001$; $t2(13)=13.985$, $p<.001$). For the subject-biasing verb condition, the effect was significant only in the subjects analysis ($t1(16)=2.256$, $p<.05$). Two further t-tests (paired-samples for the subjects analysis (*t1*) and independent-samples for the items analysis (*t2*)) revealed that there was a difference in the proportion of clitics with object antecedents between the two conditions, but the effect was reliable in the items analysis only ($t1(25)=1.439$, $p>.10$; $t2(11.177)=2.739$, $p<.05$).

Turning to the four conditions in which the fragment ended in a clitic, the mean proportions of *object* assignments for each condition appear in Figure 1. 0.5% of the observations were missing (participants failed to provide a continuation). A further 0.83% of the data was discarded (cases where the clitic had been construed as the definite article, including cases of clitic left dislocation). Instances of clitic doubling (0.67% of the data) were included in the analyses, since the clitic had been construed as such at the end of the fragment.

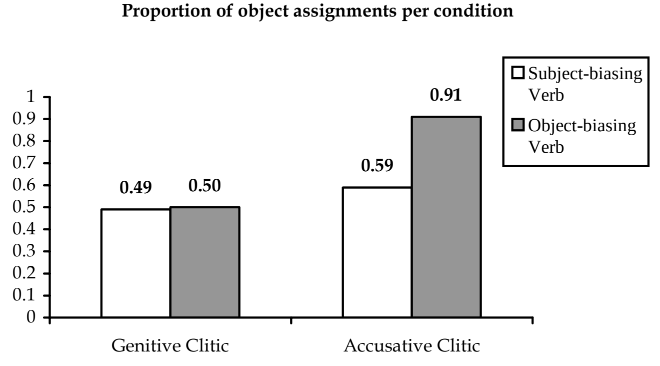


Figure 1: Mean proportions of object assignments in each of the four clitic conditions.

Two separate repeated measures ANOVAs were conducted on the proportions of object assignments, in order to test generality over subjects (*F1*) and items (*F2*). The analyses revealed a significant main effect of both variables (type of clitic: $F1(1,29)= 49.400$, $p<.001$; $F2(1,28)= 35.915$, $p<.001$; type of verb: $F1(1,29)=16.414$, $p<.001$; $F2(1,28)=6.329$, $p<.05$). These effects were qualified by a significant interaction ($F1(1,29)=17.181$, $p<.001$; $F2(1,28)=13.375$, $p<.001$).

Pairwise comparisons (employing Bonferroni's adjustment procedure to reduce the risk of Type 1 errors arising from making multiple comparisons) confirmed the obvious: there were significantly more object assignments in the condition in which an accusative clitic was used following a sentence with an object-biasing verb (OB-acc) than in all other conditions (OB-acc/OB-gen: $t1(29)=7.312$ $p<.001$; $t2(14)=6.528$ $p<.001$; OB-acc/SB-acc: $t1(29)=6.499$ $p<.001$; $t2(28)=4.943$ $p<.01$; OB-acc/SB-gen: $t1(29)=8.301$ $p<.001$; $t2(28)=6.473$ $p<.001$). None of the other comparisons yielded statistically significant results.

Finally, one sample t-tests were conducted on each condition separately, to test whether the proportion of object assignments differed significantly from chance. This set of tests yielded significant results for the OB-acc condition ($t1(29)=14.628$, $p<.001$; $t2(14)=8.078$, $p<.001$). There was also a trend for more object assignments for the SB-acc condition in the subject analysis ($t1(29)=1.697$, $p=.10$) only.

Overall, the significant two-way interaction in the ANOVAs, the significant difference between the OB-acc and all the other conditions in the pairwise comparisons, and the significant difference from chance assignments in the one-sample t-tests for the same condition indicates that only when both the bias of the verb and parallelism of syntactic role point to the same antecedent is there a significant preference for that antecedent. There was also a numerical advantage for object assignments in the SB-acc condition (where there was parallelism of syntactic role but the bias of the verb was incongruent with the parallel (object) antecedent), but statistically it was only a trend. For the conditions ending in a genitive clitic, antecedent choices were at chance and overall, there was no preference for subject antecedents in any of the conditions.

4 General Discussion

The pattern of results obtained in the study is not compatible with any of the original predictions. When fragments ended in clitics, the only circumstances under which a robust preference for a parallel antecedent was registered was when the syntactic role of the clitic and the causality bias of the verb in the previous utterance converged on a single antecedent (OB-acc condition). When the role of the clitic and the bias of the verb in the previous utterance pointed to different antecedents (SB-acc condition), there was only a tendency for more object assignments. There was no preference whatsoever for either antecedent when the clitic did not share the same syntactic role with either possible antecedent (SB-gen and OB-gen conditions), regardless of the bias of the verb in the previous utterance. Even though the small number of relevant observations in the no-clitic conditions does not allow us to draw any definitive conclusions, instead of a grammatical subject preference, as predicted, there was a grammatical object preference with spontaneously produced clitics. This effect was more reliable in the object-biasing verb condition. However, there seemed to be a difference in the proportion of clitics with object antecedents between the two conditions, suggesting a possible implicit causality effect. These results will be discussed with reference to each of the three factors hypothesised to have an effect on antecedent preferences, that is, subjecthood, parallelism, and implicit verb causality.

4.1 Subjecthood

Importantly, there was no subjecthood effect anywhere in the present experiment. In none of the conditions was there a preference for resolution of the clitic to the grammatical subject; only a null result for genitive conditions and a grammatical object preference/trend for accusative and no-clitic conditions. This finding does not support Centering Theory claims that what is identified as the center of salience on the basis of grammatical role (with grammatical subjects being assumed to rank higher than all other roles both in English (Gordon *et al.*, 1993; Hudson-D'Zmura & Tanenhaus, 1998) and in Greek (Miltsakaki, 2001)) constitutes the default assignment for *any* indeterminate pronoun.

Centering Theory was not originally meant to be a model of pronoun resolution as such; it was conceived as a theory of local discourse coherence which also attempts to model the relative salience of entities (the top-down component of processes involved in anaphora resolution) in a discourse segment. If Centering Theory were to provide an adequate model of pronoun resolution, it would need to incorporate an account of how bottom-up factors (for instance, the syntactic role of the anaphor; the position of the anaphor in the linear arrangement of constituents within the clause) interact with top-down factors to produce preferred interpretations.

Moreover, the fact that there was no subjecthood effect in the two genitive conditions contradicts previous studies on parallelism in English (Smyth, 1994;

Stevenson *et al.*, 1995). These studies have demonstrated that, when strict syntactic parallelism between clauses/sentences does not obtain, a preference for the grammatical subject is attested. However, there is a crucial difference between the present study and earlier research on parallelism: all previous studies have used English materials, where object pronouns are encountered after the verb in the VP to which the pronoun belongs. Furthermore, the tasks employed by Smyth (1994) and Stevenson *et al.* (1995) involved explicit antecedent assignment at the end of each sentence. We know from online studies (Gernsbacher *et al.*, 1989) that first mention/(subjecthood) effects are stronger with longer time delays. Even though there is no simple correspondence between on-line and off-line measures, perhaps the fact that in the present experiment participants completed fragments instead of explicitly selecting the antecedent of a pronoun at the end of a complete sentence is responsible for the lack of a subjecthood effect in those conditions.

Still, the null results for the two genitive conditions require some explanation: perhaps, there is something unnatural about a genitive clitic occurring in the configuration it did in the materials; or, in the absence of a parallel antecedent, neither antecedent was sufficiently prominent in the materials used. The former possibility is suggested by the very low occurrence of spontaneously produced genitive clitics in the no-clitic conditions of the experiments: they constituted only 2.7% of the total number of observations, in comparison to accusative clitics, which were much more frequent (14.7%). The latter explanation, although counter-intuitive, is quite plausible. A central assumption of probabilistic models of language processing (MacDonald *et al.*, 1994; Trueswell & Tanenhaus, 1994) is that interactions among factors affecting processing can be non-linear: information that does not constrain interpretations when considered in isolation becomes very constraining when considered in conjunction with other information. This is the pattern of results obtained here: in a neutral context, subjecthood in the previous utterance does not determine preferences for object pronouns; implicit causality in the absence of parallelism has no constraining effect either (genitive conditions); when implicit causality is coupled with parallelism, then it has a marked effect on antecedent preferences (accusative conditions). At any rate, this issue cannot be resolved at present, and further work is needed in order to decide between the two alternative explanations.

4.2 Parallelism

Two questions were posed regarding parallelism effects: First, do they emerge independently of whether the two utterances in question share the same global constituent structure and regardless of whether the utterances contain verbs with similar semantics? Second, do pronouns need to share exactly the same grammatical role with a potential antecedent for a parallelism bias to obtain or are non-subject anaphors resolved to non-subjects in general, as Local Focusing claims?

As regards the first question, in the experiment presented here, the sentence containing the clitic was a fragment and hence it was up to the participants whether the clitic-containing sentence would share the same global constituent structure with the preceding sentence. Moreover, the verb was missing in the fragments – again supplied by the participants – and therefore, semantic similarity between the verbs in the two adjacent utterances was not an issue. And yet, with accusative clitics, structurally parallel antecedents were preferred (significantly so only for the NP2-acc condition). Consequently, it is warranted to conclude that a parallel function strategy is evident even in the absence of semantic parallelism.

With reference to the second question, unsurprisingly, the overall pattern of results disconfirmed predictions made on the basis of Local Focusing, since there was not a uniform preference for non-subject antecedents for the clitics. The syntactic role of the clitic (and the type of verb in the preceding sentence – see Section 4.3 below) had a decisive effect on preferences, and the broad generalization that non-subject pronouns are preferentially resolved to non-subjects is not supported.

4.3 Implicit Causality

Implicit verb causality had an effect on antecedent preferences in the conditions where the fragment ended with an accusative clitic or no clitic at all. Regarding the two conditions containing an accusative clitic, when implicit verb causality did not support a parallel assignment (SB verb-accusative clitic condition), there was only a tendency for clitics to be resolved to a syntactically parallel antecedent. Conversely, when case-marking *and* implicit verb causality supported a parallel assignment, there was a highly significant preference for that assignment. So, implicit causality had the effect of reducing or enhancing parallel assignments.

Importantly, this effect of implicit causality was registered in the absence of any explicit or implicit causal link between the sentence containing the biasing verb and the sentence containing the clitic. In all the texts that contained an accusative clitic in the fragment, participants' continuations were analysed in order to determine the type of coherence relation established between the two sentences of interest. In three hundred possible contexts, there was a single instance of a causal coherence relation (the item appears in ***bold italics*** in Tables 2&3 above). This effect of the causal bias of the verb in the absence of an explicit/implicit causal link and in an *inter*-sentential context challenges recent claims that verb semantics influences antecedent assignments only intra-sententially, while inter-sentential anaphora resolution is mainly determined by structural considerations (Miltsakaki, 2002, 2003).

Moreover, Stevenson *et al.* (1994, 2000) maintain that interpersonal state verbs do not influence the selection of antecedents for pronouns unless there is a causal link between the pronoun-containing and the biasing-verb-containing clauses/sentences, the full stop being assumed to function as an implicit causal connective. We established above that in the experiment in hand the full stop did not have such a function. However, even though the majority of the verbs used were state verbs (N=25), the following five verbs were not: 'eksapato' (deceive), 'katigoro' (accuse), 'parigoro' (console), 'epeno' (praise), and 'sinhero' (congratulate). In order to address Stevenson *et al.*'s (1994, 2000) claims directly, all statistical analyses were repeated excluding the five non-stative verbs. These analyses yielded identical patterns of significance with the previous analyses for all relevant tests (ANOVAs, pairwise comparisons, and one-sample t-tests). Hence, the claims put forward by Stevenson *et al.* (1994, 2000) are not supported by our results. The full stop did not function as an implicit causal connective here, and yet verb bias affected antecedent selection.

The results for the two conditions without clitics suggest that implicit verb causality may have a top-down effect on comprehension: more object resolved clitics were produced in the context of object-biasing than subject-biasing verbs (significantly so only in the items analysis). These results, however, need to be considered with extreme caution given the limited number of observations.

5 Conclusion

The experiment presented here clarified the nature and manner of interaction of subjecthood, implicit causality, and parallelism effects for the interpretation of Greek clitics. We demonstrated that in a neutral context, subjecthood has no effect on the resolution of non-subject pronouns, a result that contradicts claims made within the Centering framework. Parallelism biases were obtained in the absence of semantic parallelism and congruence of constituent structure between the relevant sentences, but only for clitics sharing the same grammatical role with a potential antecedent (*pace* Local Focusing claims). Syntactic parallelism, when combined with implicit causality bias toward the grammatical object, had the effect of making a parallel interpretation for clitics almost obligatory. Implicit verb causality had the effect of enhancing/reducing preferences for parallel assignments. Importantly, this role of implicit causality was independent of any causal construal of the link between the sentence containing the interpersonal verb and the sentence containing the anaphor. Even though the results from the two conditions without clitics hint at a top-down effect of implicit causality, there were too few observations to support any strong claims. Finally, the lack of a preference for either antecedent with genitive clitics needs to be explored in further experimental work, as the reason for this lack is not clear.

Acknowledgments

Many thanks to William Marslen-Wilson for his invaluable help and advice, and to Magda Chourdaki and Konstantina Kousta for their help with developing the materials for this experiment and recruiting participants. I would also like to thank Helen East for reading earlier drafts of this paper and making insightful comments and suggestions for changes, as well as three anonymous reviewers for their helpful comments.

References

Anagnostopoulou, E. 1996. “Experiencer-object Predicates in Greek.” *Studies in Greek Linguistics* 16.254-265.

Arnold, J. 1998. *Reference Form and Discourse Patterns*. Ph.D. thesis. Stanford University.

Au, T. K.-f. 1986. “A Verb is Worth a Thousand Words: The Causes and Consequences of Interpersonal Events Implicit in Language.” *Journal of Memory and Language* 25.104-122.

Brennan, S., M. Friedman and C. Pollard 1987. “A Centering Approach to Pronouns.” *Proc. of the 25th Annual Meeting of the ACL*. Stanford, CA. 155-162.

Brown, R. and D. Fish 1983. “The Psychological Causality Implicit in Language.” *Cognition* 14.237-273.

Chambers, C. and R. Smyth 1998. “Structural Parallelism and Discourse Coherence: A Test of Centering Theory.” *Journal of Memory and Language* 39.593-608.

Corrigan, R. 1988. “Who Dun It? The Influence of Actor-patient Animacy and Type of Verb in the Making of Causal Attributions.” *Journal of Memory and Language* 27.447-465.

Crawley, R., R. Stevenson and D. Kleinman 1990. “The Use of Heuristic Strategies in the Interpretation of Pronouns.” *Journal of Psycholinguistic Research* 19.245-264.

Fiedler, K. and G. Semin 1988. “On the Causal Information Conveyed by Different Interpersonal Verbs.” *Social Cognition* 6.21-39.

Garnham, A., M. Traxler, J. Oakhill and M. Gernsbacher 1996. “The Locus of Implicit Causality Effects in Comprehension.” *Journal of Memory and Language* 35.517-543.

Garvey, C., A. Caramazza and J. Yates 1976. “Factors Influencing Assignment of Pronoun Antecedents.” *Cognition* 3.227-243.

Gernsbacher, M. A., D. Hargreaves and M. Beeman 1989. “Building and Accessing Clausal Representations: The Advantage of First Mention Versus the Advantage of Clause Recency.” *Journal of Memory and Language* 28.735-755.

Gordon, P. and D. Chan 1995. “Pronouns, Passives and Discourse Coherence.” *Journal of Memory and Language* 34.216-231.

----------, B. Grosz and L. Gilliom 1993. “Pronouns, Names, and the Centering of Attention in Discourse.” *Cognitive Science* 17.311-347.

---------- and K. Scearce 1995. "Pronominalization and Discourse Coherence, Discourse Structure and Pronoun Interpretation." *Memory and Cognition* 23.313-323.

Greene, S. and G. McKoon 1995. "Telling Something we Can't Know: Experimental Approaches to Verbs Exhibiting Implicit Causality." *Psychological Science* 6.262-270.

Grober, E., W. Beardsley and A. Caramazza 1978. "Parallel Function Strategy in Pronoun Assignment." *Cognition* 6.117-133.

Grosz, B., A. Joshi and S. Weinstein 1995. "Centering: A Framework for Modeling the Local Coherence of Discourse." *Computational Linguistics* 21.203-225.

Gundel, J., N. Hedberg and R. Zacharski 1993. "Cognitive Status and the Form of Referring Expressions in Discourse." *Language* 69.275-307.

Holtgraves, T. and S. Raymond 1995. "Implicit Causality and Memory: Evidence for a Priming Model." *Personality and Social Psychology Bulletin* 21.5-12.

Holton, D., P. Mackridge and I. Philippaki-Warburton 1997. *Greek: A Comprehensive Grammar of the Modern Language*. London: Routledge.

Hudson-D'Zmura, S. and M. Tanenhaus 1998. "Assigning Antecedents to Ambiguous Pronouns: The Role of the Center of Attention as the Default Assignment." M. Walker, A. Joshi, and E. Prince (eds.) *Centering Theory in Discourse*. Oxford: Clarendon Press.

Kehler, A. 2000. "Coherence and the Resolution of Ellipsis." *Linguistics and Philosophy* 23.533-575.

Kennison, S. and P. Gordon 1997. "Comprehending Referential Expressions during Reading: Evidence from Eye Tracking." *Discourse Processes* 24.229-252.

Koornneef, A., N. Waaijer and J. van Berkum 2002. "'David Praised Linda Because He (?) Was Proud'—Implicit Causality Information in Verbs Immediately Affects Sentence Interpretation." Poster presented at the *8th Annual Conference on Architectures and Mechanisms for Language Processing*, Tenerife, Canary Islands, Spain.

Long, D. and L. De Ley 2000. "Implicit Causality and Discourse Focus: The Interaction of Text and Reader Characteristics in Pronoun Resolution." *Journal of Memory and Language* 42.545-570.

McDonald, J. L. and B. MacWhinney 1995. "The Time Course of Anaphor Resolution: Effects of Implicit Verb Causality and Gender." *Journal of Memory and Language* 34. 543-566.

MacDonald, M., N. Pearlmutter and M. Seidenberg 1994. "Lexical Nature of Syntactic Ambiguity Resolution." *Psychological Review* 101.676-703.

Miltsakaki, E. 2001. "Centering in Greek." To appear in *Proc. of the 15th Int. Symposium on Theoretical and Applied Linguistics*, Aristotle University of Thessaloniki, Greece.

---------- 2002. "Towards an Aposynthesis of Topic Continuity and Intrasentential Anaphora." *Computational Linguistics* 28(3).319-355.

---------- 2003. *The Syntax-Discourse Interface: Effects of the Main-Subordinate Distinction on Attention Structure*, Ph.D. dissertation, University of Pennsylvania, PA.

Pearson, J., R. Henschel and R. Stevenson 2000. "Centering and Parallelism in Pronoun Resolution: Inanimate Referents and Sentence Structure." Poster presented at the *13th Annual CUNY Conference on Human Sentence Processing*, San Diego, CA.

Sidner, C. 1979. "A Computational Model for Co-reference Comprehension in English." Ph.D. dissertation, MIT, Cambridge, MA.

Smyth, R. 1992. "Multiple Feature Matching in Pronoun Resolution: A New Look at Parallel Function." J. Ohala, T. Neary, B. Derwing, M. Hodge, and G. Wiebe (eds.) *Proc. of the 2nd Int. Conference on Spoken Language Processing*. Edmonton: Priority Printing. 145-148.

---------- 1994. "Grammatical Determinants of Ambiguous Pronoun Resolution." *Journal of Psycholinguistic Research* 23.197-229.

Stevenson, R. 1996. “Mental Models, Propositions and the Comprehension of Pronouns.” J. Oakhill and A. Garnham (eds.) *Mental models in cognitive science: Essays in honour of Phil Johnson-Laird*. Hove: Psychology Press. 53-76.

----------, R. Crawley and D. Kleinman 1994. “Thematic Roles, Focus and the Representation of Events.” *Language and Cognitive Processes* 9.519-548.

----------, A. Knott, J. Oberlander and S. McDonald 2000. “Interpreting Pronouns and Connectives: Interactions Among Focusing, Thematic Roles and Coherence Relations.” *Language and Cognitive Processes* 15.225-262.

----------, A. Nelson and K. Stenning 1995. “The Role of Parallelism in Strategies of Pronoun Comprehension.” *Language and Speech* 38.393-418.

Stewart, A., M. Pickering and A. Sanford 2000. “The Time Course of the Influence of Implicit Causality Information: Focusing Versus Integration Accounts.” *Journal of Memory and Language* 42.423-443.

Suri, L. and K. McCoy 1993. “Comparing Local Focusing and Centering.” Paper presented at the *Workshop on Centering in Naturally-Occurring Discourse.* Institute for Research in Cognitive Science, University of Pennsylvania, PA.

---------- and ---------- 1994. “RAFT/RAPR and Centering: A Comparison and Discussion of Problems Related to Processing Complex Sentences.” *Computational Linguistics* 20(3).301-317.

----------, ---------- and J. DeCristofaro 1999. “A Methodology for Extending Focusing Frameworks.” *Computational Linguistics* 25(2).173-194.

Trueswell, J. and M. Tanenhaus 1994. “Toward a Lexicalist Framework for Constraint-based Syntactic Ambiguity Resolution.” C. Clifton, L. Frazier, and K. Rayner (eds.) *Perspectives in Sentence Processing*. Hillsdale, NJ: Lawrence Erlbaum Associates. 155-179.

Different Forms Have Different Referential Properties: Implications for the Notion of 'Salience'

Elsi Kaiser
University of Pennsylvania

I present three psycholinguistic experiments that investigate the effects of (a) word order, (b) grammatical role and (c) referential form on how a referent is subsequently referred to. Specifically, the studies test the referential properties of two third person anaphors in Finnish: the gender-neutral pronoun *hän* 's/he' and the demonstrative *tämä* 'this,' which can also be used for human referents. The results of the experiments indicate that both *hän* 's/he' and *tämä* 'this' can be sensitive to multiple factors, but in strikingly different ways. This suggests that we should not try to define the referential properties of these two forms in terms of a single unified notion of salience. I investigate a possible alternative way of coherently grouping the referential properties of *hän* and *tämä* that captures the patterns we see in the data without requiring an accessibility hierarchy.

1 Introduction[*]

One of the best-known observations in the reference resolution literature concerns the connection between the form of a referring expression and the accessibility/salience of its referent. It has often been noted that the most salient referents – i.e. those referents that are currently at the centre of attention and most prominent at that point in the discourse – are referred to with the most reduced referring expressions (e.g. pronouns in English, null *pro* in Spanish). This raises the question: What kinds of factors influence a referent's salience, i.e. make it a good candidate to be referred to with a reduced anaphoric expression?[1] In this paper, I focus on three properties that have been claimed to influence the salience of a referent – and thus these properties are also predicted to have an impact on what kinds of referring expressions can be used in subsequent discourse to refer back to this referent: (1) grammatical/syntactic role, (2) word order and (3) referential form. For example, does being in subject position make a referent more salient than being in object position? Does

[*] Thanks to Cassie Creswell, Miriam Eckert, Ritva Laury, Mikhail Masharov, Eleni Miltsakaki, Kimiko Nakanishi, Ellen Prince, Maribel Romero, Kieran Snyder, John Trueswell, Jennifer Venditti, three anonymous DAARC reviewers, and the psycholinguistics group at Penn for useful comments and insights. Earlier discussions of some of the topics addressed in this paper can be found in Kaiser (2003) and Kaiser (in press). All errors are mine.

[1] The terms 'salience' and 'accessibility' are used in somewhat different ways by different groups of researchers. In this paper, I will primarily use the term 'salience.' Thus, in this paper, the most 'salient' referent is the referent that is at the center of attention at that point in the discourse.

occurring sentence-initially increase a referent's salience? Is a pronominalized referent more salient than a full NP referent?

I will address these issues from the perspective of Finnish, a highly inflected, flexible word order language with canonical SVO order (Vilkuna, 1995). Standard Finnish has two kinds of third person anaphors: the gender-neutral pronoun *hän* 's/he' and the demonstrative *tämä* 'this'. (Dialects of colloquial Finnish have somewhat different anaphoric systems, see e.g. (Laitinen, 1992; Seppänen, 1998; Laury, to appear)). I present the results of three psycholinguistic experiments investigating the referential properties of these two anaphors, and I show that *hän* and *tämä* differ in their referential properties. As we will see, both *hän* 's/he' and *tämä* 'this' can be sensitive to multiple factors, but in strikingly different ways. The results indicate that instead of trying to define the referential properties of these forms according to a unified notion of salience, we should investigate how (and to what degree) different factors may be relevant for different referential expressions.

The structure of this paper is as follows. First, in Section 2, I review existing work on referent salience. In Section 3, I discuss the word order patterns and anaphoric system of Finnish. Sections 4 and 5 present the results of the sentence-completion experiments and discuss their implications. Conclusions and directions for future work are addressed in Section 6.

2 Salience and the form of referring expressions

There exists a general consensus that the more reduced an anaphoric expression is, the more salient its antecedent has to be, and that "pronouns are used most often when the referent is represented in a prominent way in the minds of the discourse participants, but more fully specified forms are needed when the representation of the referent is less prominent" (Arnold, 1998:4). This correlation is encoded in various accessibility hierarchies of referential forms that have been proposed in the literature (see e.g. (Gundel *et al.*, 1993; Givón, 1983; Ariel, 1990)). According to these hierarchies, overt pronouns are used for more accessible antecedents than demonstratives, and null pronouns for more accessible referents than overt pronouns, and so on.[2] In a schematic form, the hierarchy can be represented as follows: *null pronoun* > *unstressed/bound pronoun* > *stressed/independent pronoun* > *demonstrative* > *full NP*, and so on. Now, keeping these claims in mind, let us turn to the question of what makes a referent highly accessible/salient. Various factors have been claimed to have an

[2] The fact that some referential forms in some languages can provide information about things such as number, gender, animacy or 'humanness' (e.g. English *it* vs. *he/she*) is clear and I do not address it here. I focus here on choices in referential form that cannot be explained by these kinds of factors.

impact on referent salience. Here, we will focus on three: word order, syntactic role and referential form.

2.1 Syntactic function

Previous research has found a close connection between grammatical roles and salience – specifically, that subjects are more salient than objects (e.g. (Brennan *et al.*, 1987; Chafe, 1976; Matthews & Chodorow, 1988; Stevenson *et al.*, 1994)). Many researchers use anaphoric elements as a window to gain insight into what makes referents salient, and thus rely on the finding that the most reduced anaphoric element in a given language refers to the most salient referent. For example, Crawley and Stevenson (1990) conducted a sentence continuation experiment where participants were asked to write continuations for stories like “Shaun led Ben along the path and he…”. The continuations were then analysed to see how people interpret the pronoun *he*, which is assumed to refer to the most salient entity. The results show that the pronoun is interpreted as referring back to the subject significantly more often than to the object. The same subject advantage was found in reading-time studies (e.g. (Gordon *et al.*, 1993)) and corpus studies (e.g. (Brennan *et al.*, 1987, Tetreault, 2001)). Referents in subject position seem to be ‘default topics’ in that they are more likely to be the antecedents of pronouns than entities in other grammatical positions. However, in languages like English that have relatively rigid subject-object order, it is difficult to tell whether the increased salience of subjects is due to their position at the beginning of the sentence, or their semantic/thematic properties.

2.2 Word order

In order to tease apart subjecthood and first position, and to see whether word order itself influences salience, we can turn to languages with flexible word order. Existing research reveals different findings for different languages. For example, for German, Rambow (1993) found that word order in the *Mittelfeld* correlates with salience and guides pronoun resolution (see also (Choi, 1996; Lenerz, 1977; Strube & Hahn, 1999) on German word order), whereas Turan (1998) and Hoffman (1998) claim that in Turkish, the salience of a referent correlates with its grammatical (or semantic) role, and is not affected by word order. Similarly, Prasad and Strube (2000) claim that in Hindi, grammatical role, not word order, is what determines salience.

Let us first consider Rambow’s German examples in (1a,b). According to Rambow, the pronoun in the answer tends to refer to the leftmost constituent in the *Mittelfeld*. Thus, when the constituent order is changed from subject-object

(1a) to object-subject (1b), the preferred referent of the pronoun in (1c) changes.

(1) a. subject-object order
Glauben Sie, dass [eine solche Maßnahme]$_a$ [der russischen Wirtschaft]$_b$
Think you that [a such measure]-NOM [the Russian economy]-DAT
helfen kann?
help can?
"Do you think that such a measure can help the Russian economy?"
b. object-subject order
Glauben Sie, dass [der russischen Wirtschaft]$_b$ [eine solche Maßnahme]$_a$ helfen kann?
c. *answer*:
Nein, ***sie*** *ist viel zu primitiv.*
No, she is much too primitive.
"No, it (she) is much too primitive." (1a) it=measure (1b) it=economy

In contrast, in the Turkish counterparts to Rambow's examples, discussed by Turan (1998:142), the null pronoun in the answer is interpreted as referring to the subject, regardless of whether the word order is SO or OS. In other words, the claim is that in Turkish, subjects are more salient than objects even in scrambled sentences where the object linearly precedes the subject.

When faced with these data, it is worth keeping in mind that the functions of scrambling vary across languages, and even in different constructions within a single language. In my opinion, this may well be at least part of the reason for the crosslinguistic variation. In fact, Rambow (1993) shows that, in German, 'topicalized' word orders sometimes have an impact on salience and at other times they do not. He claims that in German, whether or not salience is determined by word order depends on the discourse function of the topicalization. Thus, before concluding that languages differ in terms of whether word order variation influences salience, we need to consider the functions of different word orders in those languages.

2.3 Referential form

The impact of referential form on referent salience has not received as much attention in the literature as grammatical role or word order. While it is widely accepted that the most reduced forms are used to refer back to the most accessible referents, not as much is known about the effect of a referent's form on the interpretation of reduced forms in the *following* discourse.

In other words, if being a subject or linearly sentence-initial increases a referent's salience and thus makes it more likely to be referred to with a reduced form in upcoming discourse, then we might also wonder whether being a reduced form increases a referent's salience, which in turn will influence interpretation preferences of future reduced forms. For example, does a referent become more salient for future reference when it is expressed with a pronoun, rather than a definite NP? According to Kameyama (1999), the referential form used to refer to an entity can indeed affect that entity's salience. More specifically, she argues that a pronominalized referent in non-subject position 'gains' in salience by virtue of being pronominalized, and in fact becomes so salient that it 'competes' with a non-pronominalized entity in subject position.

Similarly, Beaver (to appear) suggests an Optimality-theoretic approach to anaphora resolution that includes a constraint called SALIENT FORM. This constraint states that "If in the previous sentence discourse entity α was realized by a more minimal form than discourse entity β, then α is more salient than β" (Beaver, to appear:28).[3] In related work, Strube & Hahn (1999), on German, and Strube (1998), on English, argue that NP form is correlated with salience.

It is worth emphasizing, as Beaver points out, that the constraint SALIENT FORM is crucially different from the idea that the most salient referents are referred to with the most reduced forms, since Beaver's constraint "implies that being pronominalized makes a referent salient in the future" (Beaver, to appear:28-29 fn). In sum, there is some work claiming that referential form has an effect on salience, but this issue has not received as much attention as other factors, and so many questions remain open.

3 Finnish

The grammatical properties of Finnish make it a good testing ground for investigating how word order and grammatical role influence a referent's chances of being referring to in subsequent discourse with a reduced anaphoric expression. Finnish has flexible word order, and two third person anaphors (the pronoun *hän* 's/he' and the demonstrative *tämä* 'this'). In this paper, in addition to testing how word order and grammatical role affect the referential properties of these anaphors, we will also investigate whether the referential form used for a referent (e.g. whether a referent is expressed with a pronoun or a full NP) affects the role that referent plays in subsequent discourse. First, however, let us take a closer look at Finnish word order and the anaphoric options in Finnish.

3.1 Word order

Finnish has flexible word order and no definite or indefinite articles.[4] The canonical word order is SVO, but all six permutations of these elements are

[3] Note that in Optimality Theory, constraints are ranked with respect to each other, and low-ranked constraints can be violated in order to satisfy higher-ranked constraints. Constraint rankings can vary across languages. Thus, SALIENT FORM should not be viewed as an absolute principle.

[4] In dialects of spoken Finnish, the demonstrative pronoun *se* 'it' is evolving into a kind of definite article (Laury, 1997). However, this does not occur in standard Finnish.

grammatical in the appropriate contexts (Vilkuna, 1989, 1995:245). In this paper, we will focus on SVO and OVS orders. Why might one expect the SVO/OVS variation in Finnish to influence referent salience? A look at the discourse factors guiding this alternation provides an answer to this question. In Finnish, the discourse-status of the arguments, i.e. whether or not they have been mentioned in the preceding discourse, guides the choice between SVO and OVS order. If we combine this observation with the claim by Strube and Hahn (1999) that discourse-status determines salience (i.e. discourse-old entities are more salient than discourse-new ones), then we predict that the Finnish SVO/OVS variation determines the subject and object's salience.

Before we turn to the experiments to see if this prediction is supported, we will consider the discourse properties of subjects and objects in SVO and OVS orders in more detail (see also (Chesterman, 1991; Hiirikoski, 1995; Vilkuna, 1989, 1995; Helasvuo, 2001) for further discussion). First, let's look at subjects. Postverbal subjects – which are in a noncanonical position – are used to introduce referents that are discourse-new, i.e. have not yet been mentioned in the current discourse.[5] This is illustrated by (2), where the English original has a sentence-initial indefinite noun phrase. In Finnish, there is no indefinite article, and the noun phrase occurs postverbally. In contrast, preverbal subjects are usually discourse-old information, i.e. refer to entities that have already been mentioned in the current discourse. This is exemplified by the Finnish translation in (3). A preverbal subject NP is interpreted as being new information only when the sentence is a discourse-initial 'all new' utterance.

(2) postverbal subject

a. **A great big water-beetle** came up underneath the lily leaf.

b. *Lumpeenlehden* *alla* *ui* ***iso*** ***vesikuorianen***.
lily-leaf-GEN under swam **big-NOM** **water-beetle-NOM**.
(from Beatrix Potter's *The tale of Mr Jeremy Fisher* (1979), (Chesterman, 1991:100))

(3) preverbal subject

a. **The tyrannosaur** was very close now. (Crichton, 1995:40)

b. ***Tyrannosaurus*** *oli* *jo* *hyvin* *lähellä.* (Finnish translation:276)
"**Tyrannosaur-NOM** was already very close."

Let us now consider the discourse properties of objects. A preverbal object, followed by a postverbal subject, as illustrated in (4), is interpreted as

[5] In Finnish, the distinction between old and new information depends on the discourse status of the entities, not on whether or not they are known/old to the hearer (hearer-status). This is shown by the fact that names of family members or famous people (hearer-old) can surface as post-verbal subjects in Finnish, if they are discourse-new. (see ex. (4) below). See (Prince, 1992) for further discussion of the notions of discourse-status and hearer-status.

discourse-old information. Postverbal objects can be interpreted as new or old information, as shown in (5).

(4) OVS order

Tiedotteen	*välitti*	*julkisuuteen*	*kurdien*	*uutistoimisto*	*DEM.*
Announcement-ACC	transmitted	public-to	Kurds'	newsoffice-NOM	DEM.

"**The announcement** was made public by the Kurdish newsoffice D.E.M."
(from the newspaper *Aamulehti* 16.3.1999)

(5) SVO order

Mies	*huomasi*	***kissan***.
Man-NOM	noticed	**cat-ACC**

"The man noticed **a/the cat**."

3.2 Anaphoric forms of standard Finnish

In the previous section, we reviewed the pragmatic characteristics of SVO and OVS order, and saw that these two orders differ in terms of the discourse status of the subject and the object. Now, keeping in mind Strube and Hahn's (1999) claim that discourse status determines salience, we will take a closer look at previous work on the referential properties of the two anaphoric forms *hän* 's/he' and *tämä* 'this.' Previous work on the referential properties of the gender-neutral pronoun *hän* 's/he' supports the crosslinguistic generalization that overt pronouns (in languages that lack null pronouns) refer to the most salient entities. The pronoun *hän* has been described as referring to the most central or 'foregrounded' character (Kalliokoski, 1991) or to the character that is most important in a given situation or context (Vilppula, 1989; *inter alia*). Moreover, according to Saarimaa (1949), *hän* tends to refer to the subject of the preceding sentence because the subject is more in the 'foreground' than other referents realized in other positions.

The demonstrative *tämä* 'this' differs from the pronoun *hän* in that it can be used as a proximal demonstrative or a discourse deictic (as in English, see also (Etelämäki, 1996)), in addition to being used to refer to human antecedents. The referential properties of anaphoric *tämä* also differ from those of *hän*: Whereas the pronoun is used for foregrounded characters, *tämä* has been described as referring to characters in the background (Varteva, 1998) – which fits in with the accessibility-hierarchy based claim that pronouns are used for more accessible referents than demonstratives. In more structural terms, Sulkala and Karjalainen (1992) note that *tämä* is "used to indicate the last mentioned out of two or more possible referents" (1992:282-283). This brings up the question: Does the demonstrative refer to the last mentioned entity regardless of grammatical role? What happens with OVS order? Saarimaa (1949) argues that *tämä* 'this' refers to a recently mentioned, non-subject referent and that *hän* is used for subject antecedents. However, the question remains: In actual language use, is this the case?

A partial answer is offered by a corpus study of referential expressions in written Finnish by Halmari (1994); for spoken Finnish, see (Seppänen, 1998). Halmari's corpus contained 433 pronoun tokens, and 15 demonstrative tokens.[6] As she notes, "the huge number of pronouns in the sample skews the percentages, and this is a problem that needs to be addressed in future research" (Halmari, 1994:55). As Table 1 shows, she found that *hän* refers to subjects and the demonstrative *tämä* tends to refer to objects, and so she concludes that *hän* is used for highly salient referents, and *tämä* for less salient referents. Her corpus findings confirm the intuitions of other Finnish researchers – but do not give a conclusive answer to the question of how word order affects the referential properties of *hän* and *tämä*, as she did not analyze word order in her corpus study.[7]

	hän (pronoun)		*tämä* (demonstrative)	
Subject	314	72.5%	2	13%
Direct object	26	26%	4	27%
Indirect object	7	1.5%	2	13%
Oblique	31	7%	4	27%
Genitive	55	13%	3	20%
Total	**433**	**100%**	**15**	**100%**

Table 1: Referring expressions and grammatical role of antecedent (Halmari 1994:53)

To address the imbalance of pronoun and demonstrative tokens in Halmari's corpus, I conducted a corpus study (Kaiser, 2000) of 103 occurrences of *hän* 's/he' and 101 occurrences of anaphoric *tämä* 'this' in the novel *Tuntematon Sotilas* by V. Linna (1954/1999, Helsinki:WSOY). The results for *hän* are in Table 2. In general, *hän* 's/he' tends to refer to a preceding subject (43 out of 60 cases, 71.67%).[8] In contrast, *tämä* tends to have a non-subject antecedent (Table 3). Examples are in (7) and (8) (bolded constituents are coreferential).

[6] Halmari (1994) focused on a wide range of different referential expressions, and thus the pronoun-demonstrative distinction was not the primary focus of her investigation.

[7] Importantly, however, Halmari (1994) conducted a small survey and asked seven native speakers about sentences with different word orders and different anaphoric elements. She tested the OVS sentence *Kanan näki kissa ja {se/tämä} kuoli.* 'Chicken-ACC saw cat-NOM and {it/this} died.' People were given the sentence either with *se* 'it' or *tämä* 'this' and were asked 'Who died?' With the pronoun *se* 'it', there was a preference to interpret it as referring to the object *chicken* (presumably for pragmatic reasons, as a cat seeing a chicken is likely to result in the chicken dying, rather than the cat), and with the demonstrative *tämä* 'this', people did not give very clear responses and found the resulting sentence "extremely hard to process" (Halmari, 1994:42).

[8] These data are for cases where the anaphor and its antecedent are in distinct main clauses. Subordinate clauses were also analyzed, but are not included here, so the totals shown here are less than 101 and 103. See (Kaiser, 2000) for details.

(6) Example with *hän* (Linna, 144)

Sitten ***eversti*** *piti puheen.* ***Hän*** *koetti saada ääneensä*
Then **colonel** held speech. **He-NOM** tried get voice-into-his
tiettyä toverillista sävyä.
certain-PART friendly-PART tone-PART.
"Then the **colonel** gave a speech. **He** tried to get a certain friendly tone into his voice."

(7) Example with *tämä* (Linna, 286)

Lammio huusi ***Mielosta,*** *ja* ***tämä*** *tuli sisään*
Lammio shouted **Mielonen-PART**, and **this-NOM** came in
lähetit kannoillaan.
messengers heels-on-his.
"Lammio called for **Mielonen**, and **he** came in with the messengers on his heels."

Antecedent of *hän*		**Antecedent of *tämä***	
Role of antecedent	**Number of occurrences**	**Role of antecedent**	**Number of occurrences**
S	43 (71.67%)	S	7 (18.92%)[9]
Poss[10]	10 (16.67%)	Poss	5 (13.51%)
DO	1 (1.67%)	DO	13 (35.14%)
IO	3 (5%)	IO	1 (2.70%)
Oblique	3 (5%)	Oblique	6 (16.22%)
PP	-	PP	5 (13.51%)
Total	60	**Total**	37

Table 2: Antecedent of hän *Table 3: Antecedent of tämä*

In sum, the results of both Halmari (1994) and Kaiser (2000) show that there is a correlation between anaphoric form and grammatical role. Subjects are usually referred to with *hän*, and objects and oblique arguments with *tämä*. Should we conclude, then, that subjects have a higher level of salience than objects or obliques? Not necessarily. We know that in SVO order, the subject tends to be referred to with the 'salient anaphor' *hän*, but we don't yet know if this is due to linear order or grammatical function. To tease apart these factors, we need to look at the referential properties of *hän* and *tämä* for sentences where the object precedes the subject. However, finding sufficient numbers of such examples in an unparsed corpus is difficult. The corpus used by Kaiser (2000) did not contain any examples of transitive verbs in OVS sentences that had a human third person subject and a human third person object, followed by *hän/tämä*.

---

[9] The demonstrative *tämä* is used occasionally to refer to postverbal subjects, e.g.:
(i) *Yhdyshaudan kulman takaa häämötti* ***mies,***
trench-GEN corner-GEN behind was-vaguely-visible **man**-NOM,
ja vain silmänräpäyksen ***tämä*** *ehti epäröidä*
and only eyeblink-ACC **this-NOM** had-time to hesitate
"Behind the corner of the trench, a **man** was dimly visible, and **he** only had a moment to hesitate…" (Linna, 331)

[10] 'Poss' stands for possessive/genitive forms, e.g. [**his** book], [**Peter's** book].

To circumvent this problem, in the experiments reported here I use sentence completion tasks. In these kinds of experiments, participants are given sentences or sentence fragments and asked to provide natural-sounding continuations. The continuations are analysed to see how the participants interpreted the sentence. In Sections 4 and 5, I present the results of three written sentence completion studies which investigate how (i) the grammatical function, (ii) linear position and (iii) referential form of potential antecedents influence the referential properties of *hän* and *tämä*. As we saw in the preceding sections, the first two factors, grammatical function and linear role, have been investigated for certain languages including German, English and Hindi, but the current findings are rather contradictory. In contrast to the other two factors, the role of referential form has not received much attention in the literature, and many intriguing questions remain open. I investigate some of them in the third study.

3.3 Predicted referential patterns

In this section I discuss the predictions we can make about effects of referential form, word order, and grammatical role on the referential properties of the pronoun *hän* and the demonstrative *tämä* for SVO and OVS word orders. In this discussion, I assume that the pronoun *hän* refers to highly accessible referents and the demonstrative *tämä* to less accessible referents – i.e., as predicted by accessibility hierarchies. It will become clear later that this assumption is overly simple, but at this stage it is useful for presenting the predictions.

First, let us look at grammatical role and word order, and then focus on the effects of referential form. Let us start by hypothesizing that syntactic function is the determining factor for salience, and that subjects are more salient than objects (as illustrated schematically in option (a) in Table 4). If this is the case, we expect an occurrence of the pronoun *hän* (in subject position) in the subsequent sentence to refer to the subject of the preceding utterance, and an occurrence of the demonstrative *tämä* (in subject position) to refer to the object of the preceding utterance, in both SVO and OVS order.[11] Second, if word order is the determining factor, with constituents to the left being more salient than those to the right (option (b)), then we expect an occurrence of *hän* in the

[11] Here, and also in the experiments, we are primarily focusing on the referential properties of *hän* and *tämä* when they are in the subject position of the subsequent utterance. This was done in order to control for effects of parallelism (Sheldon, 1974; Smyth & Chambers, 1998). Since I am comparing the referential properties of the pronoun *hän* in subject position and the demonstrative *tämä* in subject position, any differences in the referential properties of the anaphoric expressions must be due to the anaphoric forms themselves. See also Experiment 2 for discussion of the effects of anaphoric form on a referent's salience in subsequent discourse.

next utterance to refer to the preverbal constituent of the preceding utterance and *tämä* to the postverbal one, regardless of grammatical role.

Salience is determined by:	*Predicted referential pattern:*
(a) Syntactic function (subjects > objects)	S…..V..…O O…..V…..S *hän* *tämä* *tämä* *hän*
(a) Word order (left > right)	S…..V..…O O…..V…..S *hän* *tämä* *hän* *tämä*

(c) Both factors are equally relevant:	S.	…V…..	O	O..	…V…..S
(i) Syntactic function	1		0	0	1
(ii) Word order	1		0	1	0
	2		0	1	1
	hän		*tämä*	*??*	*??*

Table 4. Predictions

Third, if both word order and grammatical role play a role in influencing referent salience, what happens? To represent the possibility of both factors having an impact on referent salience, in part (c) of Table 4, 'salience points' are allotted to the arguments. According to the syntactic-function approach, subjects are more salient than objects. Consequently, the subject gets a point in both SVO and OVS order. According to the word-order approach, the linearly initial constituent is more salient. As a result, in SVO order, the subject gets a point, and in OVS order, the object does. Summing the points for the two arguments, the subject clearly comes out as being the more salient argument in SVO order. In contrast, with OVS order, the point totals for the two arguments are equal. Thus, if both word order and grammatical function affect referent salience equally, it is not clear what will happen with OVS order, where the two factors are pitted against each other.[12] Of course, it could also be the case that both word order and grammatical function matter, but that one is more important than (or 'weighted more heavily than') the other. This would resolve the ambiguity for OVS order. This possibility is discussed in more detail below.

[12] This discussion raises the question: Should salience be viewed as increased 'activation' (more points) as I have done here? Or it is rather a matter of 'suppression' (see e.g. (Gernsbacher, 1990)), such that less salient referents lose points or receive negative points, so to speak? The distinction is not crucial to the present discussion, but poses interesting questions for future research.

Finally, let us consider what would happen if referential form determines salience – e.g., if pronominalized entities are more salient than entities referred to with a full NP, as Kameyama (1999) and Beaver (to appear) suggest. If we assume the strongest version of the 'pronoun salience' approach, we predict that any entity that is realized as a pronoun (regardless of this pronoun's grammatical role or linear position) will be referred to with the pronoun *hän* in the subsequent utterance, and we also predict that a nonpronominalized referent will be less salient and thus will be referred to with the demonstrative *tämä*. Of course, it might also be the case that a referent's salience is affected by multiple factors, e.g. the effects of referential form might turn out to interact with effects of syntactic function and/or word order (see also (Arnold, 1998)).

In the next sections, we will turn to the sentence completion experiments designed to investigate the effects of word order, grammatical role and referential form on the referential properties of the pronoun *hän* and the demonstrative *tämä*. The first two experiments focus on the effects of word order and grammatical role, and the third experiment addresses the impact of referential form.

4 Experiment 1

This experiment (discussed in detail in Kaiser (2003)) investigates how word order and grammatical role influence the referential properties of *hän* and *tämä*. The test items consisted of written SVO and OVS sentences (both the subject and object were full NPs), followed by the first word of the subsequent sentence, either the pronoun *hän* 's/he' or the demonstrative *tämä* 'this'. These 'prompt words' were in the nominative case, i.e. in subject position. Anaphor type and word order were crossed to create four conditions: [SVO.Hän...], [OVS.Hän...], [SVO.Tämä...] and [OVS.Tämä...]. The participants were asked to write a natural-sounding completion for the prompt. An example item is shown below.

(8) [SVO.Hän]

Lääkäri	*onnitteli*	*opiskelijaa.*	*Hän...*
Doctor-NOM	congratulated	student-PART.	S/he-NOM…

"A/the doctor congratulated a/the student. S/he..."

Thirty-two native Finnish-speakers participated in this experiment. Each participant was asked to complete 38 items: 8 critical items and 30 filler items whose order was randomised. The nouns used for the subject and object in the critical items were all 'occupational labels' or other roles (e.g. doctor, nurse, student). These types of nouns were used to make the continuations easier to interpret while coding. The verbs were agent-patient verbs (as defined by

Stevenson *et al.* (1994)), and we decided to use a unified verb group in order to control for possible focusing effects. Continuations were coded in terms of which of referent in the preceding sentence the participants chose as the referent of the anaphoric expression. When it was not clear from the continuation which referent the participant had interpreted as being the antecedent of the pronoun or demonstrative, the item was coded as 'unclear.' In addition, with *tämä*, there were some continuations where *tämä* was not used as an anaphor for one of the two characters mentioned in the preceding sentence, and was instead used as a discourse deictic (e.g. '*This* was a friendly thing to do') or in some other way. These types of uses were coded as 'demonstrative' uses, so as to set them apart from the anaphoric uses.

4.1 Results

The results are shown in Table 5. As the percentages show, word order affects the referential properties of *hän* and *tämä* differently. The pronoun *hän* 's/he' is usually interpreted as referring to the subject, regardless of whether the word order is SVO or OVS. In the SVO.Hän condition, the pronoun was interpreted as referring to the preceding subject in 62.5% of the continuations and to the object in 22% of the continuations. In the OVS.Hän condition, there were 61% subject-interpretations and 25% object-interpretations. However, in the SVO.Tämä condition, *tämä* has a strong preference to refer to the preceding object; it refers to the object in 83% of the cases. In the OVS.Tämä condition, however, *tämä* is fairly evenly split between the subject and the object (37% object-interpretations, 33% subject-interpretations).

	Subject	**Object**	**Demonstrative**	**Unclear/other**
SVO.Hän	62.5%	22%	0	15.5%
OVS.Hän	61%	25%	0	14%
SVO.Tämä	1.5%	83%	12.5%	3%
OVS.Tämä	33%	37%	16%	14%

Table 5. Referent of hän 's/he' or tämä 'this'

4.2 Adding a discourse context: Experiment 1b

Before moving onto a discussion of the results presented above, let's briefly consider a related sentence-completion experiment that situated the sentences in discourse contexts (see (Kaiser (2003) for more details). Recall that the SVO/OVS variation in Finnish is driven by the discourse status of the arguments. Thus, in Experiment 1, the OVS sentences were infelicitous, because they were presented without a preceding context. In this second experiment, Experiment 1b, a brief discourse context preceded the SVO/OVS sentences, such that the preverbal noun (S or O) was discourse-old, and the postverbal noun (O or S) was discourse-new.[13] Both were full NPs. We can view Experiment 1 as a kind of baseline experiment, one that tells us what happens in the absence of any context, and Experiment 1b as a 'fairer test' of how word order and grammatical role impact the referential properties of *hän* and *tämä.* Also, in Experiment 1b, the number of critical items was increased to sixteen, and they had the same structure as in Experiment 1: The nouns used were roles/occupational labels, and only agent/patient verbs were used. Sixteen native Finnish speakers participated in this study.

Contrary to what one might expect at first blush, the results of Experiment 1b are not quite the same as those of Experiment 1. However, for three out of four conditions, the results of this experiment largely replicate the findings of Experiment 1. In Experiment 1b, in the SVO.Hän condition and the OVS.Hän condition, we replicate for the pronoun *hän* the subject-preference from the first experiment. There are five times more subject-interpretations than object-interpretations for *hän*, regardless of whether the word order is SVO or OVS.
The demonstrative *tämä* reveals a slightly more complex pattern. In the SVO.Tämä condition of Experiment 1b, *tämä* has a very strong preference to refer to the postverbal argument of the preceding sentence (over 80% object-interpretations) – just like in the first experiment. However, with OVS order, *tämä* shows a preference for the postverbal subject over the preverbal object: there are almost five times more subject-continuations than object-continuations. This contrasts with Experiment 1, where in OVS order *tämä* was used to refer to the subject and the object almost equally often. In Experiment 1b, then, *tämä* prefers the postverbal referent regardless of that referent's grammatical role.

4.3 Discussion

If we compare the results of Experiments 1 and 1b, we see that the results are basically the same, except for the OVS.Tämä condition, as discussed above. Taken as a whole, what do these results tell us about the referential properties of *hän* and *tämä*? In addition, why does *tämä* show such a clear preference for postverbal reference in both SVO and OVS conditions in Experiment 1b, but not in Experiment 1?

Remember that one of the possible predictions sketched out in section 3.3 was that, if grammatical role determines salience, and *hän* is used for more salient referents and *tämä* for less salient referents, *hän* is used to refer back to subjects and *tämä* to objects, regardless of word order. While this prediction fits

[13] The contexts were created such that two full NPs could be felicitously used in the critical SVO/OVS sentence. This was done by means of two context sentences which mention a third referent. See (Kaiser, 2003) for details.

the pattern we saw for the pronoun *hän* in the completions, it clearly does not match what we saw happen with the demonstrative *tämä*.

With preceding SVO order, *tämä* prefers the postverbal object in both experiments, and with OVS order in Experiment 1, *tämä* is split between subject and object, but in Experiment 1b, with OVS order, *tämä* prefers the postverbal subject. Thus, in Experiment 1b, in both SVO and OVS conditions, *tämä* has a preference for the postverbal argument. I claim that the strengthening of the preference for a postverbal referent can be attributed to the presence of a preceding discourse context in Experiment 1b. In that experiment, the discourse context supports the discourse-statuses signalled by the word order (preverbal old, postverbal new). Now, let us combine this with the claim that discourse status affects salience (e.g. (Strube & Hahn, 1996, 1999)).[14] On the basis of the differences we see in the behaviour of *tämä* between Experiments 1 and 1b, and the claim that discourse status can influence salience, I hypothesize that the demonstrative *tämä* prefers referents that are low in salience. This, I would argue, is the reason why making the postverbal subject in the OVS condition more clearly discourse-new by means of the context (in Experiment 1b) makes it a better antecedent for *tämä*, and leads *tämä* to show a postverbal preference in both SVO and OVS conditions.

In sum, my hypothesis based on the data discussed here is that (1) *hän* is sensitive to grammatical role and prefers subjects, and that (2) *tämä* is sensitive to a more general notion of salience – and since salience depends on factors such as word order/discourse status (e.g. (Strube & Hahn, 1996, 1999)) and grammatical role (e.g. (Crawley & Stevenson, 1990)), *tämä* is sensitive to these factors. More specifically, according to this hypothesis, *tämä* prefers entities that are low in salience, entities that are not at the centre of attention at that point in the discourse (see also (Varteva, 1998)).

What does the disparate behaviour of *hän* and *tämä* mean for the notion of salience? These results indicate that we cannot capture the referential properties of *hän* and *tämä* by mapping them onto a unified salience scale, nor can we claim that one refers to more salient entities than the other. What, then, is the alternative? Are we going to end up defining a different kind of salience for each anaphoric form, and thereby lose a coherent concept of salience? We will return to this question in the discussion part of Section 6. Now, in Section 5, we turn to the third experiment, which looks at the role of referential form.

---

[14] *Tämä* can also refer to discourse-old referents, as is shown by corpus data. If it is preceded by a transitive sentence that contains two discourse-old arguments, which in Finnish will normally occur in S-O order, it prefers the object.

5 Experiment 2

This experiment addresses the question whether overt encoding of contextual oldness by means of a pronoun impacts the referential properties of *hän* and *tämä*. Beaver (to appear) and Kameyama (1999) discuss the possibility of pronominalization increasing a referent's salience in the subsequent discourse, and Experiment 2 tests this possibility, which has not received as much attention in the literature as the effects of word order and grammatical role. In this experiment, each item consisted of two sentences and the first word of the third sentence (*hän / tämä*). A sample item is given in (9). The first sentence introduces a referent, which is referred to with a pronoun at the beginning of the second sentence. The second sentence has SVO or OVS word order and thus the sentence-initial pronoun is either a subject pronoun (*hän* s/he-NOM') or an object pronoun (e.g. *häntä* 's/he-PART'). The second sentence has a new referent as its last word. The first word of the third sentence, *hän* or *tämä*, is in nominative case, i.e. in subject position.

(9) [SVO.Hän]

Puiston	*penkillä*	*lepäsi*	*lääkäri.*	*Hän*	*onnitteli*
Park's	bench-on	rested	doctor-NOM.	S/he-NOM	congratulated
opiskelijaa.	*Hän…*				
student-PART.	S/he-NOM…				

"On the park bench rested a doctor. S/he was congratulating a student. S/he…"

Again, there were four conditions: [SVO.Hän...], [OVS.Hän...], [SVO.Tämä...], and [OVS.Tämä...]. Thirty-two native Finnish speakers participated in this experiment, and each participant wrote continuations for 38 items (8 critical items and 30 fillers) in random order. As before, continuations were coded according to which of the referents in the preceding sentence the participants chose as the referent of the anaphoric expression. The item was coded as 'unclear/other' if it was not clear from the continuation which referent the participant had interpreted as being the antecedent of the pronoun or demonstrative, or if the pronoun or demonstrative was not used anaphorically to refer to the preceding subject or object.

5.1 Results

The results of the continuations reveal, again, that *tämä* and *hän* are affected differently by word order (see Figure 1). There are significant effects of anaphor type (*hän* vs. *tämä*) and word order (SVO vs. OVS) on reference to both subjects and objects, as well as significant interactions between anaphor type and word order (p's<.05). In other words, (i) the type of anaphoric expression (*hän* vs. *tämä*) and (ii) word order (SVO vs. OVS) have a significant effect on whether a particular anaphoric element is interpreted as referring to the preceding subject or object.

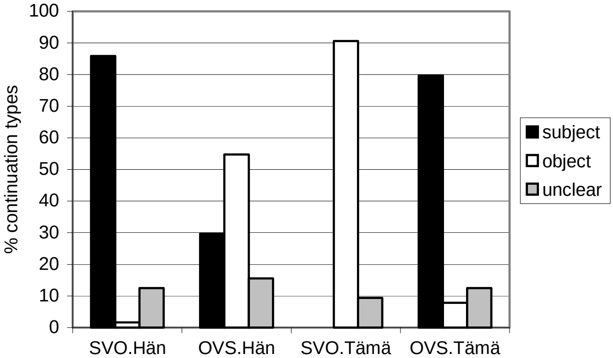


Figure 1. What does the anaphor refer to?

Let us now look at the different conditions in more detail. In the SVO.Tämä condition, we see 58/64 (91%) object-interpretations, and in the OVS.Tämä condition, there are 51/64 (80%) subject-interpretations. In other words, *tämä* has a strong tendency to be interpreted as referring to the non-pronominalized, postverbal referent. These findings are compatible with the hypothesis I formulated based on Experiments 1 and 1b, namely that *tämä* prefers referents that are low in salience.

In contrast, the referential properties of *hän* 's/he' in Experiment 2 do not match what we saw earlier. Now, word order – which, crucially, is correlated here with pronominalization – has an impact. In SVO order, the most likely antecedent for *hän* is the preceding pronominal subject, i.e. the entity that has already been mentioned (SVO.Hän = 86% subject-interpretations). In OVS order, participants' interpretations of the antecedent of *hän* are split between the full NP subject and the pronominalized object. There are 30% subject-interpretations, and 55% object-interpretations. The preference for the pronominalized object is significant ($p<.01$), which is in striking contrast to the earlier findings with two full-NP arguments, where *hän* clearly preferred the subject even in OVS order.

5.2 Discussion

Let us turn to the implications of these results. For *tämä*, the crucial factor in Experiment 2 is word order, i.e. referring to the postverbal referent. We can attribute the strengthening of the word order factor to a preference for *tämä* to refer to clearly discourse-new referents over discourse-old, pronominalized, salient referents. These results thus provide further support for the idea that *tämä* refers to entities that are low in salience.

The results for *hän* are more complex. In Experiments 1 and 1b, with two full-NP arguments in the SVO and OVS sentences, we observed that *hän* referred to the subject regardless of word order. However, in Experiment 2, what we see is in fact a preference for *hän* to refer to the pronominalized antecedent in preceding sentence – which is not always the subject. In other words, in Experiment 2 in both OVS.Hän and SVO.Hän conditions, the most likely referent of the pronoun is the preverbal, pronominalized argument. However, as Figure 1 shows, there is clearly a difference between OVS.Hän and SVO.Hän: the preference for the preverbal, pronominalized referent is greater in the SVO.Hän condition than in the OVS.Hän condition. This shows that *hän* still has some of the sensitivity to grammatical role (subjecthood) that we saw in the earlier experiments.

On the whole, then, the fact that the results of Experiment 2 differ from those of Experiments 1 and 1b reveals that other factors, beyond grammatical role, play a role in influencing what *hän* refers to – in particular, a pronoun is likely to refer to a pronominalized, preverbal referent, even if this referent is not the subject. This lends support to the claims of Beaver (to appear) and Kameyama (1999) that the referential forms of potential antecedents influence the referential properties of pronouns.

These results complicate the picture we had sketched earlier regarding the subject preference of the pronoun *hän*, and they raise a number of interesting questions. For example, what happens if the pronominalized entity is not sentence-initial? In Experiment 2, pronominalization was correlated with sentence-initial position, so even though the results tell us that referential form (full NP vs. pronoun) has an effect on *hän*, we cannot tell whether it is pronominalization alone or pronominalization combined with initial position that is responsible for the patterns we see. To address these concerns, in another version of this experiment, I disentangled sentence position and pronominalization by using only SVO sentences and varying whether the pronoun occurs in subject or object position (Kaiser, 2003: Ch. 3). The results show that sentence-initial pronouns do indeed pattern differently from non-sentence-initial pronouns, and that the unexpected preference for the pronominalized preverbal object that *hän* displayed in the OVS condition of Experiment 2 is not present when we change the order to SVO (with a full NP subject and a pronominalized object). In fact, with SVO order and a pronominalized object, *hän* is split between the subject and the object.

Comparing these results and the results of Experiments 1 and 1b, it looks like a chain of pronouns (as in Experiment 2, see ex. 9) patterns differently from a single use (Experiments 1, 1b). More specifically, I would like to suggest that the different findings can be reconciled if we make a distinction between the first occurrence of *hän* (i.e. when *hän* is used for a referent for the first time) and a second occurrence (i.e. when a referent picked out with *hän* is referred to with *hän* again). In (Kaiser, 2003), I show how the difference between first-occurrence and second-occurrences uses, as well as the differences between the two versions of Experiment 2, can be modelled using a referent-tracking system in which a second use of a pronoun can be interpreted as being anaphoric on the first use, and not directly on the referent itself. (See (Kaiser, 2003) for detailed analysis.)

6 Conclusions

In light of the data discussed in this paper, we can conclude that *hän* and *tämä* are not mirror images of one another. This suggests that we shouldn't aim to define their referential properties in terms of a single unified notion of salience. Instead, the results show that we need to explore the degree to which different factors – such as word order, grammatical role, and the form of the antecedent – are relevant for different referential expressions. Let us now return to the question raised at the end of section 4, namely, what does the disparate behaviour of *hän* and *tämä* mean for the notion of salience? If we cannot capture the referential properties of *hän* and *tämä* by mapping them onto a unified salience scale, what is the alternative? Do we define the referential properties of each form separately, and thereby lose the coherent grouping structure provided by accessibility-hierarchy type theories?

A possible alternative way of coherently grouping the referential properties of the pronoun *hän* and the demonstrative *tämä* is to hypothesize that these two forms differ in the level of linguistic representation that they access, or 'look at', in order to locate their antecedents. The idea would be that the demonstrative *tämä* accesses the discourse level, and is associated with the low-end of a salience scale, and first-occurrence uses of the pronoun *hän* accesses the syntactic level, and are associated with the high-end of a grammatical role scale. However, for this hypothesis to work, we crucially need to distinguish pronominal chains from first-occurrence uses of *hän*, as discussed above. In addition, in order to assess the validity of this approach, we need to investigate other constructions in Finnish. For instance, the grammatical-subject preference of first-occurrence uses of *hän* could be tested by looking at how it behaves in contexts in which the previous sentence contains an experiencer/psych verb or a passive construction. In the experiments reported here, the subject of the preceding sentence was also the agent. Thus, based on these experiments, we cannot tell whether the pronoun *hän* is sensitive to subjecthood or to agentivity. I plan to investigate these important questions in future work by looking at different constructions and verb types.

It would be interesting to investigate the crosslinguistic implications of my hypothesis that the 'antecedent retrieval instructions' carried by different forms are different because different forms tap into different levels of representation. Preliminary data from Estonian pronouns and demonstratives seems to resemble the Finnish data, in that the third person pronoun *ta* and the demonstratives *see/too* are sensitive to different kinds of information (Kaiser & Hiietam 2003). Clearly, in order to test the validity of the hypothesis, in-depth research on a range of languages will be necessary.

In future work, it would also be interesting to look at dialects of spoken Finnish, whose referential systems often differ from that of standard Finnish. In the most common spoken dialect, used primarily in the urban areas of southern Finland, the non-human pronoun *se* 'it' is used for human referents, but the pronoun *hän* 's/he' and the demonstrative *tämä* 'this' are also used (see e.g. (Seppänen, 1998)). This situation raises interesting questions, in particular concerning the division of labour of *se* and *hän*. This is a question that clearly merits further work.

On the whole, the results presented in this paper have interesting implications for our understanding of how referential systems work. One possible option is that the system 'assigns jobs' to the elements, such that the functions of one element are fully dependent on the functions of other elements present in the paradigm. This seems to be assumed by accessibility hierarchy-type approaches which suggest that null pronouns are used for more accessible referents than pronouns, which in turn are used for more accessible referents than demonstratives, and so on. Another option is that the different elements can also have properties of their own, independent of the system. The results discussed here seem to favour the second option, as they reveal the differences in the factors to which *hän* and *tämä* are sensitive.

References

Ariel, M. 1990. *Accessing Noun Phrase Antecedents*. London: Routledge, Croom Helm.

Arnold, J. 1998. *Reference Form and Discourse Patterns.* Ph.D. thesis. Stanford University.

Beaver, D. To appear. "The Optimization of Discourse Anaphora." *Linguistics and Philosophy.*

Brennan, S., M. Friedman and C. Pollard 1987. "A Centering Approach to Pronouns." *Proc. of the 25th Annual Meeting of the ACL*. Stanford, CA. 155-162.

Chafe, W. 1976. "Givenness, Contrastiveness, Definiteness, Subjects, and Topics." C. Li (ed.) *Subject and Topic*. Academic Press. 27-55.

Chambers, G. and R. Smyth 1998. "Structural Parallelism and Discourse Coherence: A Test of Centering Theory." *Journal of Memory and Language* 39.593-608.

Chesterman, A. 1991. *On definiteness.* Cambridge: Cambridge University Press.

Choi, H.-W. 1996. *Optimizing Structure in Context: Scrambling and Information Structure.* Ph.D. thesis. Stanford University.

Crawley, R. and R. Stevenson 1990. "Reference in Single Sentences and in Texts." *Journal of Psycholinguistic Research* 19(3).191-210.

Crichton, M. 1995. *The Lost World.* New York: Knoepf. [Finnish translation: Crichton, M. 1996. *Kadonnut Maailma.* Helsinki: Otava. Finnish translation by Jaakko Kankaanpää].

Etelämäki, M. 1996. *Keskutelu tarkoitteesta kamerataiteen oppitunnilla – ja pronominit tuo, se, tämä.* Thesis, Department of Finnish, University of Helsinki.

Gernsbacher, M. 1990. *Language Comprehension as Structure Building.* Hillsdale, NJ: Lawrence Erlbaum.

Givón, T. 1984. *Syntax: A Functional-typological Introduction*. Amsterdam: John Benjamins.

Gordon, P., B. Grosz and L. Gilliom 1993. "Pronouns, Names, and the Centering of Attention in Discourse." *Cognitive Science* 17.311-347.

Gundel, J., N. Hedberg and R. Zacharski 1993. "Cognitive Status and the Form of Referring Expressions in Discourse." *Language* 69.274-307.

Halmari, H. 1994. "On Accessibility and Coreference." *Nordic Journal of Linguistics* 17.35-59.

Helasvuo, M.-L. 2001. *Syntax in the Making – The Emergence of Syntactic Units in Finnish Conversation*. Amsterdam: John Benjamins.

Hiirikoski, J. 1995. "Correlations Between some Morphosyntactic Features and Word Order in Finnish and English: Some Preliminary Results of Testing the Transitivity Hypothesis." B. Wårwik, S. Tanskanen and R. Hiltunen (eds.) *Organization in Discourse, Proc. from the Turku Conference (Anglicana Turkuensia 14).* 289-299.

Hoffman, B. 1998. "Word Order, Information Structure and Centering in Turkish." M. Walker, A. Joshi and E. Prince (eds.) *Centering Theory in Discourse*. Oxford: Oxford University Press. 251-272.

Kaiser, E. 2000. "Pronouns and Demonstratives in Finnish: Indicators of Referent Salience." P. Baker, A. Hardie, T. McEnery and A. Siewierska (eds.) *Proc. of the Discourse Anaphora and Reference Resolution Conference (DAARC 2000).* Lancaster, University Center for Computer Corpus Research on Language (UCREL) technical papers volume, 12:20-27. Lancaster: University Center for Computer Corpus Research on Language (UCREL).

---------- 2003. *The Quest for a Referent: A Crosslinguistic Look at Reference Resolution.* Ph.D. dissertation, University of Pennsylvania.

---------- In Press. When salience isn't enough: Pronouns, demonstratives and the quest for an antecedent. R. Laury (ed.) *Minimal reference in Finnic: The use and interpretation of pronouns and zero in Finnish and Estonian discourse*. Helsinki, Finland: Suomalaisen Kirjallisuuden Seura.

---------- and K. Hiietam. 2003. "A Typological Comparison of Third Person Pronouns in Finnish and Estonian." Anne Dahl and Peter Svenonius (eds.) *Nordlyd* 31(4).654-667, *Proc. of the Workshop on Generative Approaches to Finnic Languages.* University of Tromsoe.

Kalliokoski, J. 1991. "Emphathy as Motivation for Style Shifting in Narrative." J. Verschueren (ed.) *Levels of Linguistic Adaptation. Selected papers of the Int. Pragmatics Conference.* Antwerp, August 17-22, 1987. Vol II:147-161. Amsterdam: John Benjamins.

Kameyama, M. 1999. "Stressed and Unstressed Pronouns: Complementary Preferences." P. Bosch and R. van der Sandt (eds.) *Focus: Linguistic, Cognitive and Computational Perspectives*. Cambridge: Cambridge University Press. 306-321.

Laitinen, L. 1992. *Välttämättömyys ja persoona. Suomen murteiden nesessiivisten rakenteiden semantiikkaa ja kielioppia*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Laury, R. 1997. *Demonstratives in Interaction.* Amsterdam: John Benjamins.
---------- (ed.) To appear. *Minimal reference in Finnic: The Use and Interpretation of Pronouns and Zero in Finnish and Estonian Discourse.* Helsinki: Suomalaisen Kirjallisuuden Seura.
Lenerz, J. 1977. *Zur Abfolge nominaler Satzglieder im Deutschen.* Tübingen: Gunter Narr.
Linna, V. 1954/1999. *Tuntematon Sotilas*, Helsinki: WSOY.
Matthews, A. and M. Chodorow 1988. "Pronoun Resolution in Two-clause Sentences: Effects of Ambiguity, Antecedent Location, and Depth of Embedding." *Journal of Memory and Language* 27.245-260.
Prasad, R. and M. Strube 2000. "Discourse Salience and Pronoun Resolution in Hindi." A. Williams and E. Kaiser (eds.) *University of Pennsylvania Working Papers in Linguistics* 6(3).189-208. Philadelphia: University of Pennsylvania.
Prince, E. 1992. "The ZPG Letter: Subjects, Definiteness and Information Status." S. Thompson and W. Mann (eds.) *Discourse Description: Diverse Analyses of a Fund Raising Text*. 295-325. Amsterdam: John Benjamins.
Rambow, O. 1993. "Pragmatic Aspects of Scrambling and Topicalization in German." Paper presented at the Workshop on Naturally-Occurring Discourse, Institute for Research in Cognitive Science, University of Pennsylvania, Philadelphia.
Saarimaa, E. 1949. "Kielemme käytäntö. Pronominivirheistä." *Virittäjä* 49.250-257.
Seppänen, E.-L. 1998. *Läsnäolon pronomit.* Helsinki: Suomalaisen Kirjallisuuden Seura.
Sheldon, A. 1974. "The Role of Parallel Function in the Acquisition of Relative Clauses in English." *Journal of Verbal Learning and Verbal Behavior* 13.272-281.
Stevenson, R., R. Crawley and D. Kleinman 1994. "Thematic Roles, Focus and the Representation of Events." *Language and Cognitive Processes* 9.519-548.
Strube, M. 1998. "Never Look Back: An Alternative to Centering." *Proc. of Coling-ACL '98.* 1251-1257.
---------- and U. Hahn 1996. "Functional Centering." *Proc. of ACL '96*. 270-277.
---------- and ---------- 1999. "Functional Centering: Grounding Referential Coherence in Information Structure." *Computational Linguistics* 25(3).309-344.
Sulkala, H. and M. Karjalainen 1992. *Finnish*. London: Routledge.
Tetreault, J. 2001. "A Corpus-Based Evaluation of Centering and Pronoun Resolution." *Computational Linguistics* 27(4).507-520.
Turan, Ü. 1998. "Ranking Forward-Looking Centers in Turkish." M. Walker, A. Joshi and E. Prince (eds.) *Centering Theory in Discourse*. Oxford: Oxford University Press. 139-160.
Varteva, A. 1998. Pronominit hän ja tämä tekstissä. *Virittäjä* 2/1998:202-223.
Vilkuna, M. 1989. *Free Word Order in Finnish: Its Syntax and Discourse Functions.* Helsinki: Suomalaisen Kirjallisuuden Seura.
---------- 1995. "Discourse Configurationality in Finnish." Katalin Kiss (ed.) *Discourse Configurational Languages*. New York: Oxford University Press.244-268.
Vilppula, M. 1989. "Havaintoja hän- ja he-pronominien käytöstä suomen murteissa." *Virittäjä* 93.389-39.

Referential Accessibility and Anaphor Resolution: The Case of the French Hybrid Demonstrative Pronoun *Celui-Ci/Celle-Ci*[1]

Marion Fossard[1] and François Rigalleau[2]
[1]*Laboratory of Experimental Psychology, University of Sussex*
[2]*Laboratoire LACO, Université de Poitiers*

Most psycholinguistic studies dealing with the concomitant effects of discourse structure (focusing certain entities more than others) and different types of referential expressions on sentence processing in discourse contexts mainly aim to characterize the referential opposition between anaphoric pronouns and full repeated NPs (Garrod *et al.*, 1994; Gordon *et al.*, 1993, 1995). The aim of our research is to study – in French – another type of referential contrast, that exists between the anaphoric pronoun and the 'hybrid' demonstrative pronoun *celui-ci/celle-ci*. In two experiments using reading time measurements, we tested the contrast between these pronouns. The results suggest that both these pronoun types are sensitive to entity focusing, but in opposite ways: indeed, whereas an anaphoric pronoun is expected to signal referential and attentional continuity, the demonstrative pronoun would rather indicate a shift in attention focus (Gundel, 1998). The results also indicate that it is necessary to postulate a distinction between these two linguistic forms in terms of the processing instructions that they carry. We suggest that the presuppositional constraints could be stronger for certain forms than for others.

1 Introduction

As a linguistic procedure "designed to bring into conformity speaker's and addressee's model of the current discourse by maintaining the saliency level of some discourse referent already presented within it" (Cornish, 1999:5), anaphora is a very important phenomenon for the psychology of language, in general, and for studies dealing with the nature and role of mental representations used during discourse comprehension, in particular (Garnham, 1997).

Indeed, because it is "more a type of use than a category" (Charolles, 1991), anaphora may be realized via a whole range of referential expressions,[2] from zero forms or unaccented pronouns (typically, a 3rd person pronoun), to demonstrative pronouns, demonstrative descriptions, and definite descriptions.

---

[1] This paper is a revised and extended version of a paper entitled "Cognitive aspects of pronominal anaphora: the case of the French hybrid demonstrative pronoun *celui-ci/celle-ci'*", which was presented at the DAARC 2002 (Estoril, Portugal, 18-20 September 2002). We would like to thank Francis Cornish and Harriet Dunbar for their helpful comments on the revised version.

[2] In fact, it is not only the referential expression used which realizes anaphora, but also the whole clause in which it occurs (Cornish, 1999; Kleiber, 1994).

These expression types, however, are not equivalent in terms of the procedural instructions they carry, and a better understanding of the way in which each of these forms selects its referent should contribute to improving our knowledge concerning the cognitive processes involved in the resolution of referential expressions.

For example, how do the functional specificities inherent in each type of expression guide sentence processing in different ways? Or how does the choice of one expression (rather than another) affect the way in which the addressee builds his/her own discourse model, as well as the way in which s/he distributes his/her attention within this model?

A number of linguists supporting their work on cognitive accessibility,[3] propose that the different types of referential expressions, via their specific meaning which consists in marking what the cognitive status of the intended referent is in the speaker's and addressee's mental discourse model, would signal different ways in which a sentence may – or should – be resolved (Ariel, 1990, 1996; Chafe, 1994; Givon, 1983; Gundel *et al.*, 1993; Gundel, 1998).

Models such as Ariel's *Accessibility Marking Hierarchy* (1990, 1996) or Gundel *et al.*'s *Givenness Hierarchy* (1993) claim, indeed, that the use of any particular expression, far from being a random choice, is closely connected to the level of accessibility or activation that the mental representation of the referent is assumed to have in the addressee's mental model of the discourse under construction. On the basis of the various cognitive statuses which are assigned to discourse referents, these models provide a formal basis that makes it possible to justify and even to predict the presence or use of a particular marker.

Henceforth, when the referent targeted is assumed to be highly accessible, highly active in the addressee's mental discourse representation, a morphologically and phonologically attenuated form such as a zero and unaccented third-person pronoun should be used, a form whose use signals precisely the '*in focus'* status of the referent (in Gundel *et al.*'s terminology). On the other hand, where the referent is not in focus, but is '*activated'* (in Gundel *et al.*'s terminology) or enjoys a level of '*medium accessibility*' (in Ariel's terminology), a demonstrative pronoun or accented third-person pronoun, whose use codes this cognitive status, will tend to be used. Finally, lexically and accentually more substantial expressions (such as definite full NPs

[3] That is on the notion whereby the form of referential expressions used to refer to entities in a discourse depends, to a great extent, on the way in which the mental representation of these entities has been previously established.

or proper nouns) are expected to indicate referents bearing a low level of accessibility.

Following Givenness Hierarchy of Gundel *et al.* (1993), six cognitive statuses are then recognized and arranged from most restrictive (currently *'in focus'*) to least restrictive (only '*type identifiable'*). These statuses are claimed to represent the conventional meanings of the different determiners and pronominal forms placed at each point on the hierarchy (see Table 1).

In focus	>	Activated	>	Familiar	>	Uniquely identifiable	>	Referential	>	Type identifiable
{it}		*{that, this, this N}*		*{that N}*		*{the N}*		*{indefinite this N}*		*{a N}*

Table 1: Gundel et al.'s Givenness Hierarchy (1993)

Therefore, in order to propose valuable distributional patterns of anaphoric expressions, considerable effort has been devoted to characterizing – from a linguistic point of view – the relations between anaphor informativeness and referent accessibility. This work, however, is very much in contrast with the study of processing in the psycholinguistic literature given that, up to now, most psycholinguistic studies dealing with the interaction between the accessibility of discourse referents and referential expressions has essentially been interested in the differences between *only* two types of expressions: unaccented 3rd person pronouns and full NPs such as definite descriptions or repeated proper names (Garrod *et al.*, 1994; Gordon *et al.*, 1993; Gordon & Scearce, 1995; Hudson-D'Zmura & Tanenhaus, 1998; Sanford *et al.*, 1988).

Psycholinguistic models of anaphor processing also assign a crucial importance to the accessibility factor. However, the six levels defined by Gundel *et al.* (1993) are usually not distinguished by these models, maybe because experiments contrast only two anaphoric devices (but see (Garrod, 1994), for an attempt in this sense).

For instance, Greene *et al.* (1992) only suggest that entities vary in accessibility, and that highly accessible entities are placed in the focus of attention. Those entities are normally referred by a pronoun. Therefore, "highly accessible" and "in focus" are synonyms under this approach. The relatively loose usage of the Focus notion in psycholinguistic models is related to the heterogeneity of the factors that can affect the accessibility of an entity during discourse comprehension. Some researchers have centred their attention on the effect of syntactic factors on the accessibility ("structural focusing", (Grosz *et al.*, 1995)), whereas others have argued for focusing based on background knowledge of the topic of discourse (Sanford & Garrod, 1981). Yet, others have proposed that many different factors interact to determine what is in focus

(McKoon *et al.*, 1993; Marslen-Wilson *et al.*, 1993). Despite this variability, all these authors agree that, in a well-constructed discourse, the predicate of a pronoun can often be attached to the most accessible entity. According to McKoon *et al.* (1993), this process is sufficient to resolve most of the anaphoric pronouns, reducing the need to compute difficult inferences to establish the correct referent.

There is growing psychological evidence that anaphoric pronouns (like *he/she*) are very sensitive to entity focusing. They appear to act, indeed, as 'pointers' to the discourse focus (i.e. the highly-focused entity). Experimental studies indicate: (1) that sentences containing a pronoun which refers back to the highly-focused entity are easier to process and are read faster than sentences containing a pronoun which refers back to the less-focused entity (Garrod *et al.*, 1994; Gordon & Scearce, 1995; Sanford *et al.*, 1988); and (2) that the use of a repeated name for referring back to the highly-focused entity increases the reading time compared with the use of a pronoun in this same condition – *the repeated-name penalty effect* – (Gordon *et al.*, 1993; Gordon & Scearce 1995; Hudson-D'Zmura & Tanenhaus, 1998). However, with more explicit anaphors (such as repeated names), such a difference of processing (between differently focused entities) was not found (see in particular, (Fossard, 1999; Garrod *et al.*, 1994; Sanford *et al.*, 1988)), which suggests that the focus status of discourse entities has little influence on the processing of repeated full NPs. Indeed, contrary to an anaphoric pronoun – whose interpretation is not independent of the immediate context of use –, a repeated noun is better able to identify its referent in a descriptive fashion (Garrod, 1994). Consequently, the interpretation of repeated full NPs would be less bound by entity focusing.

Now, what about demonstrative pronouns? Turning to Gundel *et al.*'s work (1993), this expression type for which the accessibility level of the intended referent is 'midway' between the *'in focus'* status (typically coded by *'it'*) and the *'uniquely identifiable'* status (typically coded by *'the N'*),[4] could be sensitive to entity focusing, but in a quite original fashion. Indeed, as suggested by Gundel (1998:186): "Demonstratives (especially demonstrative pronouns) and stressed personal pronouns typically imply that the referent is not in focus, i.e. they imply *a focus shift*".

Also – and given this fact –, we wondered about the constraints likely to affect the interpretation of one particular demonstrative pronoun: the French 'hybrid' demonstrative pronoun *celui-ci/celle-ci*.[5] This original indexical

[4] See Table 1, above.

[5] *The latter* or *this one* in English. However, unlike English which does not possess the category of grammatical gender, the French demonstrative pronoun *celui-ci/celle-ci* (plural: masc. *ceux-ci*; fem. *celles-ci*) is a gender- and number-variable pronoun which presupposes that its intended referent is a

expression, which in Kleiber's phrase would consist in "showing something new in terms of something already known" (1994:177), could be particularly well-adapted to triggering a focus shift.

Following Kleiber's (1994) analysis, this pronoun is indeed atypical, hybrid, because it combines both anaphoric and deictic elements. It is an anaphoric expression by virtue of its pronominal element *lui* which is responsible for the anaphoric aspect of the interpretation of the whole expression, and it is also a deictic expression by virtue of the combination of *ce* and the demonstrative particle modifier *–ci* which is responsible for its deictic value.

So, in the case of anaphoric reference, as illustrated in the example below (1), where the referent has already been introduced into the discourse (that is, when the referent is not 'new' but already 'known', i.e. *Melle Vatnaz*), the use of this pronoun, through its deictic value, would have the effect of isolating, and hence highlighting, the referent in question from within its background context.

(1) L'ouvrière redit naïvement son mensonge à Melle Vatnaz; celle-ci en vint à parler au brave commis. (Flaubert; Kleiber 1994:186)
"The female worker naïvely repeated her lie to Melle Vatnaz; the latter (f.sg) ended up speaking to the good clerk (about it)."

According to Kleiber (1994) and Cornish (1999), this expression type would signal a change in the attention focus already established, that is, it would shift the attention focus from the highly-focused entity towards a less-focused entity in order to bring the latter into the foreground. In brief, the demonstrative pronoun *celui-ci*, also called an 'anaphoric-deictic pronoun', could be specially adapted (more so than a repeated noun) to "drawing the addressee's attention to the member of the set of salient entities already evoked which enjoys the *lower* level of focus at the point of occurrence, i.e. Melle Vatnaz" (Cornish, 1999:67). As Kleiber and Cornish pointed out, putting the ordinary third-person pronoun *elle* ('*she*') in the place of the demonstrative pronoun *celle-ci* – in example (1) above – would have led to a completely different referent (i.e. "the female worker", which constitutes the main protagonist, the most salient referent).

Also, it appears that an additional constraint intervenes in the functioning of the pronoun *celui-ci*. As it is an anaphoric tool for a reference in favour of a background entity, at least two entities must be present. This constraint, of

---

discrete entity, having already been categorized via the use of a noun whose head provides the gender value of the pronoun (Kleiber, 1994; Cornish, 1999). When *celui-ci* is used as a pure demonstrative pronoun (i.e. without textual antecedent), it must agree with the gender of the noun corresponding to the entity. For instance, when somebody shows a table (feminine gender in French), this person must use *celle-ci* to be understood. If *celui-ci* is used instead, the pronoun is not correct. This requirement suggests that people quickly activate the gender associated to an entity that is not explicitly mentioned.

course, is not relevant for the ordinary pronoun which can be used if only one entity is mentioned.

Consequently, we think that studying the more subtle contrast which exists between the 3rd person anaphoric pronoun and the 'anaphoric-deictic' pronoun *celui-ci/celle-ci* could be worthwhile. Indeed, as a distinct object of study, *celui-ci* could enable us to obtain further data on the influence of entity focusing on the processing of referential expressions.

We therefore make the assumption that the processing of the demonstrative pronoun *celui-ci/celle-ci* should be sensitive to the focus status of discourse entities (i.e. its referential behaviour should be bound by entity focusing), but in an opposite way to that of the 3rd person anaphoric pronoun. In other words, whereas the 3rd person anaphoric pronoun, insofar as it indicates that attention is to be maintained on the highly-focused entity, is a preferred marker in accessing discourse focus, the demonstrative pronoun *celui-ci/celle-ci*, since it serves to re-orient attention to a less-prominent discourse referent, could be specialized in the foregrounding of an entity which did not have this status previously.

To test the contrast between both these pronouns, we used a self-paced reading time paradigm in two experiments, with either agreement in gender with only one name (Experiment 1: unambiguous gender cue) or agreement in gender with two names (Experiment 2: ambiguous gender cue).

2 Experiment 1

Experiment 1 used a self-paced reading time task to test the intended contrast between the 3rd person anaphoric pronoun (*il/elle–'he/she'*) and the demonstrative pronoun (*celui-ci/celle-ci*) in accessing differently focused discourse entities. In accordance with our hypotheses, our aim was to highlight:

1) a specialization of the 3rd person anaphoric pronoun in referring back to the highly-focused entity (i.e. the target sentence containing this pronoun should be easier to process and therefore, should be read faster when it refers to the highly-focused entity than when it refers to the less-focused entity), and conversely;

2) a specialization of the demonstrative pronoun to refer back to the less-focused entity (i.e. the target sentence containing this pronoun should be easier to process and, therefore, read faster when it refers to the less-focused entity than when it refers to the highly-focused entity). In this experiment, gender cue was unambiguous.

2.1 Method

Subjects

A total of forty subjects (mean age 22 ± 2.1 years) participated in Experiment 1. They were all native speakers of French, and were not paid for their participation.

Materials

Pronoun Type		Entity Type	
		Entity 1	Entity 2
3rd person anaphoric pronoun		Le panier de linge était rempli de vêtements. *(The linen basket was full of clothes.)* Salomé(i) écoutait la radio en (i)repassant les chemises d'Hervé./ *(Judy was listening to the radio while ironing John's shirts) /*	Le panier de linge était rempli de vêtements. *(The linen basket was full of clothes.)* Hervé(i) écoutait la radio en (i)repassant les chemises de Salomé./ *(John was listening to the radio while ironing Judy's shirts) /*
	TS:	De fatigue, elle s'allongea sur le lit./ *(Tired, she lay down on the bed) /*	De fatigue, elle s'allongea sur le lit./ *(Tired, she lay down on the bed) /*
demonstrative pronoun		Le panier de linge était rempli de vêtements. *(The linen basket was full of clothes)* Salomé(i) écoutait la radio en (i)repassant les chemises d'Hervé./ *(Judy was listening to the radio while ironing John's shirts) /*	Le panier de linge était rempli de vêtements. *(The linen basket was full of clothes)* *Hervé(i) écoutait la radio en (i)repassant les chemises de Salomé./ *(John was listening to the radio while ironing Judy's shirts) /*
	TS:	De fatigue, celle-ci s'allongea sur le lit./ *(Tired, this one (FEM.SG) lay down on the bed) /*	De fatigue, celle-ci s'allongea sur le lit./ *(Tired, this one (FEM.SG) lay down on the bed) /*

Note. The referent-entity (entity 1 or entity 2) is underlined for expository purposes. In the text presented to participants, nothing was underlined. In the same way, the co-indexation sign (i) did not appear. The slash indicates the text presentation on the screen. The target sentence (TS) was presented separately on the screen. The English translation of this example is indicated in brackets in italics. **Hervé* is a man's name. For clarity, the translated names are different but have same gender as the French names.

Table 2: Example of an experimental text in all four conditions in Experiment 1

For the experimental texts, a total of 40 three-sentence texts was constructed on the following pattern: the first sentence was an introductory, scene-setting sentence. The second introduced two discourse entities (two characters), one of whom was the highly-focused entity (entity 1), and the other was the less-focused entity (entity 2). The third sentence, the target sentence, referred to one of the two characters mentioned in the second sentence either via the 3rd person anaphoric pronoun (*il/elle*), or via the demonstrative pronoun (*celui-ci/celle-ci*). Each text ended with a 'true-false' statement which probed comprehension of the target sentence.

A total of four experimental conditions (2 entity types × 2 pronoun types) were then created for each experimental text (see Table 2).

For each text, the differential level of focusing between the two referent-entities mentioned in the second sentence was obtained as follows:

- the highly-focused entity (entity 1: the first character) was always introduced as the most topical argument of the utterance: it occurred as initial mention and in grammatical subject position in the sentence, being, therefore, the "main protagonist" of the situation described (Gordon *et al.*, 1993; Garrod *et al.*, 1994).
- On the other hand, the expression designating the less-focused entity (entity 2: the second character) was embedded within a participial clause, always occurring in object position of a preposition. This second entity, more deeply embedded in the sentence structure, only plays a very peripheral, minor role within the situation described. Moreover, to increase the topicality of the first entity and therefore to increase the focus differential between the two entities, entity 1 was re-evoked by means of a zero form (PRO in generative analyses) as grammatical subject of the participial clause.

In this first experiment, the pronominal reference was unambiguous, the pronoun agreed in gender with only one of the two character-entities. On the other hand, the semantic information carried by the predicate of the target sentence was relatively compatible on a pragmatic level with both potential referents. This information was 'neutral' in relation to the two characters; hence, readers could not use such pragmatic inferences like 'argument-predicate' to resolve a given pronoun. The mean length of the target sentences for all the experimental texts was 7.6 words.

In addition to the forty experimental texts, there were forty-eight filler texts. Half of them were similar to the experimental texts, the other half were different, but the 'true-false' statements did not test the interpretation of pronouns for the fillers. These statements were designed to encourage comprehension of either the first or the second sentence.

Design and Procedure

Four lists of materials were constructed to ensure that each experimental text occurred in each of the four conditions. The subjects were assigned randomly to lists with the restriction that each list was assigned to an equal number of subjects. Overall, each subject saw 40 experimental texts, 10 in each condition.

Thus, each subject read each text only once, but each text appeared in the four conditions equally often across subjects. The order of presentation of the 88 texts (40 experimental texts + 48 filler texts) was individually randomised for each subject.

Subjects were tested individually. They were instructed to read the text at a normal rate and answer the statements as accurately and as rapidly as they could. The text appeared in the centre of the screen in two successive stages. The first two sentences were presented together. After reading the first two sentences, the subjects pressed the space bar, the two sentences disappeared and the third sentence – the target sentence – was displayed. At the end of the text, the target sentence disappeared when the statement was presented. For half the statements, the intended answer was 'true'; for the other half, it was 'false'. After answering the statements by pressing one of two keys marked 'true' and 'false', subjects were prompted to start the next trial.

Before the presentation of the experimental materials there were ten practice trials, whose primary purpose was to familiarize the subjects with the self-paced reading procedure. The task lasted about forty minutes.

2.2 Results and Discussion

The means of the reading times (RTs) for the target sentences were calculated for each subject and each item in each condition. RTs were analysed after eliminating outlier reading times (less than 500 msec or greater than 15 sec), affecting 0.3% of the data. Any data points that were more or less than 2 standard deviations from the mean for a particular subject were replaced with the 2 standard deviations' cut-off values (5.1% of the data). Then, RTs were normalized in order to take target sentence length (number of characters) into account.

Pronoun Type		Entity Type	
		Entity 1	Entity 2
3rd pers. Pron.	*Error %*	4.75%	5%
	Mean RT	40.6	45.8
	SD	(10.3) [5.3]	(13) [6,6]
Dem. Pron.	*Error %*	6.75%	6.25%
	Mean RT	46.1	44.6
	SD	(12.2) [6.8]	(11) [6.2]

Note. The reading times are given in milliseconds per character. The values in parentheses are the standard deviations with subjects as the random factor; the values in brackets are the standard deviations with items as the random factor.

Table 3: Means and standard deviations of the reading times of the target sentences and mean percent of errors for the statements in Experiment 1

Table 3 shows the means and the standard deviations of the reading times of the target sentences and the mean percent of errors for the statements in all conditions.

A 2 × 2 analysis of variance (ANOVA) of Entity-Type (entity 1 vs. entity 2) × Pronoun-Type (*il/elle* vs. *celui-ci/celle-ci*) was conducted separately for subjects as the random factor (F_1), and for items as the random factor (F_2).

Accuracy of answer for the statements

For all the statements, there was a 94.3% correct answer rate. The results revealed no main effect[6] of either entity-type, *F1, F2* < 1, or pronoun-type, *F1, F2* < 1. The interaction between these two factors was also not significant, *F1, F2* < 1, the different versions of a text did not influence the accuracy of the answer for the statement in a significant way.

Reading times of the target sentences

The results revealed a main effect of entity-type, $F_1(1,39)=12.7$, $p < .002$; $F_2(1,39)=7.2$, $p < .02$, reflecting the fact that processing was faster when the target sentence referred back to the highly-focused entity (entity 1) (*M*= 43.3 msec/char.) than when it referred back to the less-focused entity (entity 2) (*M*= 45.2 msec/char.).

There was also a main effect of pronoun-type, significant by subjects, $F_1(1,39)=14.1$, $p < .001$, but not by items, $F_2(1,39)=2.4$. Target sentences containing a 3rd person pronoun were read faster (*M*=43.2 msec/char.) than those with a demonstrative pronoun (*M*=45.2 msec/char.).

Most crucially, the interaction between entity-type and pronoun-type was significant, $F_1(1,39)=8.4$, $p < .02$; $F_2(1,39)=41.4$, $p < .001$. There was a strong effect of focus (i.e. the highly-focused entity) on the processing of the 3rd person pronoun: target sentences referring back to entity 1 were read faster with a 3rd person pronoun than with a demonstrative pronoun, $F_1(1,39)=29$, $p < .001$; $F_2(1,39)=13.5$, $p < .001$. As predicted, using a demonstrative pronoun to refer back to entity 1 is penalizing. There was also a specific referential functioning of the 3rd person pronoun in referring back to entity 1 rather than to entity 2, the difference being significant, $F_1(1,39)=20.1$, $p <.001$; $F_2(1,39)=37.6$, $p <.001$.

So, according to our hypotheses, the 3rd person anaphoric pronoun seems to act effectively as a ‘pointer’ to discourse focus.

---

6 The ‘Main effect’ of a factor concerns the global difference between the different levels of this factor (i.e. the difference between Entity 1 and Entity 2), without taking into account the other factor (i.e., Type of pronoun). For more precisions, see (Winer *et al.*, 1996).

On the other hand, the ‘opposite’ effect of focus (in favour of the less-focused entity) which was predicted for the processing of the demonstrative pronoun was not significant, F_1, F_2 < 1. In spite of a small reading time difference which goes in the expected direction (see Table 3), the target sentences referring back to entity 2 were not significantly easier to process with a demonstrative pronoun than with a 3rd person anaphoric pronoun. In the same way, the results did not indicate significant differences for the processing of the demonstrative pronoun referring either to entity 2, or to entity 1, $F_1(1,39)=1.2$, $p > .2$; $F_2(1,39)=2.8$, $p > .09$.

Except for the results concerning the processing of the 3rd person anaphoric pronoun, this first experiment failed to highlight a specialization of the demonstrative pronoun in referring to the less-focused entity. An advantage in reading times was noted in favour of the demonstrative pronoun for referring back to entity 2 when compared with entity 1 (respectively, M = 44.6 msec/char vs. M = 46.1 msec/char). Similarly, when Entity 2 was referred to, the demonstrative pronoun seemed advantageous when compared with the 3rd person anaphoric pronoun (respectively, M = 44.6 msec/char vs. M = 45.8 msec/char). However, the two comparisons were not significant.

Why then did subjects not take advantage (or not sufficiently) of the highly specialized ‘mode of givenness’ associated with the demonstrative pronoun *celui-ci* in retrieving the less-focused entity? The answer may be found in (Charolles, 1995:89) who notes that “*celui-ci* is one of a set of anaphoric forms specialized in the avoidance of risk of ambiguity whose purpose is to select a referent in terms of a contrast within a set of potential candidates”.

In the materials used in Experiment 1, any risk of ambiguity was minimal because the gender cue was unambiguous: the pronoun agreed in gender with only one of the two names mentioned. So, to re-test our hypotheses concerning the processing of the pronoun *celui-ci*, we constructed another experiment taking the notion of *avoidance of risk of ambiguity* into account. In this new experiment, gender cue was irrelevant (i.e., ambiguous).

3 Experiment 2

Experiment 1 failed to highlight a specialization of the pronoun *celui-ci* in referring to the less-focused entity. We think this failure could be due to the unambiguous gender cue. In order to remedy this, Experiment 2 was constructed with ambiguous gender cue (with agreement in gender with either of the two proper names mentioned). The major aim of Experiment 2 was to bring out the specialization of the demonstrative pronoun *celui-ci* in referring to the less-focused entity. Our hypotheses, then, are unchanged in relation to those of Experiment 1. In this second experiment, we also used a self-paced reading time task to test the predicted contrast between the 3rd person anaphoric pronoun (*il/elle–'he/she'*) and the demonstrative pronoun (*celui-ci/celle-ci*) in accessing differently focused entities.

3.1 Method

Subjects

A total of forty subjects (mean age 23 ± 2.7 years) participated in Experiment 2. They were all native speakers of French, and were not paid for their participation.

Materials

For Experiment 2, a set of forty new experimental texts was constructed. Indeed, although initially we wanted to modify the texts of Experiment 1 so that the two characters would be ambiguous from a gender point of view (with two first names of the same gender), we quickly became aware of the difficulty of changing the target sentence verbs in order to coherently orient the predication within the target sentence, either towards entity 1 or entity 2. Therefore, new texts in which the second character (i.e. entity 2) was more involved in the situation described were constructed. In this way, it was easier to find verbs with the desired referential orientation.

As in Experiment 1, the experimental texts had three sentences (see Table 4). The first sentence was an introductory sentence. The second sentence introduced two same-sex characters, one of whom was the highly-focused entity (entity 1: coded in subject position via a first name), and the other was the less-focused entity (entity 2: coded in indirect object position via a description of his or her role in the setting, e.g. *the schoolmistress*). The third sentence, the target sentence, referred to one of the two characters mentioned in the second sentence, either via the 3rd person pronoun (*il/elle*), or via the demonstrative pronoun (*celui-ci/celle-ci*).

As the gender cue was not relevant in the processing of the pronominal expressions, two types of target sentence with a different predicative component were constructed for each entity type, orienting the whole predication within the target sentence either towards entity 1 or entity 2. Both types of target sentence were roughly equal in length (a mean of 5.8 words). Each text ended by a 'yes-no' question which probed comprehension of the target sentence. A total of four experimental conditions (2 entity types × 2 pronoun types) were created for each experimental text. As in Experiment 1, forty-eight filler texts were added. Half of them were identical to the experimental texts, the other half were different, but the 'yes-no' questions did not test the interpretation of pronouns for the fillers.

Pronoun Type	Entity Type	
	Entity 1	Entity 2
3rd person anaphoric pronoun	Les élèves de l'école se défoulaient pendant la récréation. *(The schoolchildren were letting off steam at playtime.)* Marie a donné un coup de pied à la maîtresse dans la cour./ *(Marie kicked the schoolmistress in the playground) /*	Les élèves de l'école se défoulaient pendant la récréation. *(The schoolchildren were letting off steam at playtime.)* Marie a donné un coup de pied à la maîtresse dans la cour./ *(Marie kicked the schoolmistress in the playground) /*
TS:	Elle a été sévèrement punie./ *(She was severely punished) /*	Elle a eu un gros hématome./ *(She got a nasty bruise) /*
demonstrative pronoun	Les élèves de l'école se défoulaient pendant la récréation. *(The schoolchildren were letting off steam at playtime.)* Marie a donné un coup de pied à la maîtresse dans la cour./ *(Marie kicked the schoolmistress in the playground) /*	Les élèves de l'école se défoulaient pendant la récréation. *(The schoolchildren were letting off steam at playtime.)* Marie a donné un coup de pied à la maîtresse dans la cour./ *(Marie kicked the schoolmistress in the playground) /*
TS:	Celle-ci a été sévèrement punie./ *(This one (FEM.SG) was severely punished) /*	Celle-ci a eu un gros hématome./ *(This one (FEM.SG) got a nasty bruise) /*

Note. The referent-entity (entity 1 or entity 2) is underlined for expository purposes. In the text presented to participants, nothing was underlined. The slash indicates the text presentation on the screen. The target sentence (TS) was presented separately on the screen. The English translation of this example is indicated in brackets in italics.

Table 4: Example of an experimental text in all four conditions in Experiment 2

Design and Procedure

The experimental design and the self-paced reading procedure were the same as those for Experiment 1, except that the subjects answered 'yes-no' questions rather than 'true/false' statements.

3.2 Results and Discussion

The means of the reading times (RTs) for the target sentences were calculated for each subject and for each item in each condition by the same method as for Experiment 1 (5.5% of the data were eliminated). As previously, RTs were normalized in order to take target sentence length (number of characters) into account (See Table 5).

Pronoun Type		Entity Type	
		Entity 1	Entity 2
3rd pers. Pron.	*Error %*	2.75%	4.25%
	Mean RT	49.4	59.1
	SD	(11.4) [8.1]	(13.8) [11,4]
Dem. Pron.	*Error %*	7%	5.5%
	Mean RT	58.9	49.5
	SD	(16.3) [8.9]	(12.3) [9.2]

Note. The reading times are given in milliseconds per character. The values in parentheses are the standard deviations with subjects as the random factor; the values in brackets are the standard deviations with items as the random factor.

Table 5: Means and standard deviations of the reading times of the target sentences and mean percent of errors in Experiment 2

A 2 × 2 analysis of variance of Entity-Type × Pronoun-Type was conducted separately for subjects as the random factor (F_1), and for items as the random factor (F_2).

Accuracy of answer for the questions

For all the questions, there was a 95% correct answer rate. As in Experiment 1, neither the main effect of entity-type nor the interaction between entity-type and pronoun-type were significant, $F1, F2 < 1$. However, unlike Experiment 1, the main effect of pronoun-type was significant in this experiment ($F1(1,39)=8.125$, $p <.01$; $F2(1,39)=9.286$, $p <.01$) with more errors when a demonstrative pronoun was used in the target sentence.

Finding worse performance for questions dealing with the identification of the demonstrative pronoun referent could suggest that, when gender cue is not relevant, the use of a demonstrative pronoun would become much more sensitive to the focusing constraint. Therefore, the subjects would find 'breaking' this constraint less acceptable. Indeed, the 'unusual' use of the demonstrative pronoun to refer back to the highly-focused entity seems to have particularly constricted the subjects (maximal error rate in this condition: 7%), which could explain the increase in error rate for the demonstrative pronouns.

Reading times of the target sentences

Contrary to those of Experiment 1, the reading time results revealed no main effect for either entity-type or pronoun-type, F_1, $F_2 < 1$. However, as in the previous experiment, there was a significant interaction effect between the entity-type and the pronoun-type, $F_1(1,39)= 39.3$, $p < .001$; $F_2(1,39)= 145.4$, $p < 001$. As in Experiment 1, this interaction indicated a strong effect of focus (i.e. the highly-focused entity) on the processing of the 3rd person anaphoric pronoun, $F_1(1,39)=26.4$, $p < .001$; $F_2(1,39)=81$, $p < .001$, as well as a specific referential functioning of the 3rd person anaphoric pronoun to refer back to entity 1 rather than to entity 2, $F_1(1,39)=38.2$, $p < .001$; $F_2(1,39)=17$, $p < .001$.

The most notable point in the results, however, was that the demonstrative pronoun, this time, was sensitive to entity focusing. Not only was the simple effect of entity 2 conditions within pronoun-type significant, $F_1(1,39)= 32.7$, $p < .001$; $F_2(1,39)=58.3$, $p < .001$ (the target sentences referring back to entity 2 were read faster with a demonstrative pronoun than with a 3rd person pronoun), but it was also true for the simple effect of the demonstrative pronoun conditions within entity-type, $F_1(1,39)=19.8$, $p < .001$; $F_2(1,39)=24.2$, $p < .001$ (the target sentences containing a demonstrative pronoun were read faster when they referred back to entity 2 than when they referred to entity 1).

In brief, the 'lack' of any relevant gender cue in this experiment seems to have revealed the specialization of the demonstrative pronoun *celui-ci/celle-ci* in accessing the less-focused entity.

More notably than in Experiment 1, these last results suggest that both the pronoun types tested here are sensitive to the focus status of discourse entities, but in opposite ways: if the 3rd person pronoun is specialized for maintaining the addressee's attention on the highly-focused entity, the demonstrative pronoun *celui-ci/celle-ci* would be specialized in re-orienting the addressee's attention towards the less-focused entity.

4 General Discussion

The results reported here suggest that the referential functioning of the demonstrative pronoun *celui-ci/celle-ci* is constrained in terms of entity focusing, but in an opposite way in comparison with that of the 3rd person pronoun *il/elle*.

Is the pronoun *celui-ci* really the mirror image of the pronoun *il*, for all that? Apparently not. Indeed, we saw from the results of Experiment 1 (cf. Table 3) that the convergence of both focus constraint and gender cue towards only one referent was clearly not sufficient for the processing of the demonstrative pronoun to give rise to a rapid identification of its referent (namely, Entity 2).[7] Unlike the 3rd person pronoun, for which it seems that both these constraints (focus and gender) are sufficient to make an early commitment to resolution in favour of Entity 1 (Arnold *et al.*, 2000; Garrod *et al.*, 1994; Rigalleau &

[7] The difference of reading times between both the versions of the demonstrative pronoun was not significant in this first experiment.

Caplan, 2000; Sanford & Garrod, 1989) the demonstrative pronoun would be more ‘demanding’ in terms of cognitive effort.

Thus, we believe that *celui-ci* is not a converse copy of the pronoun *il*. The way in which this demonstrative pronoun selects its referent is not simply the same as that of the anaphoric pronoun, but inverted.[8] We believe, instead, that *celui-ci* must also conform to another constraint, namely the one related to context which provides a specific class of entities on the basis of which the pronoun *celui-ci* will be able to isolate – and this by virtue of its deictic value – the least focused entity among the set of entities previously evoked.

This additional constraint to which the demonstrative pronoun must conform, which we called *‘the selection–presupposition constraint’*, would seem to account for the fact that the ‘demonstrative capture’ that the pronoun *celui-ci* operates on its referent is not without consequences on the cognitive processing of this pronoun. This additional presuppositional constraint which is related to the contrastive function of the pronoun *celui-ci* – whose effect is to isolate an element from a class of similar elements – would seem to be specific to this type of pronoun and it is not assumed to exist for the 3rd person anaphoric pronoun.

Indeed, unlike the demonstrative pronoun, the anaphoric pronoun does not select its referent in terms of a contrast, and so, its resolution does not need to be based on a specific class of entities. In any case, it functions very well with only one (salient) entity, which is not true of the demonstrative pronoun.[9] The latter, on the contrary, given its hybrid nature, being both anaphoric and deictic, must conform to the selection-presupposition constraint in order to be used in an appropriate manner. As it selects its referent in terms of a contrast, by extracting it from among a set of potential candidates, the demonstrative pronoun would appear to be more demanding in terms of cognitive effort.

Taking the presupposition of selection constraint into account could then explain the additional processing load observed in the case of the French demonstrative pronoun studied here. This proposition, obviously, would need to be explored in greater depth, using a more direct methodology like eye-tracking

---

[8] See, also, Kaiser and Trueswell’s study (2003) in Finnish on the referential contrast between the gender-neutral pronoun *hän* (he/she) and the demonstrative *tämä* (this, he/she). Using an eye-movement paradigm, authors show that these two forms are not mirror images of one another. Unlike the pronoun *hän* which tends to refer to subjects, the demonstrative pronoun *tämä* would look at the discourse/pragmatic level and would be used for low-salience referents. “The referential properties of *hän* and *tämä* are not subject to a single common factor, and an accessibility scale is not sufficient to explain the division of ‘referential labour’ between these two expressions” (Kaiser & Trueswell, 2003:6).

[9] Unlike the pronoun *celui-ci*, the pronoun *il* functions very well with only ONE salient entity. Example: *‘Marie pleure. Elle est triste (Mary is crying. She is sad)’*. But, *‘Marie pleure. * Celle-ci est triste (Mary is crying. * This one (FEM.SG) is sad).*

or even evoked response potentials (ERP). These techniques have a higher level of resolution than the one used here. Therefore, our results, obtained from a segment-by-segment reading, are preliminary. A word-by-word reading procedure allowing the observation of the pronoun processing effect (as a specific element) would also have been possible, however, a preference has been given to longer segments giving a more natural reading.

Finally, the conclusions of this study about the pronoun *celui-ci* may not be extended to demonstratives in general, in particular the most frequent and unmarked demonstratives such as English *this* or French *ce*. *Celui-ci* and similar devices in other languages are somewhat related to demonstratives, but they are a very specialized kind of demonstratives.

References

Ariel, M. 1990. *Accessing Noun Phrase Antecedents*. London: Routledge, Croom Helm.

---------- 1996. "Referring expressions and the +/- coreference distinction." T. Fretheim and J. Gundel (eds.), *Reference and referent accessibility*. Amsterdam: John Benjamins. 13-25.

Arnold, J., J. Eisenband, S. Brown-Schmidt and J. Trueswell 2000. "The Rapid Use of Gender Information: Evidence of the Time Course of Pronoun Resolution from Eyetracking." *Cognition* 76.B13-B26.

Chafe, W. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.

Charolles, M. 1991. "L'Anaphore: Définition et Classification des Formes Anaphoriques." *Verbum* 14.203-216.

---------- 1995. "Comment Repêcher les Derniers? Analyse des Expressions Anaphoriques en 'ce dernier'." *Pratiques* 85.89-112.

Cornish, F. 1995. "Référence Anaphorique, Référence Déictique, et Contexte Prédicatif et Énonciatif." *Sémiotiques* 8.31-55.

---------- 1999. *Anaphora, Discourse and Understanding. Evidence from English and French*. Oxford, Linguistics: Clarendon Press.

De Mulder, W. 1998. "Celui-ci et Celui-là : Encore un Couple Mal Assorti?" C. Schnedecker (ed.) *Les corrélats anaphoriques*, 97-129. Recherches Linguistiques, 22. Université de Metz.

Fossard, M. 1999. "Traitement Anaphorique et Structure du Discours. Etude Psycholinguistique des Effets du 'Focus de Discours' Sur la Spécificité de Deux Marqueurs Référentiels: Le Pronom Anaphorique 'il' et le Nom Propre Répété." *Cognito* 15.33-40.

---------- 2001. "Aspects Cognitifs de L'Anaphore Pronominale. Approche Psycholinguistique et Ouverture Neuropsycholinguistique Auprès de Deux Patientes Atteintes de Démence Type Alzheimer." Ph.D. thesis, Université de Toulouse-le Mirail.

Garnham, A. 1997. "Representing Information in Mental Models." M. Conway (ed.), *Cognitive Models of Memory*, 149-172. Hove, East Sussex: Psychology Press.

Garrod, S. 1994. "Resolving Pronouns and Other Anaphoric Devices: The Case for Diversity in Discourse Processing." Clifton, Frazier and Rayner (eds.) *Perspectives on Sentence Processing*, 339-359. Englewood: LEA.

----------, D. Freudenthal and E. Boyle 1994. "The Role of Different Types of Anaphor in the On-line Resolution of Sentences in a Discourse." *Journal of Memory and Language* 33.39-68.

Givon, T. 1983. “Topic Continuity in Spoken English.” Givon (ed.) *Topic Continuity in Discourse*. Amsterdam: John Benjamins. 345-363.

Gordon, P., B. Grosz and L. Gilliom 1993. “Pronouns, Names, and the Centering of Attention in Discourse.” *Cognitive Science* 17.311-347.

---------- and R. Hendrick 1998. “The Representation and Processing of Coreference in Discourse.” *Cognitive Science* 22.389-424.

---------- and K. Scearce 1995. “Pronominalization and Discourse Coherence, Discourse Structure and Pronoun Interpretation.” *Memory and Cognition* 23(3).313-323.

Greene, S., G. McKoon and R. Ratcliff 1992. “Pronoun Resolution and Discourse Models.” *Journal of Experimental Psychology: Learning, Memory and Cognition* 18(2).266-283.

Grosz, B., A. Joshi and S. Weinstein 1995. “Centering: a Framework for Modeling the Local Coherence of Discourse.” *Computational Linguistics,* 21(2): 203-225.

Gundel, J. 1998. “Centering Theory and Givenness Hierarchy: Towards a Synthesis.” M. Walker, A. Joshi and E. Prince (eds.), *Centering Theory in Discourse*. Oxford: Clarendon Press. 183-198.

----------, N. Hedberg and R. Zacharski 1993. “Cognitive Status and the Form of Referring Expressions in Discourse.” *Language* 69(2).274-307.

Hudson-D’Zmura, S. and M. Tanenhaus 1998. “Assigning Antecedents to Ambiguous Pronouns: The Role of the Center of Attention as the Default Assignment.” M. Walker, A. Joshi, and E. Prince (eds.) *Centering Theory in Discourse*. Oxford: Clarendon Press. 199-226.

Kaiser, E. and J. Trueswell 2003. “Dividing Up Referential Labor: Finnish Pronouns and Demonstratives in On-line Processing.” *Proc. of the 9th Annual Conference on Architectures and Mechanisms for Language Processing* (AMLaP-2003, Glasgow, 25-27 August 2003: 6-7).

Kleiber, G. 1994. *Anaphores et Pronoms. Champs Linguistiques*. Duculot. Louvain-la-Neuve.

---------- 2001. “Regards sur l’Anaphore et la Givenness Hierarchy.” Kronning, Noren, Noven, Ransbo, Sundell and Svane (eds.) *Langage et Reference, Melanges Offerts à Kerstin Jonasson*. Acta Universitatis Upsaliensis, Uppsala, 311-322.

Marslen-Wilson, W., L. Tyler and J. Koster 1993. “Integrative Processes in Utterance Resolution.” *Journal of Memory and Language* 32.647-666.

McKoon, G., S. Greene and R. Ratcliff 1993. “Discourse Models, Pronoun Resolution, and the Implicit Causality of Verbs.” *Journal of Experimental Psychology: Learning, Memory and Cognition* 18.1040-1052.

Rigalleau, F. and D. Caplan 2000. “Effects of Gender Marking in Pronominal Coindexation.” *The Quarterly Journal of Experimental Psychology,* 53(A), 1: 23-52.

Sanford, A. and S. Garrod 1981. *Understanding Written Language: Explorations Beyond the Sentence.* Chichester: John Wiley.

---------- and ---------- 1989. “What, When and How: Questions of Immediacy in Anaphoric Reference Resolution.” *Language and Cognitive Processes* 4.263-287.

----------, K. Moar and S. Garrod 1988. “Proper Names as Controllers of Discourse Focus.” *Language and Speech* 31.43-56.

Strauss, S. 2002. “*This, That,* and *It* in Spoken American English: A Demonstrative System of Gradient Focus.” *Language Sciences* 24.131-152.

Walker, M., A. Joshi and E. Prince 1998. “Centering in Naturally Occuring Discourse: An Overview.” M. A. Walker, A. K. Joshi and E. F. Prince (eds.). *Centering Theory in Discourse*: Oxford: Clarendon Press. 1-28.

Winer, B., D. Brown and K. Michels 1996. *Statistical Principles in Experimental Design*. New York: McGraw-Hill.

Section III

Corpus-Based Studies

The Predicate-Argument Structure of Discourse Connectives: A Corpus-Based Study

Cassandre Creswell[1], Katherine Forbes[2], Eleni Miltsakaki[1], Rashmi Prasad[1], Aravind Joshi[1] and Bonnie Webber[3]

[1]University of Pennsylvania, [2]University of Pittsburgh, [3]University of Edinburgh

Discourse connectives can be analysed as encoding predicate-argument relations whose arguments derive from the interpretation of discourse units. These arguments can be anaphoric or structural. Although structural arguments can be encoded in a parse tree, anaphoric arguments must be resolved by other means. A study of nine connectives, annotating the location, size, and syntactic type of their arguments, shows connective-specific patterns for each of these features. A preliminary study of inter-annotator consistency shows that it too varies by connective. Results of the corpus study will be used in the development of resolution algorithms for anaphoric connectives.

1 Introduction

Discourse connectives can be analysed as encoding a relation between two discourse segments. In other words, the semantic interpretation of a discourse connective is a predicate that takes discourse units as its arguments. These arguments can be derived anaphorically or structurally. We describe this distinction below in more detail. Roughly, structural arguments can be encoded in a parse tree, but anaphoric arguments must be resolved by other means.

The distinction between anaphoric and structural arguments is a theoretical one based on a discourse lexicalised tree-adjoining grammar (DLTAG). In DLTAG, the compositional part of discourse meaning (projected by the tree structures) is divided from the non-compositional contributions due to general inferencing and anaphora. This division is a key insight of the DLTAG approach to discourse structure which simplifies the set of structures that can be assigned to a discourse.

With respect to any particular connective, its categorization as taking its arguments structurally or anaphorically is an empirical question. Because only structural arguments can be derived from a DLTAG discourse structure, the location of anaphoric arguments is an additional issue that requires empirical investigation of linguistic data. This corpus study is undertaken as a preliminary attempt to annotate discourse connectives' arguments in order to provide evidence for 1) whether the arguments of discourse connectives can be reliably annotated; 2) whether to classify particular connectives as structural or anaphoric; and 3) whether anaphoric arguments of connectives display properties that would allow development of robust resolution algorithms for locating them.

The results of this corpus study of nine connectives, where the location, size, and syntactic type of their arguments were annotated, sheds light on all three of these issues. First, the data do provide evidence for characterizing certain connectives as anaphoric or structural. In addition, with respect to the features examined here, we found a range of connective-specific behaviours. Finally, a preliminary study of inter-annotator consistency shows that its reliability also varies by connective. The results of this corpus study will be useful for parsing discourse structure, for developing resolution algorithms for anaphoric arguments of connectives, and for revising the annotation guidelines in preparation for a large-scale study of discourse connectives and their arguments.

The structure of the paper is as follows: In Section 2, we provide the theoretical background necessary to understand the distinction of interest here between *structural* and *anaphoric* connectives. This background includes a brief introduction to LTAG and DLTAG. Then, in Section 3, we describe the corpus study undertaken, including its guidelines, results, and an assessment of the reliability of the annotation. In Section 4, we examine the implications that variation in the annotations has for the ability to develop resolution algorithms. We conclude in Section 5 with a discussion of future annotation and algorithm development efforts.

2 Theoretical Background: LTAG and DLTAG

The theoretical background of this study of discourse connectives is Discourse Lexicalised Tree Adjoining Grammar (DLTAG) (Webber *et al.*, 2003). DLTAG builds an intermediate level of discourse structure directly on top of the clause. DLTAG's syntax is currently modelled using the structures and structure-building operations of a lexicalised tree-adjoining grammar (LTAG) (Joshi *et al.*, 1975), which is widely used to model the syntax of sentences.

2.1 LTAG

Briefly, an LTAG is a lexicalised extension of a tree-adjoining grammar (TAG). The object language of an LTAG is a set of trees, allowing the underlying structure of a surface string to be represented, as well as the string itself. An LTAG consists of a finite set of *elementary* trees and operations for combining them. Elementary trees are associated with at least one lexical item, called an *anchor*. They represent extended projections of the anchor and encode its subcategorization frames. An anchor may be associated with more than one tree; each tree in this tree *family* reflects a different syntactic construction in which that anchor can appear. For example, the verb *eat* may anchor either a transitive or intransitive tree.

There are two types of elementary trees in an LTAG: *initial* trees, which encode basic predicate-argument relations, and *auxiliary* trees, which encode optional modification and must contain a non-terminal node (called the *foot* node) whose label matches the label of the root. The rightmost tree in Figure 1 is an auxiliary tree, all the others are initial trees.

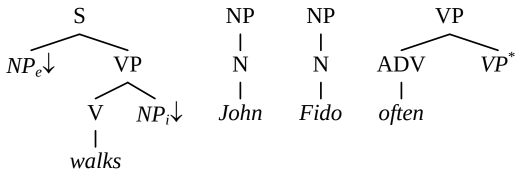


Figure 1: Elementary LTAG trees

There are two structure-building operations in LTAG that create complex trees called *derived* trees: *substitution* and *adjunction*. As shown in Figure 1, substitution sites are indicated by ↓ and adjunction sites, by *.

Substitution consists of replacing the node marked ↓ with the tree being substituted. Only initial trees or trees derived from initial trees can be substituted, and the root node of the tree being substituted must match the label of the node being replaced. For example, the tree anchored by *Fido* in Figure 1 can substitute for the internal argument (NP_i) in the tree anchored by *walks*, and the tree anchored by *John* can substitute for the external argument (NP_e) in the tree anchored by *walks*. The result of these substitutions is shown in Figure 2(a).

Adjunction is restricted to non-terminal nodes not already marked for substitution, building a new tree from an auxiliary tree β and any other tree τ (initial, auxiliary, or derived). To combine β and τ by adjunction, the root node of β must match the label of the node *n* in τ to which it is to be adjoined. The root node of β is identified with *n*; the subtree dominated by *n* is attached to the foot node of β, and the rest of the tree that dominated *n* now dominates the root node of β. For example, the tree anchored by *often* in Figure 1 can adjoin to the VP node in Figure 2(a), producing the derived tree in Figure 2(b).

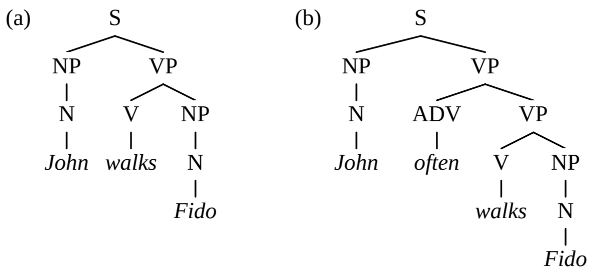


Figure 2: LTAG derived trees after substitution and adjunction

2.2 DLTAG

DLTAG is an extension of LTAG in which the elementary trees are anchored by discourse connectives. Discourse connectives can be analysed as encoding predicate-argument relations whose arguments are the interpretations of discourse segments. The elementary trees anchored by connectives combine with each other and with derived trees to create a structure for a multi-sentence discourse. That is, DLTAG is a grammar for combining sentences into a discourse rather than for combining words into sentences. A lexicalised grammar at the discourse level can capture the inter-sentential relations encoded by connectives and allow an extension of compositional semantic representations from the sentence level to the discourse level.

Just as at the sentential level, arguments to these discourse relations can be found structurally or anaphorically. Here, *structurally* means the semantic content of the argument must be derivable locally. At the sentential level, an example of a relationship with a strictly structural basis is the relationship between a reflexive pronoun and its antecedent, as in (1), where *himself* must co-refer with *John*, the subject of the sentence.

(1) John saw **himself** in the mirror.

The reflexive pronoun and its antecedent must have a particular relationship to each other in the syntactic tree, one where they are both present in the same clause, i.e. arguments of a single predicate. The antecedent of a free pronoun, however—although there are positions in which it cannot appear with respect to the pronoun—does not need to have any particular local syntactic relationship with the pronoun, and so it can be found within or outside the same sentence. This is illustrated in (2) where *she* may be coreferent with either *Jan* or *Fran*.

(2) Jan called. Fran said that **she** might come over later.

Locating the antecedent – and therefore computing the interpretation – of a free pronoun relies on anaphoric and inferential mechanisms. This difference between a structural and an anaphoric relationship at the sentential level is analogous to the one found with the arguments of discourse connectives at the discourse level.

Every discourse connective will find at least one of its arguments structurally, the argument that substitutes into one of the leaf nodes in the tree anchored by the discourse connective. Its other argument may be found either structurally or anaphorically. We will refer to connectives that find one of their arguments anaphorically as *anaphoric connectives*; the others as *structural connectives*.[1] The difference between the two types can be most easily seen in an example where multiple connectives appear together (Webber *et al.*, 2000), like (3).

(3) S_1: Sally rarely eats meat and subscribes to *Vegetarian Times*.
S_2: Lately, she's raised the ire of her vegan friends
S_3: because she nevertheless enjoys the occasional bacon cheeseburger.

In (3), *becaus*e, a structural connective at the discourse level, is the predicate expressing the causal relation between two eventualities, P = RAISE IRE (SALLY, FRIENDS) and Q = ENJOYS (SALLY, CHEESEBURGER). This is encoded formally with the two argument nodes appearing in the same elementary tree, shown in Figure (3).[2]

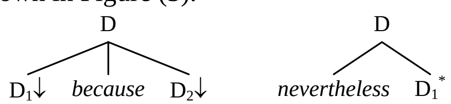


Figure 3: Elementary tree: because and nevertheless

In contrast, the connective *nevertheless* in S_3 finds only a single argument structurally Q = ENJOYS (SALLY, CHEESEBURGER). Its other 'left-hand' argument is derived anaphorically from S_1. The formal way of capturing this difference is assigning a different type of elementary tree to *nevertheless*, also

[1] We view this as a lexical property of a particular connective. If a connective can ever be found with an non-adjacent, non-contiguous, or only inferentially-derivable (rather than textually-derivable) anaphoric argument, it is an anaphoric connective. In a given use of an anaphoric connective, however, its anaphoric argument might occur in the immediately preceding text. As such, to decide what category a particular connective falls into, if a convincing example of an anaphoric use cannot be constructed, then many naturally-occurring examples may also be needed to make this designation.

[2] In general, the theory has only been applied to monologic text, primarily written rather than spoken. This analysis will have to be extended to account for structural connectives that appear in sentence fragments in dialog, as in *Because I said so.* We suspect that fragments containing structural connectives like *because* will pattern with their structural counterparts in written text rather than with anaphoric connectives, but a detailed study remains for future work.

shown in Figure (3); here, the discourse clause to which the *nevertheless* tree adjoins, D_1, is its sole structural argument.[3]

Not all discourse segments (elementary or complex) are related via a lexically explicit discourse connective. In DLTAG, such relations are handled by an auxiliary tree anchored in a lexically-empty discourse connective that conveys *continuation* of the description of the larger tree to which it is attached. Although a more specific relation may be inferred, the relation provided by the syntax alone is semantically underspecified, analogous to the semantics of noun-noun compounds.[4] In the discourse tree derived from (3), S_2 is attached to the previous discourse with an auxiliary tree anchored by a lexically-empty connective.

The full derived tree for the discourse in (3) is shown in Figure (4).

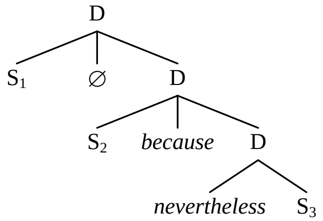


Figure 4: Full discourse tree

We can see that the arguments of structural connectives are encoded directly in a parse tree and, therefore, are relatively easy to identify.[5] The non-structural argument of an anaphoric connective must be resolved by other means.[6] Once again, this is similar to the case of bound versus free pronouns.

---

[3] Adverbial connectives may appear sentence initially, medially or finally. The position of the adverbial connective in the sentence affects the scope of the connective and is often associated with the information structure partitioning of the sentence into focus and ground (Kruijff-Korbayová & Webber, 2001). With respect to parsing discourse structure using a DLTAG, in cases of medial and final discourse adverbials, a sentence-initial copy of the adverbial is added during parsing. This makes it possible to use the same elementary tree structure anchored in that lexical connective no matter where the connective appears at the sentence level. An index is retained inside the sentence to retain information about its clause internal position. The discourse-syntactic role then remains the same regardless of its sentence-level syntax. See (Forbes *et al.*, 2003) for more detail about the use of DLTAG in parsing.

[4] There may be some limits on the types of relations that may be inferred without the specific use of a discourse connective. Presumably, this will depend on the contributions of sentential semantics, syntax and prosody to the inferential process.

[5] In fact, structural connectives are associated with attachment ambiguity in the parse tree, and so although once a parse tree is created, identifying them is trivial, the determination of the parse is not itself trivial.

[6] Although the missing argument will not be in the syntactic tree, it will be represented in the semantics of that tree, e.g. *e* in the semantics for *nevertheless*:

(1) NEVERTHELESS(x, $[[e_i]]^{ac}$)

Here the semantics links the *x* argument to an address in the tree, but the *e* argument is not linked to an address; it is represented using an assignment function. Assignment functions have already been used to

Another property discourse anaphoric connectives share with other types of discourse anaphora is that their anaphoric arguments may be found intra- or inter-sententially, as in (4) and (5), respectively.

(4) A person who seeks adventure might, for example, try skydiving.
[Webber *et al.* (2000)'s ft.8 (i)]

(5) Some people seek adventure. For example, they might try skydiving.

Because discourse connectives are some of the clearest indicators of discourse structure, annotating the arguments of the relations they convey provides information both about those arguments and about the range of possible discourse structures. In order to use information about discourse connectives to parse discourse structure, we need to know for any particular discourse connective whether it is structural or anaphoric. In order to make this designation, a systematic empirical study which shows the behaviour of the connective over a significant number of cases is required. In addition, for anaphoric connectives, in order to develop a resolution algorithm, symbolic or statistical, which can identify anaphoric arguments, a corpus study which provides evidence for patterns of location and properties of anaphoric arguments is a necessary first step. The corpus study undertaken here is very exploratory, but its general goal is to provide evidence to characterize particular connectives as structural or anaphoric, and if the latter, identify features characteristic of their anaphoric arguments.

3 Corpus Study

This work is a subset of a larger discourse annotation project whose main goal is to create a large corpus reliably annotated for discourse structure for further scientific research and development of NLP applications (e.g. question-answering, text summarization) (Miltsakaki *et al*., 2004). Each overt or null discourse connective in the corpus will be marked with the minimal textual unit in the preceding discourse which contains the source of its left-hand argument. Although for the purposes of both the present study and the larger corpus study a strictly textual antecedent is being identified, this is a practical simplification of the theory. In fact, the anaphoric argument is more accurately treated as an Abstract Object (AO) in the discourse model, the same kind of entity that can be accessed through a demonstrative pronoun (or *discourse deictic*), as argued

represent pronominal reference (Heim & Kratzer, 1998), i.e. an pronoun denotes an entity *e* via an index *i* that is mapped to *e* relative to an assignment function *a*, where *a* is determined by a context *c* (e.g. $[[\text{you}_i]]^{ac}$, where *c* might yield *Tom, Dick, Harry* for *i*).

for in (Forbes, 2003).[7] However, in light of the fact that many successful current approaches to anaphor resolution of NPs, apply to surface elements, from an engineering perspective, identification of the textual material which gives rise to the AO is a more realistic task. The success of the overall project will contribute to our ability to understand and deal with an important aspect of discourse meaning, i.e. discourse relations.

3.1 Corpus Annotation

The work we report here is a first attempt to taxonomize the set of discourse connectives and their properties. To the best of our knowledge, annotation of the arguments of discourse connectives has not been previously attempted. As such, the annotation here is necessarily experimental and explorative, and to some extent the annotation guidelines were developed and altered as the annotation took place. We began with a set of nine connectives picked from three semantic classes: resultatives *(as a result of, so, therefore*), concessives *(nevertheless, yet, whereas*), and additives *(also, in addition, moreover*). They are all adverbials, which may, by definition, modify phrasal constituents (ADJP, PP, VP) or an entire clause.[8] [9]

For each of the nine connectives, seventy-five tokens (for a total of 675 tokens) were extracted from a variety of corpora: Brown, Wall Street Journal, Switchboard and 58 transcribed oral histories from the online Social Security Administration Oral History Archives (SSA).[10] The 675 tokens were split in three groups (each group containing a connective from each semantic class) and annotated by three annotators (225 tokens per annotator).

Each token was annotated with tags that encoded information about (a) the connective's left argument (ARG), and (b) the clause containing the connective (CONN). Table 1 shows the ARG and CONN tag(sets) in the top and bottom box respectively. Both ARG and CONN were annotated with a REF tag that encoded an ID number which linked the two parts of the single token. ARG was

---

[7] Note also that, just as with bridging reference for NPs, the argument of a discourse adverbial may be an abstract object derived from, but not identical to, an AO already in the discourse model (Webber *et al.*, 2003).

[8] In other words, lexical items that function as connectives also have other syntactic roles at the sentential level (e.g. *he is* **otherwise** *occupied,* **hereafter** *happy to eat tofu,* **so** *tire*d, etc.). This study excludes these other uses on the grounds that they must be accounted for as part of sentential syntax.

[9] Although *whereas* can be used as an adverbial connective, in our data it mostly appears as a subordinate conjunction because a clause introduced by it can appear before or after its other clausal argument, as shown in example (1).

(1) Whereas persons of eighth grade education or less were more apt to avoid or be shocked by nudity, those educated beyond the eighth grade increasingly welcomed and approved nudity in sexual relations. (Brown)

[10] The Brown, Wall Street Journal and Switchboard corpora are available from LDC, http://www.ldc.upenn.edu. The SSA corpus is available at http://www.ssa.gov/history/orallist.html

further tagged with a TYPE tagset that identified the extent of the argument. The tags under TYPE were as follows: MAIN if the argument was contained in a full sentence (including subordinate clauses); MAIN-MULT if the argument was contained in a sequence of sentences; SUB if the argument was contained in a subordinate clause; and XP if the argument was contained in a phrasal constituent. The variation in the size of the argument was thus specified as a structural description.

In the TYPE tagset, two additional tags were added during the annotation. The category OTHER was added in order to describe cases where the left argument of the connective could not be identified. The category NONE was added for cases where both arguments are to the right of the connective, and therefore, there is no left-hand argument. This tag applies only to cases of subordinate conjunctions, and so it only appeared in the annotation of *whereas* here.

The set of tags used for the type of the left-hand argument were selected in order to enable us to identify statistically useful information about the type of the antecedent of anaphoric connectives, which will ultimately allow the selection of features for use in a statistical or a symbolic anaphora resolution algorithm. In particular, the distinction between MAIN/MAIN-MULT and SUB/XP combined with the LOC tag (discussed in Section 3.3) will help determine optimal structural descriptions for the connectives that will be useful for systems like the DLTAG parser (Forbes *et al.*, 2003). For example, connectives found to take only contiguous MAIN/MAIN-MULT arguments can be associated with a tree taking two structural arguments, thus maximizing compositional semantic representations derived directly from the syntax of discourse.

The clause containing the connective, CONN, was annotated with two tagsets: COMB and POSITION. The tag COMB was used to identify punctuation marks (PERIOD, COMMA, etc.), coordinating conjunctions (‘AND’ and ‘BUT’), and adverbial connectives (‘YET’, ‘SO’, etc.) that can co-occur with the connective.[11] Information about co-occurrence with punctuation

[11] Because all the adverbial connectives were annotated separately, in cases where *so* occurred with one of the adverbial connectives in the study, the token could be annotated both as an instance of that adverbial connective and as an instance of *so*. However, because *so* is a much more frequent connective than any of the adverbials it can potentially combine with, the first 75 instances of adverbial connective *so* that were encountered and used for the study did not include any cases where it combined with another adverbial (i.e. there was no overlap in the set of tokens of *so* annotated as *so* and the set of tokens of other connectives where *so* appeared; the latter were annotated as instances of the other connective.) There were five cases where *therefore* appeared with *so*. Here the effects of *so* and *therefore* on the location of the left-hand argument can be somewhat distinguished by comparing the location of the argument in the actual token with the location of the argument in an example identical but for the absence of *so*. The exploratory

and other (mainly structural) connectives will also be useful for determining syntactic properties of connectives. In DLTAG, *and* and *but* are structural connectives anchoring elementary trees. That is, both their arguments must be realized structurally. Co-occurrence with *and* and *but* may be an indication that a connective cannot take both its arguments structurally because such a structure would be underivable[12] or would require the assignment of computationally complex structural descriptions.

The tag ZERO in the COMB tagset is primarily relevant for tagging tokens of *whereas*. It describes cases where the conjunction combines with no punctuation marks or other connectives. However, in most cases, the presence of this tag indicates that the connective is a subordinate conjunction. Subordinate conjunctions do not combine with a punctuation mark – because of punctuation conventions in written English – or other connectives when the subordinate clause appears after the main clause. The ZERO tag applies much less frequently to adverbial connectives, like *also*.

Finally, we found it useful to make special tags for combinations with a complementizer (COMP) and a subordinate conjunction (SUB) because several connectives appear in complement and subordinate clauses. This creates ambiguity in their interpretation, discussed below in Section 4.1.

For the purposes of anaphora resolution, co-occurrence with punctuation combined with the results of the argument-size (i.e. TYPE) annotation will be useful features in guiding an automated search for anaphoric arguments. Also, certain types of punctuation, e.g. dashes and parentheses, may indicate that the text containing the argument of the connective is not adjacent to the clause containing the connective.

Co-occurrence with other connectives also raises the question of the semantics of the combined connective and its relationship to the semantics of the individual contributors, as for example, in the combinations *and in addition* or *yet nevertheless*.

For CONN, we also defined a POSITION tagset which identified the position of the connective in its clause (INITIAL, MEDIAL, FINAL). As

---

nature of the present study did not allow for full investigation of these effects, but because they were part of the annotation schema, they can be given more attention in a future study.

[12] In other words, the combination of two structural connectives (i.e. appearance within a single clause) cannot be derived under the current framework. This could be an empirically desirable result if there is additional, separate evidence supporting the inability of combining two structural connectives. On the other hand, if independent evidence for characterizing two connectives as structural exists and these connectives can be combined, then possibly the formal framework may have to be altered. The results of this study appear to favor the latter conclusion because *so* appears to be a structural connective, and it can clearly co-occur with *an*d. The necessary revision of the formal framework to account for this phenomenon and/or more detailed investigation of the behavior of *so* remain for future work.

mentioned above, the position of the connective in the clause is relevant for the information structure partitioning of the clause (Forbes *et al*., 2003; Kruijff-Korbayová & Webber, 2001), and thus this is useful information to keep track of with respect to particular connectives for parsing purposes.

A sample full annotation of an instance of *therefore* is shown in example (6). Here the left-hand argument, or ARG, is a main clause (TYPE=MAIN) that immediately precedes the sentence which contains *therefore*. *Therefore* itself appears medially (POSITION=MEDIAL) in a clause introduced by *and* (COMB=AND).

(6) <ARG REF=27 TYPE=MAIN> Philip Lee was the Chancellor of the campus at San Francisco </ARG>
<CONN REF=27 COMB=AND POSITION=MED> and he was **therefore** the person who hired me for the post as Director of the Medical Center. </CONN >

The complete set of tags we defined is given in Table 1, and an example of each tag is provided in examples (7–9).

ARG:	REF	ID #
	TYPE	MAIN = sentence MAIN-MULT = multiple sentences SUB = subordinate clause XP = phrasal constituent OTHER = no argument NONE = no *left* argument
CONN:	REF	ID #
	COMB	PERIOD COMMA COLON SEMI-COLON DASH 'AND' 'BUT' 'YET' 'SO' ZERO COMP SUB
	POSITION	INITIAL MEDIAL FINAL

Table 1: Annotation tagsets

(7) a. MAIN: <ARG>On the basis of the applications, social security cards had been issued to people</ARG> <CONN> and two records therefore came to Baltimore.</CONN> One was the application form, the SS-5; the other was the office record form. (SSA)

b. MAIN-MULT: Well, John Corson as assistant executive director was a wonderful, wonderful foil for Frank Bane, because Frank Bane never likes to say "no" to anybody, you know, even the executive director. And he could always say, "Yes," when the answer was yes. <ARG>But when the answer was, "No," John Corson would always give the answer. And John, knowing that this was the role for the second man, would handle it,</ARG> <CONN> and therefore all the onus that built up in the organization when a bureau director or staff member didn't get what he wanted fell on John and not on Frank.</CONN> (SSA)

c. SUB: And then these people would argue <ARG> we no longer need that sort of effect</ARG> <CONN> and, therefore, we don't need a retirement test any more. </CONN> (SSA)

d. XP: Of course, the contractors were to be <ARG> out there, </ARG> <CONN> and therefore part of the field, </CONN> (SSA)

e. OTHER: *Claims precedent lacking.* After reading his statement discharging the 23d ward case , Karns told Wexler that <CONN> if the seven cases scheduled for trial also involved persons who had been subpoenaed, </CONN>he would dismiss them. (Brown)[13]

f. NONE: <CONN> Whereas most men were a bit ambivalent about the sex scandals (though they were furious about Recruit), </CONN> <ARG> women were upset about both and surged to the polls. </ARG> (Brown)

(8) a. PERIOD: Well that gave me sort of an insight, so I made it a practice to contact all of the funeral directors, which in those days was forbidden. **Nevertheless**, I went ahead and contacted them anyway. (SSA)

b. COMMA: Although Sam Rayburn affects a gruff exterior in many instances, **nevertheless** he is fundamentally a man of warm heart and gentle disposition. (Brown)

c. SEMI-COLON: I am thoroughly convinced that most watercolors suffer because the artist expects nature will do his composing for him; **as a result**, such pictures are only a literal translation of what the artist finds in the scene before him. (Brown)

d. DASH: The 1954 Amendments completely changed the financing of the vocational rehabilitation program, providing for a three-part grant structure – for (1) basic support; (2) extension and improvement; and (3) research, demonstrations, training and traineeships for vocational rehabilitation — and **in addition** for short-term training and instruction. (SSA)

e. 'AND': But it is still a quasi-Independent Agency **and therefore** the ability to be able to speak one's mind is certainly more than it is for traditional cabinet-level officials or senior political officials who serve at the pleasure of the President. (SSA)

f. 'BUT': **But nevertheless** consultation is the prime instrumentality that you use to get support. (SSA)

g. 'SO': **So therefore**, if you have some situations that arise when maybe an ALJ put someone on that the DDS didn't think was disabled, you've got to show the person improved over what the ALJ said before you can take the person off. (SSA)

---

[13] The missing argument here is roughly *The 23rd ward case involved persons who had been subpoenaed.* This proposition is not expressed explicitly anywhere in the article.

h. 'YET': This, plus the habit of many schools of simply adding interior design to the many subjects of their home economics department, **yet**, **nevertheless**, claiming that they teach interior design , has contributed to the low repute of many university courses in interior design . (Brown)

i. ZERO: The Controller's charge of rigging was the latest development in an investigation which **also** brought these disclosures Tuesday : [...] (Brown)

j. SUB: After the first few weeks, it was obvious that rules had to be made, laid down and obeyed — even if our popularity ratings became subnormal **as a result**. (SSA)

k. COMP: Moritz said Monday that his leg feels fine and, **as a result**, he hopes to start practicing field goals this week. (Brown)

(9) a. INITIAL: **Nevertheless** he had ample opportunity to contest the statement before the appeal board. (Brown)

b. MEDIAL: Only those who were actually investors, **therefore**, were eligible for a lump-sum return at reaching age 65 or the widow would receive it at his death. (SSA)

c. FINAL: But it is true, **nevertheless**. (Brown)

Although each tag was explained and illustrated with examples like those above, each annotator was guided wholly by their intuition when determining the values of each tag for each anaphoric argument they annotated. Below in Section 4, we discuss how this intuitive guideline can be further refined: by studying patterns that emerge across similarities and differences between all the annotators' intuitive decisions, we develop a set of heuristics that both improve the guidelines and improve the inter-annotator agreement.

3.2 Annotation Results

Table 2 shows the results of the preliminary annotation for the nine connectives. The table contains percentages of the tags TYPE and POSITION along with the actual number of occurrences of the tags in brackets. In the COMB tagset, a connective could combine with more than one of the categories of the group, so no percentages are given because the numbers do not add up to 75 for each category.

For most connectives, there is a strong tendency for the left argument to be identified locally (in the structural sense) – either in the immediately preceding sentence or in an immediately preceding sequence of sentences, (e.g. the preceding paragraph).[14] Most notably, *so* always takes a sentence or a sequence of sentences (i.e. a segment made up of multiple sentences) immediately preceding it as its left argument, indicating that it may tentatively be treated as a

[14] No limited window was set as a search space for a potential argument. This allowed annotators to look as far back as needed in a particular text to find the location of the argument. In a few rare cases of the 675 tokens examined in the first part of this study, the left argument spanned multiple paragraphs. As can be seen from the results to be presented in Section 3.3, left arguments non-contiguous with their connective are also exceptional. From this, we can conclude that in future development of resolution algorithms, setting the window to be examined to at most the paragraph containing the connective and one or two preceding paragraphs would not harm accuracy greatly.

structural connective. *In addition*, *yet*, *moreover*, *as a result* and *also* tend to take their left argument locally but they demonstrate a larger syntactic variety of potential arguments such as subordinate clauses or phrasal constituents. *So*, *nevertheless* and *moreover* are more likely than the others to take larger discourse segments as arguments, adjacent in the case of *so* and not necessarily adjacent in the case of *nevertheless* and *moreover*. The connective *therefore* often takes its left-hand argument from a subordinate clause. Larger discourse segments appear to lead to vagueness in resolving anaphora (cf. Section 4). For example, it was often difficult to determine the extent of a multi-sentence left-hand argument of *nevertheless*. *Nevertheless* can also find its anaphoric argument in an intra-sentential constituent (XP).

Regarding the position of connectives, *so* appears only in initial position. This supports the claim that *so* is a structural connective because the quintessential structural connectives — subordinate and coordinate conjunctions — are restricted to the initial position. *Also*, on the other hand, frequently appears in medial positions, while the semantically similar *in addition* prefers the initial position.

Connective	in addition	so	yet	nevertheless	moreover	therefore	as a result	whereas	also
Type									
Main	65.3% (49)	45.0% (34)	53.3% (40)	37.3% (28)	42.7% (32)	25.3% (19)	78.6% (59)	46.7% (35)	69.3% (52)
Main-Mult	18.7% (14)	55.0% (41)	33.3% (25)	36.0% (27)	45.3% (34)	21.3% (16)	18.7% (14)	4.0% (3)	9.3% (7)
Sub	5.3% (4)	0.0% (0)	2.7% (2)	9.3% (7)	8.0% (6)	31.0% (24)	2.7% (2)	16.0% (12)	12.0% (9)
XP	10.7% (8)	0.0% (0)	10.7% (8)	17.3% (13)	4.0% (3)	21.3% (16)	0.0% (0)	1.3% (1)	4.0% (3)
(none)	--	--	--	--	--	--	--	32.0% (24)	--
(other)	--	--	--	--	--	--	--	--	5.3% (4)
Comb									
Period	65	33	33	47	68	28	55	26	49
Comma	9	22	14	5	2	1	0	36	7
Semicolon	1	2	8	0	0	0	3	5	0
Dash	1	0	4	0	0	0	0	0	0
And	12	2	8	1	4	41	14	0	7
But	0	0	0	17	1	0	1	0	4
Yet	0	0	0	2	0	0	0	0	0
So	0	0	0	0	0	5	0	0	0
Zero	0	0	0	3	2	0	0	8	1
Comp	0	0	0	0	0	0	1	0	8
Sub	0	0	0	0	0	0	1	0	0
Pos									
Initial	92.0% (69)	100.0% (75)	98.7% (74)	78.6% (59)	82.7% (62)	88.0% (66)	90.7% (68)	100.0% (75)	17.3% (13)
Medial	8.0% (6)	0.0% (0)	1.3% (1)	18.7% (14)	17.3% (13)	12.0% (9)	2.7% (2)	0.0% (0)	80.0% (60)
Final	0.0% (0)	0.0% (0)	0.0% (0)	2.7% (2)	0.0% (0)	0.0% (0)	6.6% (5)	0.0% (0)	2.7% (2)

Table 2: Annotation results for 9 connectives, 75 tokens each

The results of this initial annotation project are promising because they reveal interesting variation in distribution patterns. To further revise the annotation tags and guidelines and, crucially, test inter-annotator reliability, we focused our attention on three connectives *as a result*, *in addition* and *nevertheless*, one from each of the three semantic classes. Another twenty-five tokens of each of the three connectives were extracted to add up to a total of one hundred per connective and give an indication of intra-annotator consistency. The annotation of the complete set of three hundred tokens for the three connectives appears in Table 3. Comparison of Tables 2 and 3 shows that the relative percentages of each tag remained stable, indicating that the anaphoric arguments of each of these connectives display patterns that can be recognized via a large-scale annotation project, and be used to lead to reliable annotation algorithms. What remains to be shown is that this annotation is reliable, such that the same patterns are perceived across annotators.

Connective	In addition	Nevertheless	As a result
Type			
Main	63% (63)	36% (36)	68% (68)
Main-Mult	19% (19)	35% (35)	26% (26)
Sub/Comp	10% (10)	10% (10)	5% (5)
XP	8% (8)	18% (18)	0% (0)
Other	0% (0)	0% (0)	1% (1)
Comb			
Punctuation	101	78	80
Dash	1	0	0
And	12	1	17
But	0	2	1
Conn	0	2	0
Comp	0	0	10
Sub	0	0	1
Pos			
Initial	94% (94)	82% (82)	91% (91)
Medial	6% (6)	16% (16)	3% (3)
Final	0% (0)	2% (2)	6% (6)

Table 3: Annotation results for 3 connectives, 100 tokens of each

3.3 Inter-Annotator Agreement

Our studies in the prior section suggest that a human can identify and find patterns in the arguments of the connectives studied. The study presented in this section suggests that this identification and the patterns found are reliable. To test the reliability of our annotation, three additional annotators annotated 25 of the original 100 tokens of each of the three connectives *(in addition*, *as a result*, *nevertheless*), yielding a total of four annotations of 25 tokens of each of these connectives. Each connective and its anaphoric argument were, as in the prior study, assigned an ID. However, in order to focus on the ability of multiple annotators to agree on the unit from which the anaphoric argument is derived, we employed only one tag, LOC. Each annotator labelled the anaphoric argument with one of the four possible values of this tag shown in Table 4.

LOC:	SS = same sentence PS = previous sentence PP = previous paragraph NC = non-contiguous

Table 4: Values for argument tag LOC

The LOC tag defines the sentence as the relevant atomic unit from which the anaphoric argument is derived. A *sentence* is minimally a main clause and all (if any) of its attached subordinate clauses. The semantic argument of the connective could thus be derived from the single sentence containing the connective (SS), the single prior sentence (PS), a sequence of adjacent sentences (PP), or a sequence of sentences not contiguous to the clause containing the connective (NC). In other words, we did not ask the annotators to distinguish sub-clausal constituents or subordinate clauses; we did not distinguish the exact boundaries of sequences of sentences when we marked more than one sentence as the argument; and we did not distinguish whether a non-adjacent argument comprised one clause or a sequence of them. In this sense, the LOC tag can be viewed as a generalization of the TYPE tag; however, it adds the additional information of whether the anaphoric argument is contiguous to the clause containing the connective. Reasons for employing the LOC tag will be discussed in Section 4.

The Table in the Annex shows the annotations for each set of 25 connective tokens using the LOC tag. The first column indicates the token number being annotated. Then, for each inter-annotation, the first four columns contain the particular LOC tag given to that token by each annotator, and the fifth column shows the proportion of annotators who agreed on a LOC tag for that token, i.e. 4/4 represents the case in which all four annotators produced the same tag, 3/4 represents the case in which three out of four annotators produced the same tag, 2/4 represents the case where two out of four annotators produced the same tag but the remaining two annotators had different tags, and <2,2>/4 represents the case where two annotators produced one tag, and the other two annotators produced another tag.

A summary of the inter-annotator results for the 25 tokens for these three connectives produced using the LOC tag is shown in Table 5. The first column indicates the connective, and the remaining columns contain the percentage of tokens in which a particular pattern of agreement was found for each connective. Again, the first column (4/4) represents the case in which all four annotators produced the same tag, the second column (3/4) represents the case in which three out of four annotators produced the same tag, the third column (2/4) represents the case where two out of four annotators produced the same tag but the remaining two annotators had different tags, and the fourth column

(<2, 2>/4) represents the case where two annotators produced one tag, and the other two annotators produced another tag. That there is no "0" column reflects the fact that in every case, there was some agreement among annotators, e.g. there was no case in which each annotator selected a different tag.

Connective	4/4	3/4	2/4	<2,2>/4
in addition	76% (19)	16% (4)	4% (1)	4% (1)
as a result	84% (21)	12% (3)	0	4% (1)
nevertheless	52% (13)	36% (9)	0	12% (3)

Table 5: Inter-annotator agreement by raw percentages

As Table 5 shows, four-way inter-annotator agreement is greater than 50% in every case, and majority agreement (three-way or better) is 92% for *in addition*, 96% for *as a result*, and 88% for *nevertheless*. Inspection of the individual annotations in the Table in the Annex further demonstrate that the annotators almost always agreed on the use of the SS tag. In other words, the annotators were in agreement when distinguishing anaphoric arguments in the same sentence as the connective from anaphoric arguments farther back in the discourse. The most difficult distinction found across all the connectives concerned whether the anaphoric argument was contained in the prior sentence (PS) or some larger chunk of the prior contiguous discourse (PP). This table also shows that the anaphoric argument was almost always agreed to be contiguous to the clause containing the connective, i.e. the NC tag was rarely used.

Tables 6-8 break down these inter-annotation agreement results by pairs of annotators, using the Kappa statistic. Kappa values are used to measure the degree to which two annotators concur in their respective sortings of N items into k mutually exclusive categories. In the present study, 25 tokens are sorted into one of 4 categories, represented by the 4 values of the LOC tag.[15] Note that these tables show Cohen's unweighted Kappa value for each pair of annotators, for each connective, e.g. the value located in the third row and fourth column of Table 6 shows that the annotations of ANN_k and ANN_c had a Kappa value of 0.88.

One can alternatively compute weighted Kappa values, and this may in fact be more appropriate to this study; however, weighted Kappas require that one can accurately determine how to weight each category. For unweighted Kappa, category weightings are by default set to '1'. Alternative weightings can be determined by the imputed relative distances between categories. At this point, we use unweighted Kappa because determining how to weight each of our LOC tags is an unresolved empirical question. It may be, for example, that there is a

[15] See (Carletta, 1996) for details on computing Kappa values.

tendency to prefer the closest likely discourse unit over others farther away in the discourse as the anaphoric argument. We discuss such issues further in the next section, but the overall question is still an open one.

	ANN_e	ANN_k	ANN_c	ANN_r
ANN_e	--	0.61	0.66	0.74
ANN_k	--	--	0.88	0.81
ANN_c	--	--	--	0.81

Table 6: Kappa values for in addition annotation across 4 annotators

	ANN_e	ANN_k	ANN_c	ANN_r
ANN_e	--	0.67	0.76	0.84
ANN_k	--	--	0.91	0.74
ANN_c	--	--	--	0.83

Table 7: Kappa values for as a result annotation across 4 annotators

	ANN_e	ANN_k	ANN_c	ANN_r
ANN_e	--	0.58	0.59	0.58
ANN_k	--	--	0.65	0.53
ANN_c	--	--	--	0.76

Table 8: Kappa values for nevertheless annotation across 4 annotators

As shown, Kappa values for the *in addition* annotation range from 0.61–0.88, and yield an average Kappa across the 4 annotators of 0.75. Kappa values for the *as a result* annotation range from 0.67–0.91, and yield an average Kappa across the 4 annotators of 0.79. Kappa values for the *nevertheless* annotation range from 0.53–0.76, and yield an average kappa across the 4 annotators of 0.62. Across all three connectives, Kappa values range from 0.53–0.91 and yield an average of 0.72.

Overall, both the raw percentages and the Kappa-statistic evaluations of our inter-annotation agreement reflect the fact that *nevertheless* was more difficult to annotate than either *in addition* or *as a result*. As the project expands, we will likely continue to find both more and less difficult annotation cases. Based on what we've seen so far, however, we conclude that the anaphoric arguments of discourse connectives can be reliably annotated.

In the next section, we discuss how investigating of annotator disagreements can be used to develop a resolution algorithm for the anaphoric arguments of discourse connectives.

4 Towards a Resolution Algorithm

A closer look at 1) how the annotations vary in the inter-annotator study and 2) the results of the more complex annotations in the individual annotation studies, reveals certain issues relevant to developing a resolution algorithm, including the need for a minimal argument heuristic, the existence of true ambiguity in identifying arguments, and the issue of whether anaphora resolution can guide decisions about parsing discourse structure.

4.1 Minimal argument heuristic

As mentioned above, we employed the LOC tag instead of the TYPE tag in the study of inter-annotator agreement. By additionally asking each annotator to record the boundaries of the units she identified as the “exact” unit from which the anaphoric argument was derived, we were able to derive the values for the TYPE tags from each of the four annotations. For the purposes of inter-annotator agreement we found that “exact match” was not a useful comparison, due to differences in the implicit guidelines each annotator was individually following.[16] However, the exact match comparison, combined with the data from the first study, is useful for elucidating these differences and understanding why they arise. The differences between the annotations fall into two main categories, the extent of the argument and the syntactic form of the argument. Both concern the annotator’s understanding of the properties of the unit that are necessary to derive the semantic argument of the connective.

Consider the discourse in (10). When deciding on the anaphoric argument of *as a result*, one annotator might decide that the decrease in blood pressure is the result of the decrease in stress and so tag the argument as PS. Because the decrease in stress about money is a result of Lee winning the lottery, however, another annotator might tag the argument as PP, e.g. as including both the first and second sentences.

(10) Lee won the lottery. So, he was less stressed about money. As a result, his blood pressure went down.

Similarly, consider the discourse in (11). When deciding on the anaphoric argument of as *a result*, one annotator might decide that John’s being a man is the cause of his being drafted (females not being drafted in America historically), and thereby tag the argument as NC because John’s living in the US and registering for the selective service are an elaboration on the concept of being a male American. However, another annotator might tag the argument as PP, e.g. as including the first three clauses.

(11) John is a male American. He has lived in the US his whole life. At 18, he registered for the selective service. As a result, he was drafted.

[16] Because of the exploratory nature of the annotation project, the initial set of guidelines used for annotation were not detailed enough to prevent differences in annotation which would affect our ability to make use of string matching comparisons across annotators in any interesting way. For example, one annotator might systematically include punctuation or other connectives within an argument, while another excludes it.

Finally, consider the sentence in (12). When deciding on the anaphoric argument of *as a result*, one annotator might decide that because *as a result* modifies an adjective on the right, its left argument should be (using the TYPE tag) an XP, e.g. *overworked*. Another annotator might interpret *tired* as a small clause, or a clause with a deleted subject and verb, and so he might tag the entire clause *Kim is overworked* as the anaphoric argument of *as a result* using the MAIN tag. (Note that this issue is avoided when the LOC tag SS is employed.)

(12) Kim is overworked, and as a result, tired.

What all of these cases have in common is the question of how large to make the argument. What they also have in common, however, is that in each case it is possible to select a minimal unit as the argument, and allow the relations between that unit and the surrounding context to complete the interpretation. In (10), if the annotator selects *So, he was less stressed about money* as the argument of *as a result*, the relation between Lee winning the lottery and being less stressed will not be lost because *so* will take as its anaphoric argument the semantic interpretation of *Lee won the lottery*.

Similarly, in (11) if the annotator selects only the clause *John is a male American* as the argument of *as a result*, the relation between John living in the US and registering for the selective service and John being drafted can still be recovered. The empty connective signalling basic elaboration will link the first two arguments to *John is a male American* structurally; their relation to John being drafted will be an indirect one through the resultative relation of John being drafted and John being a male American.[17]

An additional complication that arises in the annotation of examples like (12) is the role of the lower-level syntactic annotation. In the Penn Treebank, from which the majority of our data is drawn, there is no principled parsing of such cases, in that it is left to the annotator to decide whether a particular use of a gerund, adjective, etc., should be parsed as a clause with missing elements when it is modified by an adverbial discourse connective. Therefore, we cannot reliably invoke the syntactic parse to decide when to label the left argument as a clause or an XP. We could, however, draw an analogy with coordinating conjunctions, which are commonly parsed with two XP arguments (e.g. *Sue is happy and tired*), although at the semantic level, two propositions are arguably being conjoined. If we allow the syntactic XP unit to represent a full proposition in the semantics, then we can invoke the minimal unit heuristic here

[17] Note that these same issues arise for a series of elaborations followed by *in additio*n, and in the same way a minimal unit can be selected, under the assumption that the remainder of the connections can be reconstructed through all the links between minimal units.

too. The annotator could be instructed to choose the smallest possible unit as the argument, in (12) the AdjP *overworked*, and then the full prepositional content of the argument, *Kim is overworked*, could be extracted from the sentential syntax and semantics. This would have the additional benefit of retaining parallelism in the surface syntactic form of the arguments of the connectives in such constructions.

Another potential heuristic in resolving the arguments of anaphoric connectives is their ability to combine with particular structural connectives, such as *but* and *and*. An auxiliary tree anchored with one of these connectives must be adjoined to its left-hand argument. Another connective, like *nevertheless, therefore,* or *moreover,* adjoined into this structure at the same point will frequently take the structural connective's left-hand argument as its own anaphoric argument (e.g. (13), where *and* and *therefore* share their lefthand argument).[18]

(13) He believed that <ARG> the Federal Security administrator had the authority and the responsibility for actions taken throughout the agency, </ARG> <CONN>and therefore he should be apprised of them and should play a part in the decisions.</CONN> (SSA)

A similar heuristic could be used for determining the size of the left-hand argument. In particular, when the right argument is a constituent smaller than a full clause (e.g. the second of two conjoined VPs), the left argument appears to consistently be the same size (e.g. the first of two conjoined VPs), as in (14).

(14) Jasper arrived late and therefore got no dinner.

An investigation of the variations in exact match labelling using the LOC tag and the individual labelling using the TYPE and COMB tags leads us to expect that if these heuristics are employed in the annotation, inter-annotator agreement will improve substantially. These minimal unit and connective combination cases, however, are distinguished from other issues that arise during the annotation of anaphoric arguments of discourse connectives, in that they are not cases of true ambiguity because principled heuristics can be introduced to resolve them. There are true cases of ambiguity, where such heuristics are not possible. One such case is discussed in the following section.

4.2 Ambiguity in Complement Clause

Cases of true ambiguity in identifying the left argument of a connective were found in connectives contained in complement clauses, mostly complements of verbs of saying. A connective in a complement clause may connect the complement clause with either the preceding sentence or with the main clause

[18] But not always, as the examples, like that in (3) above, which motivate the distinction between anaphoric and structural connectives, demonstrate.

containing the verb of saying. To illustrate the point, consider example (15). This example is ambiguous between analyses (15a) and (15b).

(15) Moritz said Monday that his leg feels fine and, as a result, he hopes to start practicing field goals this week.
  a. Moritz said Monday [that [his leg feels fine] [and, as a result, he hopes to start practicing field goals next week].]
  b. [Moritz said Monday that his leg feels fine] [and, as a result, he hopes to start practicing field goals this week.]

In (15a), both arguments of *as a result* are embedded under *said*. The left argument is the first complement clause and is annotated as SS (same sentence) because both it and the connective clause are the conjoined object of the matrix clause verb. In (15b), the clause containing the connective is a main clause by itself. On this interpretation, *as a result* was not part of what Moritz said but was added by the writer. More generally, connectives appearing after a complement clause are ambiguous with respect to whether they are part of the embedded complement clause immediately preceding them (i.e. it is their left argument) or whether they are conjoined to the main clause (i.e. this higher clause is their left argument.)

4.3 Low Attachment

As stated above, one reason we used the LOC tag in inter-annotator agreement was because the TYPE tag did not distinguish contiguous from non-contiguous arguments. This is an important distinction to make, because such arguments cannot be modelled structurally, thus indicating that they must be resolved anaphorically.

Because anaphoric connectives do not retrieve their left argument structurally, the clause containing them must attach to the prior discourse tree via a tree anchored by an empty structural connective. The DLTAG parser (Forbes *et al.*, 2003) currently employs the procedure of always adjoining this empty connective tree to the leaf node on the right frontier of the growing tree. If the anaphoric argument could be identified through a resolution mechanism, the parser could use this information to decide to instead attach this empty connective to the node on the right frontier which dominates the anaphoric argument. This would mean that the resolution of the argument would in a sense be captured in the discourse structure tree.[19] However, the anaphora resolution would have to precede the attachment decision here, so the structure cannot be thought of as in anyway determining the anaphora resolution.

[19] The precise identity of the anaphoric argument would potentially remain ambiguous depending on the level where the anaphoric connective and its right argument were attached because that node might dominate several discourse segments.

Moreover, examples like S_3 in the discourse in (3), show that this heuristic may very well not apply in cases where the anaphoric connective co-occurs with a lexical structural connective, rather than an empty connective.

5 Conclusions and Future Work

Discourse connectives are easily identified cues to discourse structure. But the actual discourse structure and relation that any particular connective indicates is not a fully-understood area of linguistic theory. By developing an annotated corpus of the discourse relations that individual connectives communicate through the anaphoric and structural connections they indicate, we can create an empirical picture of their behaviour which can be utilized in automatic detection of discourse structure.

We have reported the results of a preliminary corpus analysis of (primarily) anaphoric discourse connectives and the location and type of their arguments. The annotation provided information about whether particular connectives typically subcategorize for structural vs. anaphoric arguments. In addition, it provided detailed information about what the arguments look like and where they are found. This information will be useful for parsing discourse with a DLTAG. In addition, our results indicate that it will be possible to develop a resolution algorithm for identifying arguments that cannot be derived from the parse tree directly.

This study and the annotation guidelines developed as part of it are the starting point for a more extensive study which is creating a layer of annotations on top of both the Penn Treebank (syntactic) annotations and Prop Bank (semantic) annotations (Kingsbury & Palmer, 2002). Therefore, in the future we will be able to capture additional syntactic and also semantic properties of the sources of anaphoric arguments. These properties will be able to be automatically extracted from the annotated data. Additional annotation work on the discourse connective *instead* (Miltsakaki *et al.*, 2003) indicates that semantic properties of anaphoric arguments will be very useful for distinguishing them from non-arguments.

The annotation effort begun here, then, is a crucial first step towards automatic detection of the syntactic and semantic relations between propositions in discourse.

References

Carletta, J. 1996. “Assessing Agreement on Classification Tasks: The Kappa Statistic.” *Computational Linguistics* 22.249-254.

Forbes, K. 2003. “Discourse Semantics of S-modifying Adverbials.” Doctoral Dissertation, University of Pennsylvania.

----------, E. Miltsakaki, R. Prasad, A. Sarkar, B. Webber and A. Joshi 2003. “D-LTAG System – Discourse Parsing with a Lexicalized Tree-adjoining Grammar.” *Journal of Logic, Language and Information* 12.261-279.

Heim, I. and A. Kratzer 1998. *Semantics in Generative Gramma*r. Oxford: Blackwell.

Joshi, A., L. Levy and M. Takahashi 1975. “Tree Adjunct Grammars.” *Journal of the Computer and System Sciences* 10.136-163.

Kingsbury, P. and M. Palmer 2002. “From Treebank to Propbank.” *3rd Int. Conference on Language Resources and Evaluation (LREC-02*). Las Palmas.

Kruijff-Korbayová, I. and B. Webber 2001. “Information Structure and the Semantics of ‘Otherwise’.” In *ESSLLI’2001 Workshop on Information Structure, Discourse Structure and Discourse Semantic*s, 61-78. Helsinki.

Miltsakaki, E., C. Creswell, K. Forbes, A. Joshi and B. Webber 2003. “Anaphoric Arguments of Discourse Connectives: Semantic Properties of Antecedents versus Non-antecedents.” *Proc. of the Computational Treatment of Anaphora Workshop, EACL 200*3. Budapest.

----------, R. Prasad, A. Joshi and B. Webber. 2004. “The Penn Discourse Treebank.” In *Proceedings of the Language Resources and Evaluation Conference*. Lisbon, Portugal.

Webber, B., A. Knott and A. Joshi 2000. “Multiple Discourse Connectives in a Lexicalized Grammar for Discourse.” R Muskens, H. Bunt and E. Thijsse (eds.) *Computing Meanin*g, vol. 2 of *Studies in Linguistics and Philosophy Serie*s, 229-245. Dordrecht: Kluwer Academic Publishers.

----------, M. Stone, A. Joshi, and A. Knott. 2003. “Anaphora and Discourse Structure.” Computational Linguistics 29.545-587.

Annex

	***In addition* inter-annotation**					***As a result* inter-annotation**					***Nevertheless* inter-annotation**				
Token	**Ann.**$_e$	**Ann.**$_k$	**Ann.**$_c$	**Ann.**$_r$	**Agreement**	**Ann.**$_e$	**Ann.**$_k$	**Ann.**$_c$	**Ann.**$_r$	**Agreement**	**Ann.**$_e$	**Ann.**$_k$	**Ann.**$_c$	**Ann.**$_r$	**Agreement**
1	SS	NC	SS	SS	3/4	PS	PS	PS	PS	4/4	SS	SS	SS	SS	4/4
2	SS	SS	SS	SS	4/4	PS	SS	SS	SS	3/4	SS	SS	SS	SS	4/4
3	SS	SS	SS	SS	4/4	PS	PS	PS	PS	4/4	SS	SS	SS	SS	4/4
4	PS	PS	PS	PS	4/4	SS	SS	SS	SS	4/4	PS	PS	PS	PS	4/4
5	SS	SS	SS	SS	4/4	PP	PS	PS	PP	<2,2>/4	PS	PS	PS	PS	4/4
6	SS	SS	SS	SS	4/4	PP	PP	PP	PP	4/4	PS	NC	NC	PS	<2,2>/4
7	PS	PS	PS	PP	3/4	PP	PP	PP	PP	4/4	PS	PP	PS	PS	3/4
8	PS	PS	PS	PS	4/4	PP	PP	PP	PP	4/4	PP	PP	PP	PP	4/4
9	SS	SS	SS	SS	4/4	PS	PS	PS	PS	4/4	PS	PS	PS	PS	4/4
10	SS	SS	SS	SS	4/4	PS	PS	PS	PS	4/4	SS	SS	SS	SS	4/4
11	PS	NC	NC	NC	3/4	PS	PS	PS	PS	4/4	SS	SS	SS	SS	4/4
12	PS	PS	PS	PS	4/4	PS	PS	PS	PS	4/4	SS	SS	SS	SS	4/4
13	PS	PS	PS	PS	4/4	PS	PS	PS	PS	4/4	PP	PP	PP	PS	3/4
14	PS	PP	PP	PP	3/4	SS	SS	SS	PS	3/4	PP	PP	PS	PS	<2,2>/4
15	PS	PS	PS	PS	4/4	PS	PS	PS	PS	4/4	PP	PS	PP	PP	3/4
16	PS	PS	PS	PS	4/4	PS	PS	PS	PS	4/4	PP	PS	PP	PP	3/4
17	PS	PS	PS	PS	4/4	PS	PS	PS	PS	4/4	PS	PP	PP	PP	3/4
18	PS	PP	PP	PS	<2,2>/4	PS	PS	PS	PS	4/4	SS	SS	SS	SS	4/4
19	PS	PS	PS	PS	4/4	PS	PS	PS	PS	4/4	PS	PP	PP	PP	3/4
20	PS	PS	PS	PS	4/4	PP	PP	PP	PP	4/4	PS	PS	PS	PP	3/4
21	PS	PS	PS	PS	4/4	PS	PS	PS	PS	4/4	PP	PP	PP	PP	4/4
22	PS	PS	PS	PS	4/4	PP	PS	PP	PP	3/4	PS	PP	PP	PP	3/4
23	PS	PP	PP	PP	3/4	PS	PS	PS	PS	4/4	PS	PS	PP	PP	<2,2>/4
24	PP	PP	PP	PP	4/4	PP	PS	PS	PS	3/4	PS	PS	PP	PS	3/4
25	NC	PP	PS	NC	2/4	PS	PS	PS	PS	4/4	PS	PS	PS	PS	4/4

Combining Centering-Based Models of Salience and Information Structure for Resolving Intersentential Pronominal Anaphora

Costanza Navarretta
Centre for Language Technology, Copenhagen University

In many computational approaches for resolving intersentential pronominal anaphora, the degree of salience of entities is identified by their degree of givenness in the addressee's discourse model, so that *given* (i.e. known, familiar) entities are assigned the highest degree of salience. The most salient entities are chosen as antecedents of pronouns. Centering-based resolution models also adopt this approach (Grosz *et al.*, 1995). A different point of view is taken by Hajičová *et al.* (1990) who assume that discourse elements in the focal part of an utterance in Information Structure terms have the highest degree of salience. These elements often correspond to new information. Analysing Danish discourse we found that these apparently contrasting interpretations of salience are both valid, but in different contexts. We propose a unified approach combining Centering-based models of salience with Hajičová *et al.*'s proposal.

1 Introduction

Cognitive-based theories on the use of referring expressions presuppose that the speaker makes some assumptions about the status of entities in the addressee's mental state and that these assumptions influence her/his choice of referring expressions, i.a. (Prince, 1981; Ariel, 1994; Gundel *et al.*, 1993; Givón, 1976, 1979, 1983). Although the theories focus on different aspects, they all conclude that pronominal anaphors refer to those entities in discourse which are most easily accessible because they are more *given* (known) in the addressee's discourse model. The different theories classify these entities as being *familiar* (Prince, 1981, 1992), *in focus* (Gundel *et al.*, 1993, 2001), *topic prominent* (Givón, 1979) or *accessible* (Ariel, 1994).

The fact that pronouns usually refer to the most accessible elements in discourse is also presupposed by computational approaches for resolving intersentential pronominal anaphora.[1] In these approaches, the degree of accessibility of elements in the addressee's discourse model is connected to their degree of salience, so that the most salient discourse elements are also the most easily accessible. Inspired by various cognitive-based theories, computational approaches use different models of salience but, as we will discuss in the paper, the majority of models identify high degree of salience

[1] In this paper we only look at entities introduced in discourse by nominals and by pronominal antecedents, we simply mean the nominals in the preceding utterance which corefer with the intersentential anaphors.

with high degree of *givenness*, where *given* means *known*. This is also the case for the various algorithms which have been proposed inside the popular Centering framework (Grosz *et al.*, 1995), which we choose as examples of givenness-based models of salience.

Analysing the use of intersentential pronominal anaphora in Danish written and spoken discourse, however, we found a number of cases, not involving grammatical parallelism, in which nominals providing new information were the preferred antecedents of pronominal anaphors, although they competed with known candidate antecedents. Many of these antecedents providing new information occurred in specific types of syntactic construction and in these constructions the preferred pronominal antecedents were more frequently nominals providing new information than "known" candidate antecedents. The percentage of this phenomenon in the analysed Danish texts and dialogues is given in Section 4. An example of a "not given" nominal being the pronominal antecedent instead of the more known antecedent candidate is given in (1).

(1) [Dommeren i Hørsholm]$_i$ fængslede i onsdags [en 28-årig mand fra Århus]$_k$. [Han]$_k$ sigtes for sammen med [en 44-årig, der blev fængslet før påske,]$_j$ at være gået ind på samlingen og uantastet at have taget de to billeder ned fra væggen.
"[The judge on duty in Hørsholm]$_i$ arrested last Wednesday [a 28-year old man from Århus]$_k$. [He]$_k$ is charged for, together with [a 44-year old man, who was arrested before Easter,]$_j$ having entered the gallery and unchallenged having taken the two pictures from the wall." [BERLINGSKE]

In the example the antecedent of the pronoun *han* (he) is the indefinite object *en 28-årig mand fra Århus* (a 28-year old man from Århus) and not the subject definite candidate *dommervagten i Hørsholm* (the judge on duty in Hørsholm). The latter nominal is considered to be the most given and then also the most accessible element in most models of salience. The only model which assigns the highest degree of salience to elements which are not the most given ones is the one proposed by Hajičová & Vrbová (1982) and Hajičová *et al.* (1990). In this model, Hajičová *et al.* presuppose that elements in the focal part of utterances in Information Structure terms have the highest degree of salience. These focal elements represent, in most cases, new information. Hajičová *et al.*'s proposal is very interesting because it can account for data like the one presented in example (1). However, it is also problematic especially because it is not always true that nominals in the focal part of an utterance are the most salient elements.

We believe that the two apparently contrasting models of salience proposed by Grosz *et al.* (1995) and Hajičová *et al.* (1990) are both valid, but in different contexts. In this paper we present a novel approach combining the two models. Although in this paper we focus on the degree of givenness of entities in discourse as a measure of their degree of salience, we do not state that givenness is the only factor influencing salience, neither that salience is the only aspect to be considered when resolving pronominal anaphora. This is discussed further in Sections 5 and 7.

The paper is organised as follows. In Section 2, we describe how the degree of salience of discourse elements is modelled in various Centering-based approaches and how all these models assign the highest degree of salience to given elements. In Section 3, we present Hajičová *et al.*'s (1990) model of salience and discuss why it is problematic from an applied point of view.

In Section 4, we look at some examples of Danish discourse containing anaphors whose antecedents are not the most given candidate elements, and discuss how different Centering-based models resolve these anaphors. Finally, we present the results of an analysis of pronominal antecedents in Danish written and spoken corpora. In Section 5, we propose our approach combining a Centering-based model of salience with Hajičová's proposal and describe the results of a survey of the uses of pronominal anaphors which confirm our proposal. In Section 6, we shortly outline other pronominal anaphora resolution approaches which assign high prominence to some types of focal information. Finally, in Section 7, we give a summary of the paper and make some concluding remarks.

2 Modelling Salience in Centering

The Centering theory (Grosz *et al.*, 1995) has been quite influential because of its simplicity and because some of its basic assumptions are quite intuitive, are confirmed by cognitive studies of pronominal anaphors and can account for many anaphoric occurrences as shown by a number of tests on more languages, i.a. (Strube, 1998; Tetreault, 2001). The Centering theory assumes that discourse tends to be "about" few salient entities at a time, the so-called *centres*, and that intersentential pronominal anaphors often refer to the most salient of these centres. The theory presupposes Grosz and Sidner's (1986) three-level discourse model according to which the intentions behind discourse allow to divide discourse in discourse segments which exhibit global coherence. Centering models local reference, i.e. entities inside a discourse segment.

In the original formulation of Centering, Grosz *et al.* (1995) assign to an utterance U_n a set of forward-looking centres, $C_f(U_n)$, corresponding to the entities which can be referred to in the following utterance. The elements in $C_f(U_n)$ are partially ordered according to their prominence (salience). The highest ranked element in $C_f(U_n)$ is called the *preferred centre* $C_p(U_n)$, following Brennan *et al.* (1987). The highest ranked element in $C_f(U_n)$ which was also realised in the preceding utterance U_{n-1} is called the backward-looking centre *($C_b(U_n)$)*.

If one of the elements in $C_f(U_{n-1})$ is realised by a pronoun in U_n, then the $C_b(U_n)$ must also be realised by a pronoun.

Grosz *et al.* also assume that, inside a discourse segment, the addressees perceive utterances in which the speakers continue speaking about the same entities as more coherent than utterances in which speakers change the focus of attention. This assumption is implemented by the ranking of center transition states between pairs of utterances. In (Grosz *et al.*, 1995), the highest ranked transition is *centre continuation*. In centre continuation the backward-looking centre is the preferred centre in U_n and co-refers with the backward-looking centre in U_{n-1}, i.e. $C_b(U_n)=C_b(U_{n-1})=C_p(U_n)$. *Centre retainment*, $C_b(U_n)=C_b(U_{n-1})\neq C_p(U_n)$ is ranked less than *centre continuation* in the transition ranking hierarchy, but it precedes *centre shifting*, where the C_b in two adjacent utterances are not the same. The transition state hierarchy is illustrated below:

continue > retain > centre

Figure 1: Hierarchy of transition states

Brennan *et al.* (1987) extend the hierarchy of transition states, but they still assume that centre continuation and centre retainment are more coherent than centre shifting states. In our opinion, the assumption that continuing speaking about the same elements is perceived as more coherent than shifting centre of attention is not very intuitive. It is true that discourse tends to be about some few entities at a time, but it is also natural that speakers change the focus of attention. We will show later in this paper that this shift is often as coherent as centre continuation, because it is announced to the addressee.

Grosz *et al.* (1995) recognise that many factors contribute to the ordering of the forward-looking centres, but for practical reasons all Centering-based algorithms use simple models of salience. Grosz *et al.* rank elements according to their order of occurrence in the utterance. In English, this order often corresponds to the hierarchy of grammatical roles. This hierarchy has been proposed as the preferred syntactic structure for describing the topics in discourse, i.e. the elements discourse “is mainly about”, see i.a. Givón’s (1979) Topicality hierarchy.

Brennan *et al.* (1987) and Kameyama (1998) adopt the hierarchy of grammatical roles to rank forward-looking centres in their Centering-based algorithms.[2] The hierarchy of grammatical roles used by Brennan *et al.* is illustrated in Figure 2.

subject > first-object > second-object > other complements > adjuncts

Figure 2: Brennan et al.'s hierarchy of grammatical roles

Strube and Hahn (1996) present a so-called functional Centering model for ordering discourse elements according to their degree of salience. They use the information structure terms proposed by Daneš (1968) who distinguishes between given information, *theme*, i.e. the already known information that the discourse is mainly about, and *rheme* or new information, i.e. information that has just been introduced in discourse.

In Strube and Hahn's interpretation of Centering, the $C_b(U_n)$ corresponds to given information, while the highest ranked element in U_n, the $C_p(U_n)$, is the theme of the utterance. The *theme/rheme* hierarchy is determined by the elements in U_n and U_{n-1}. Elements which are contained in both $C_f(U_{n-1})$ and in $C_f(U_n)$ are thematic and Strube and Hahn call them *bound elements*. Bound elements are ranked higher than rhematic, or *unbound elements*, i.e. elements that are in $C_f(U_n)$ but not in $C_f(U_{n-1})$. Strube and Hahn also propose a ranking order for the various types of bound element, while they rank elements belonging to the same type according to their order of occurrence in the utterance, so the leftmost elements have the highest prominence. The three-levels of ranking in Strube and Hahn's model are given in Figure 3.

bound elements > unbound elements
anaphora > (possessive pronoun *xor* elliptical antecedent)> (elliptical expression *xor* head of anaphoric expression)
nom head 1 > nom head 2 > . . .> nom head n

Figure 3: Ranking of information structure patterns

In all three ranking levels, the elements introduced in discourse earlier, and thus with a higher degree of givenness, are ranked higher than those elements which have just been introduced in it. Strube and Hahn's approach has the advantage of extending the Centering framework to free word-order languages, such as German, where the order of discourse elements does not correspond to their grammatical role in utterances. Although Strube and Hahn assume an information structure based model for measuring salience, they do not consider the possibility of unbound elements having higher degree of salience than bound elements.

---

[2] Kameyama's (1996) model is much more complex than that proposed by Grosz *et al.* (1995) and Brennan *et al.* (1987). Kameyama distinguishes, among other things, an input and an output attentional state. In her model the hierarchy of grammatical roles is used to rank discourse elements in the output attentional state.

An approach similar to that proposed by Strube and Hahn is followed by Hoffman (1998). She investigates the pronominal anaphors in Turkish taking into account the information structure of sentences. Hoffmann concludes that in Turkish, the backward-looking centre preferentially co-refers with discourse elements in the topic part of utterances.

Another functional model of salience degree is used by Strube (1998). Strube's model is an operationalised version of the *Familiarity scale* proposed by Prince's (1981). In the *Familiarity scale*, Prince models to which degree information is assumed by the speaker to be known to the addressee, where known means familiar in the hearer's model:

Evoked > Unused > Noncontaining Inferable > Containing Inferable > Brand-New Anchored > Brand-New Unanchored

Figure 4: Prince's Familiarity scale

Strube's operationalised version of the model is shown below:

OLD (pronominal and nominal anaphors, previously mentioned proper names, relative pronouns, appositives, proper names, titles) **> MEDIATED** (inferables) **> NEW** (indefinites)

Figure 5: Strube's model

In Strube's model, discourse elements classified as OLD are ranked higher than those classified as MEDIATED and NEW. Two elements of the same type are reciprocally ranked according to their order of occurrence in the utterance, the element mentioned earliest being assigned the highest prominence. Therefore, also in Strube's model given (OLD) elements are always ranked higher than non-given (NEW) elements.

In conclusion, in all the centering-based models of salience we have discussed, the criterion for ranking elements in the utterance connects high degree of salience of discourse elements with high degree of givenness in the addressee's model. Because the different Centering-based models identify degree of givenness with different phenomena, they sometimes rank elements differently. However, none of these models takes into account the fact that speakers can mark as salient elements that do not have a high degree of givenness in the addressee's discourse model.[3] In the following section, we present Hajičová *et al.*'s model of salience where the highest degree of salience of entities is not necessarily connected with the highest degree of givenness.

[3] In our account we have only taken into consideration how Centering-based approaches model the degree of salience of entities. The various approaches also follow different resolution strategies and they do not cover exactly the same types of phenomenon. However, discussing these aspects is out of the scope of the present paper.

3 Hajičová's model of salience

Hajičová and Vrbová (1982) propose a model that assigns the highest degree of salience to discourse elements which represent new information. The model is operationalised in (Hajičová *et al.*, 1990).

Hajičová *et al.* presuppose the information structure of utterances proposed by Sgall *et al.* (1986). Sgall *et al.* recognise a topic/focus dichotomy in sentences, which they call the topic/focus articulation, **TFA**, of the sentences. The terms *topic* and *focus* are used with various meanings not only in different fields, but also in the information structure literature. In the following, we use them as in (Sgall *et al.*, 1986), where *topic* is assumed to correspond to given, known, bound information, the *theme* in (Daneš, 1968), while *focus* corresponds to new, unbound information the *rheme*, or as suggested by Vallduví and Engdahl (1995: 462) to "the informative, newsy, dominant, or contrary-to-expectation part" of an utterance. Thus, focus is used differently than in (Gundel *et al.*, 1993) where elements which are "in focus" are those elements in an utterance whose referents are at the current centre of attention (corresponding to the *focus* of attention in (Grosz & Sidner, 1986)). According to Sgall *et al.*'s (1986) context-unbound nominals are always part of the focus, while context-bound nominals are, in most cases, part of the topic. Elements in the topic or in the focal part of an utterance can be ordered according to their degree of dynamism in the sentence.

In the algorithm proposed by Hajičová *et al.* (1990), the degree of accessibility of elements in the addressee's discourse model is identified with their degree of salience and is implemented by weights. Differing from other salience models, Hajičová *et al.* assign the highest degree of salience to the nominal phrases referred to in the focal part of an utterance *U*, with these phrases being given the highest accessibility weight, $w{=}max$. The nominal phrases in the topic part of *U* are activated one degree less than those referred to in its focal part, i.e. $w{=}max\text{-}1$. A pronominal reference to an element in the topic part of *U* retains the degree of activation of the element in the discourse model.

The activation of elements not mentioned in *U* fades away and it fades away most quickly for those elements which had the highest activation in the model. Elements whose activation weight differs only by one compete as preferred antecedents of pronominal anaphors. This is exactly the case for antecedents in the focal and in the topic part of *U*.

The suggestion that elements in the focal part of an utterance, which often correspond to new information, have the highest degree of salience distinguishes Hajičová *et al.*'s approach from other salience models. Hajičová *et al.*'s approach is original and can account for some types of anaphor, such as the one discussed in Section 1, which are often incorrectly resolved by givenness-based salience models. However, Hajičová *et al.*'s proposal is problematic from an applied point of view. In the first place, it is difficult to determine the TFA of all utterances. Secondly, focal candidate antecedents are ranked highest in Hajičová *et al.*'s model, but they compete with topic candidate antecedents in their resolution system, because their accessibility weights differ only by one. Finally, the data does not confirm that all entities referred to in the focal part of an utterance have the highest degree of accessibility. On the contrary, the analysis of the Danish data indicates that there are only a restricted number of constructions where the entities referred to in the focal part of an utterance have so high degree of salience that they should be proposed as the preferred antecedents in applied systems. We discuss these empirical data in the following section.

4 Danish Data

We have analysed the occurrences of intersentential pronominal anaphora in Danish texts and transcriptions of naturally occurring conversations. The texts are taken from newspaper collections, Berlingske Tidende 1992 and 1999, henceforth BERLINGSKE, a collection of computer manuals and novels. The analysed dialogues belong to the BYSOC corpus, collected under "Projekt Bysociolingvistik" (Project Urban Sociolinguistics) (Gregersen & Pedersen, 1991; Henrichsen, 1998), to the PID corpus collected under "Projekt Indvandrerdansk" (Project Immigrant Danish) (Jensen, 1989) and to the corpus "Samtale hos Lægen" ("Talking with the doctor") (Duncker & Hermann, 1996), henceforth SL. The BYSOC and the PID collections consist of the recordings of everyday conversations. The SL collection consists of the recording of 43 dialogues between Danish adult patients and their GPs. In the texts and dialogues, we found a number of intersentential pronominal anaphors with more candidate antecedents where the antecedent chosen by two humans, also on the basis of the context, is the least given nominal according to givenness-based salience ranking. We only considered examples occurring inside discourse segments. In the texts, discourse segments were identified by paragraphs as in i.a. (Strube, 1998; Tetreault, 2001) while in the dialogues they were manually marked mainly on the basis of the context and cue words. As indicated in Section 1, we did not consider cases involving grammatical parallelism, according to which in adjacent utterances with parallel grammatical complements, the preferred antecedent of an anaphor in the second utterance is the linguistic expression in the first utterance with the same grammatical function. Parallelism has been discussed in i.a. (Asher, 1993; Kameyama, 1996; Kehler, 2000). We discuss examples of parallelism in Section 5.

In (2), examples from the Danish corpora of pronominal antecedents presenting new information are given.

(2) a. *[Chefen]$_i$ fik kun [en søn]$_k$ og [han]$_k$ gad i hvert fald ikke*
[Boss-defin]$_i$ got only [one son]$_k$ and [he]$_k$ wanted surely not
videreføre familieforetagendet.
carry on familybusiness-defin.
"[The boss]$_i$ got only [one son]$_k$ and [he]$_k$ surely did not want to carry on the family business". [SL]

b. *Med [Peter]$_I$ sad der altid [en enkelt mand fra "den lokale" I*
With [Peter]$_I$ sat there always [one man from "the local pub" in
Flensburg]$_k$ og [han]$_k$ var aldrig med til udekampene.
Flensburg]$_k$, and [he]$_k$ came never to away matches-defin.
"There was always sitting [one man from "the local pub" in Flensburg]$_k$ with [Peter]$_{i,}$ and [he]$_k$ never came to the away matches." [BERLINGSKE]

c. *[Igor]$_I$ talte med [en mand]$_k$ udenfor Irma. [Han]$_k$ var stor og havde uredt hår.*
"[Igor]$_I$ spoke with [a man]$_k$ outside Irma. [He]$_k$ was big and had ruffled hair."
[BERLINGSKE]

d. *Og så var der [patient-chaufføren Duddi]$_{i,}$ [der]$_i$ kørte*
And then was there [patient-chauffeur-defin Duddi]$_i$, [who]$_i$ drove
[en mand]$_k$ hjem fra sygehuset. [Han]$_k$ havde været indlagt,
[a man]$_k$ home from hospital-defin. [He]k had been hospitalised,
for [han]$_k$ fik [sin]$_k$ fod i plæneklipperen.
because [he]$_k$ got [his]$_k$ foot in lawn mower-defin.
"And then there was [the patient-chauffeur Duddi]$_i$, [who]$_i$ drove [a man]$_k$ home from the hospital. [He]$_k$ had been hospitalised, because [he]$_k$ had got a foot in the lawn mower." [BERLINGSKE]

e. speaker 1: *hvem hvem arbejdede [din mor]$_i$ med*
whom... whom worked [your mother]$_i$ with
"with whom... whom did [your mother]$_i$ work"
speaker 2: *[Hun]$_i$ arbejdede med [vores nabo]$_k$*
"[She]$_i$ worked with [our neighbour]$_k$"
[Hun]$_k$ var enke ... havde tre sønner
"[She]$_k$ was a widow... had three sons" [BYSOC]

In example (2a), the antecedent of the pronominal anaphor *han* (he) is the indefinite nominal *en søn* (one son), the object of the preceding utterance. The second candidate antecedent is the definite nominal *chefen* (the boss), the subject of the utterance. All centering-based approaches, discussed in Section 2, prefer the definite subject *chefen* (the boss) as antecedent. Grosz *et al.*, (1995) rank it highest because it is the first occurring candidate antecedent in U_{n-1}. Brennan *et al.*, (1987) and Kameyama (1996) rank *chefen* highest because it is the subject. In Strube and Hahn's (1996) model, context-bound elements are preferred to context-unbound ones as anaphoric antecedents. The definite *chefen* (the boss) is bound while the indefinite *en søn* (a son) is not. Similarly, in Strube's model definite nominals are classified as OLD information, which are ranked higher than indefinite nominals, classified as NEW.

In example (2b), the antecedent of the pronoun *han* (he) is the indefinite nominal *en enkelt mand fra "den lokale" i Flensburg* (a man from "the local pub" in Flensburg) and not the more given proper noun *Peter*. Both Kameyama's and Brennan *et al.*'s models indicate the indefinite nominal as the antecedent because it is the subject of the utterance. In Grosz *et al.*'s model, the topic-fronted nominal[4] *Peter* is chosen as antecedent because it occurs in the utterance before the competing subject nominal. In the models proposed by Strube and Hahn and by Strube, proper names are ranked higher than indefinite nominals and therefore *Peter* is chosen as the pronominal antecedent. Concluding the discussed Centering-based models resolve the anaphor in example (2b) in different ways.

In example (2c), all Centering-based models rank the subject proper name *Igor* higher than the indefinite object *en mand* (a man).

In example (2d), the antecedents of the pronoun *han* (he) is the indefinite object *en mand* (a man) and not the subject relative pronoun *der* (that) co-referring with the nominal *patientchafføren Duddi* (the patient chauffeur Duddi).[5] All Centering-based algorithms rank the subject relative pronoun highest.

In the last example, (2e), the antecedent of the second occurrence of the pronoun *hun* (she) is *vores nabo* (our neighbour), the object in the preceding utterance. Instead, all Centering based models choose as antecedent the entity referred to by the first occurrence of the pronoun *hun*, which co-refers with the nominal *din mor* (your mother). In conclusion, in all the examples in (2), the antecedents of pronominal anaphors are less "given" than the competing candidate antecedents in the utterances. All these less given antecedents occur in the focal part of the utterances. In the first four examples, (2a)-(2d), the focal antecedents are context-unbound elements. In example (2a), the indefinite nominal *en søn* (a son) follows the rhematiser or focusing adverbial *kun* (only) (Quirk *et al.*, 1985). In example (2b), the indefinite nominal *en enkelt mand fra "den lokale" i Flensburg* (a man from "the local pub" in Flensburg) occurs in an existential construction. In examples (2c) and (2d), the focal entity is the

4 We propose that topic-fronted nominals – usually not prosodically marked – should be distinguished from topicalised entities, which are prosodically marked. Only the latter are focal elements. This distinction is noticed for Swedish by Vallduví and Engdahl (1995) and, in our opinion, is also valid in Danish.

5 In this example we presuppose that the Centering algorithms are applied to intrasentential clauses as proposed by (Kameyama, 1998). This is not necessary for Strube's (1998) algorithm which applies to both intrasentential and intersentential anaphors.

indefinite nominal object, *en mand* (a man). In example (2e), the antecedent *vores nabo* (our neighbour) is context-bound, but is less "given" than the competing personal pronoun *hun* (she). Furthermore, the nominal *vores nabo* also presents new information, i.e. provides the information which was asked for in the preceding question and is thus the focus of the utterance. Usually focal information is also prosodically marked in spoken language, see i.a. (Sgall *et al.*, 1986; Vallduví & Engdahl, 1995).

The anaphors in (2) can be accounted for by Hajičová and Vrbová's (1982) and Hajičová *et al.*'s (1990) model which assigns the highest degree of salience to elements in the focal part of an utterance. However, not all elements in the focal part of an utterance are the antecedents of pronominal anaphora. The data indicates that only in specific contexts, nominals in the focal part of an utterance have so high degree of salience that they should be chosen as the preferred anaphoric antecedents. More precisely, the majority of the elements in the focal part of an utterance which are the antecedents of pronominal anaphors in ambiguous contexts, i.e. in contexts with more given candidate antecedents, occurred in a restricted number of construction types in our data. Most of these constructions have also been recognised as focus-marking in the English and/or Danish information structure literature i.a. (Vallduví & Engdahl, 1995; Togeby, 1993; Paggio, 1997; Kruijff, 2001). They comprise *there*-constructions,[6] topicalised constructions, clefts, nominals providing information asked for in the preceding question, nominals preceded by a focusing adverbial, nominal indefinite objects occurring in particular positions in the utterances. In Table 1, the percentage of focal antecedents preferred to more given antecedents in each type of the above constructions in the analysed texts and dialogues is indicated.[7]

Focal antecedents preference

Construction type	texts	Dialogues
There-constructions	98%	97%
Clefts	100%	100%
topicalised constructions	86%	91%
constructions with focusing adverbs	100%	97%
context-marked focus in question/answer	--	100%
constructions with indefinite nominal objects in object-nominal position	59%	65%

Table 1: Focal preference in focal-marked constructions

[6] *Der*-constructions in Danish.

[7] There were two examples in the data where a topicalised nominal preceded a *there*-construction. In one case, the pronominal antecedents were the topicalised nominals; in the other, the indefinite nominals in the *there*-construction.

As the values in the table indicate, focality preference is not equally stronger in all types of construction and is weakest in the case of indefinite nominal objects in object-nominal position.

In the following section, we propose an account of the anaphors in these constructions combining centering with focal preferences.

5 Our Proposal

On the basis of the analysed data, we propose that nominal elements identified as focal in the particular constructions presented in Section 4 not only belong to the focal part of an utterance, but are the main focus of it. The main focus of an utterance may correspond to what Sgall *et al.* (1986) call *focus proper*, i.e. the most dynamic element in the focal part of an utterance, "carrying the intonation centre" (Sgall *et al.*, 1986:178). The focus proper can be considered the opposite of what Sgall *et al.* call the *topic proper*, i.e. the less dynamic element in the topic part of the utterance.

In our opinion, only nominals which are the *focus proper* in an utterance have the highest degree of salience. On the basis of the analysed data we propose the following tentative list of entities which can usefully be ranked as the most prominent candidate antecedents of pronominal anaphors.

1. Entities referred to by nominals which are focally marked structurally. In Danish, structural marking of focus occurs in clefts, existential and topicalised utterances. The preferred antecedent in example (2b) is the indefinite nominal in an existential construction.
2. Entities referred to by nominals that follow focusing adverbs. These adverbs, include additives such as *også* (also) and restrictives such as *kun* (only). The antecedent in example (2a) belongs to this group.
3. Entities focally marked by prosodic marking and/or by the context.[8] This is the case in question/answer pairs as in example (2e). In this example the focus proper, *vores nabo* (our neighbour), provides the information asked for in the preceding question. These types of anaphoric antecedents are quite frequent in our dialogues.
4. Objects which have just been introduced in the discourse by indefinite nominals and which occur in the object-nominal position (Togeby, 1993).[9] Examples (2c) and (2d) belong to this group.

It is relatively easy to recognise the majority of these constructions in Danish. We propose that the accessibility of discourse elements is by default connected with the concept of givenness as assumed by the Centering theory.

---

[8] It should be noticed that in the dialogue transcriptions used above prosodic information was only available in some cases.

[9] The Danish word order has been described using the so-called *Feltskema* (Field schema) proposed by Diderichsen (1957; (1946)).

This is also the case for entities referred to in the focal part of utterances, which are not the focus proper. The accessibility of given elements, however, is overridden by the accessibility of elements which are the focus proper in the utterances they occur in. In a few cases, the focus proper is indicated by the context, while in the majority of cases, the speakers explicitly change the degree of accessibility of elements in utterances by marking them as salient with information structure related devices. These devices, in Danish as in many other languages, comprise word order, prosodic marking and syntax.

Our tentative list of focus-marking constructions is mainly based on empirical data and it is confirmed by the information structure literature. However, it is not always clear in texts and dialogues whether the focal-marked antecedents are chosen because of their salience or because of the context. To verify our hypothesis that main foci have the highest degree of accessibility we conducted a survey of the use of intersentential pronominal anaphora. In the survey, we isolated groups of preference types in constructed examples. Some of these examples are discussed in Section 5.1. Our work is inspired by Kameyama (1996) who in a survey of English pronouns studies how Centering-based preferences interact with parallelism and common sense knowledge.

5.1 Verifying Our Hypothesis

In our survey of the use of Danish pronouns, we asked 32 native speakers of Danish, the *informants* henceforth, to choose the preferred antecedents in a number of constructed examples. Less than half of the informants were linguists. If the informants could not choose a preferred antecedent, they had to signal this impossibility. Most of the constructed examples are variations of utterances found in our texts or dialogues.

In the survey, among other things, we investigated the relation between givenness and focality preferences in examples where a pronominal anaphor has two competing candidate antecedents, one being the focus proper, the other being a nominal which is more given according to givenness-based definitions of salience. In particular, we considered cases where the focus proper is a NEW entity (an indefinite nominal) and the competing antecedent is an OLD entity (a proper name or a definite nominal), according to the Familiarity scale proposed by Prince (1981) and implemented in Strube's centering algorithm (Strube, 1998). The reason to focus on these cases was that, as indicated by Table 1, focality preference is less strong in cases where a given subject nominal competes with an indefinite object nominal in object-nominal position. We also investigated (i) cases where the NEW focus proper and the OLD candidate antecedents have different syntactic roles and/or occur in different positions and (ii) the relation between givenness, focality and parallelism. In the survey, we also examined other factors influencing anaphora resolution, such as animacy, recency and lexical knowledge. We do not discuss these aspects in this paper. World knowledge and conventional presuppositions are of course the strongest preference of all, but we attempted to minimize their influence in our examples by constructing as "neutral" and context-isolated utterances as possible.[10]

The survey examples relevant to this paper are listed in the following.

A. *Der sad en mand ved siden af Peter i toget Han så træt ud.*
There sat a man next to Peter on train-defin. He looked tired.
"A man sat next to Peter on the train. He looked tired."

B. *Peter snakkede med en gammel mand i toget. Han så meget sur ud.*
"Peter talked with an old man on the train. He looked very angry."

C. 1. speaker:
Hvem mødte Peter på gaden i går?
"Whom did Peter meet in the street yesterday?"
2. speaker:
Peter mødte Søren. Han havde travlt.
"Peter met Søren. He was busy."

D. *En journalist genkendte forsvarsministeren. Han begyndte at løbe.*
"A journalist recognised the minister of defence. He began to run."

E. *Forsvarsministeren blev genkendt af en journalist. Han var meget overrasket.*
"The minister of defence was recognised by a journalist. He was very surprised."

F. *Peter mødte Søren på gaden. Han hilste på ham.*
"Peter met Søren in the street. He greeted him."

G. *Peter så en mand på gaden. Han råbte til ham.*
"Peter saw a man in the street. He shouted at him."

H. *Peter mødte en mand på gaden. Han hilste på ham.*
"Peter met a man in the street. He greeted him."

I. *Peter mødte en mand i toget. Maria hilste på ham.*
"Peter met a man on the train. Maria greeted him."

In example **A**, the focus proper of the first utterance is the indefinite subject nominal *en mand* (a man) in the *there*-construction. Strube and Hahn's and Strube's algorithm chooses the proper name *Peter* as antecedent of the pronoun *han* (he). Brennan *et al.*'s, Kameyama's and Grosz *et al.*'s algorithms choose *en mand* (a man) as antecedent.

In **B**, the focus proper is the prepositional object *en gammel mand* (an old man), while the given candidate antecedent is the subject *Peter*, which is proposed as antecedent by all Centering algorithms.

In **C**, both candidate antecedents of the pronoun *han* (he) are proper names, thus OLD elements, but the focus proper, *Søren*, occurs after the subject antecedent *Peter*, which is chosen as antecedent in all Centering algorithms.

[10] The whole survey is described in Navarretta (2002).

In example **D**, as in **A**, the subject is an indefinite nominal, but in **D** this subject does not occur in a typical focal position. Brennan *et al.*'s, Kameyama's and Grosz *et al.*'s algorithms choose the indefinite subject nominal *en journalist* (a journalist) as antecedent of the pronoun *han* (he), while in Strube and Hahn's and in Strube's algorithm the preferred antecedent is the definite object *forsvarsministeren* (the ministry of defence).

The first utterance of example **E** contains a passive construction, thus the syntactic subject is not the agent. The agent is expressed by an indefinite nominal. The subject is a definite nominal. Passivisation alters the information structure of the active clause. In all the considered algorithms, the syntactic definite subject *forsvarsministeren* (the minister of defence) is chosen as the antecedent of the pronoun *han* (he).

In examples **F-I**, the relation between focality preference and parallelism is investigated.

In example **F**, the two candidate antecedents are both proper names. The first utterance in **F** has the same semantic content as the second utterance in **C**, but the topic/focus articulation of the two utterances is quite different. In examples **G-H**, there are two pronominal anaphors in parallel position to the two candidate antecedents in the preceding utterance. One candidate is the focus, the other candidate is a more given entities as in examples **A** and **B**. In example **I**, we investigate a case where parallelism competes with subject-antecedent preference which, according to the results presented in (Kameyama, 1996), can overrule parallelism.

All the Centering-based algorithms choose the subject proper nominal *Peter* as antecedent of the first pronoun *han* (he) in examples **G-H**, and of the pronoun *ham* (him) in example **I**, because it is the subject, precedes the other nominal or is the most given candidate antecedent.

5.2 Results of the Survey

The results of the survey are shown in Table 2. In the table the number of informants that chose each candidate antecedent in the examples is shown. The sign '*?*' indicates that the informants could not choose a preferred interpretation. In the last two columns of the table, we give the $\chi^2_{df=1}$ significance and the level of preference *p* for each example. The $\chi^2_{df=1}$ significance is computed by adding an evenly divided number of the answer "unclear" (in Table 2 indicated by "?") to each explicitly selected answer.

Significance is calculated with Pearson's correlation coefficient, see (Woods *et al.*, 1986). Being a two-sided test we have doubled the *p*-value, before calculating significance. Preference is considered significant if $p < .05$, weakly significant if $.05 < p < .10$ and insignificant if $.10 < p$.

The interpretations of examples **A-E** confirm that foci have the highest degree of salience. Kameyama (1988) consider empathy as a salience factor explaining the pronominal antecedents in Japanese utterances similar to examples **A** and **B**. Empathy is defined in Kuno (1987: 206) as “the speaker’s identification, which may vary in degree, with a person/thing that participates in the event or state that he describes in a sentence”. It is possible that the focal nominals in examples **A** and **B** can also be accounted for as cases of empathy, but they are still focal entities. Furthermore, it is not possible to determine cases of empathy in Danish without a deep analysis of discourse. In the paper we have exclusively focused on phenomena which can be identified on the basis of syntactic phenomena.

In example **D**, there is no significant preference for any antecedent. The antecedent subject *en* journalist (a journalist) is an indefinite nominal, thus a newly introduced entity, but it is not the focus proper, as it competes in salience degree with the *given* object *forsvarsministeren* (the minister of defence).

	Answers						Preference	
	1. antecedent		2. antecedent		unclear		$\chi^2_{df=1}$	*p*
A	mand	32	Peter	0	?	0	32	*p< .001*
B	Peter	2	gammel mand	30	?	0	24.05	*p< .001*
C	Peter	0	Søren	32	?	0	32	*p< .001*
D	journalist	14	Forsvarminister	13	?	5	0.03	*.40< p< .50*
E	forsvarminister	26	journalist	5	?	1	16.53	*p< .001*
F	Peter hilste S.	31	Søren hilste P.	1	?	0	28.12	*p< .001*
G	Peter råbte til m.	27	mand råbte til P.	5	?	0	15.12	*p < .001*
H	P. hilste mand	29	mand hilste P.	2	?	1	22.78	*p < .001*
I	Peter	4	mand	16	?	12	4.50	*.10<p<..20*

Table 2: Survey results

In example **E**, the known subject nominal *forsvarsministeren* (the minister of defence) is preferred to the agent *en journalist* (a journalist) which is unknown, but which has a high status because of its thematic role. This example confirms many cases in the data which indicate that in Danish, the hierarchy of grammatical complements is more relevant to anaphora resolution than the hierarchy of thematic roles. This is also the case in English according to Kameyama (1996). In both **D** and **E**, common sense knowledge may also have influenced the results.

In the examples **F-H**, the two pronouns *han* (he) and *ham* (him) are interpreted according to parallelism preference. The answers to **G** and **H**, compared to those in examples **A** and **B**, show furthermore that parallelism preference overrules focality preference. In the interpretation of **I**, parallelism only competes with subject antecedent preference (the preference is not significant).

In conclusion, the results of the survey confirm the hypothesis that focality preference is a stronger preference than givenness and that both preferences are overridden by parallelism. Other cases of parallelism in Danish are discussed in (Navarretta, 2002).

5.3 Combining Givenness and Focality

We have proposed that givenness preference is valid by default when resolving pronominal anaphora. However, this preference is overridden by focality preference, i.e. the salience ranking proposed by a givenness-based model is overridden if there is a focus proper candidate antecedent in the explicitly focally-marked constructions we have tentatively listed. In the following we give an example of how givenness preference, implemented by a Centering-based model of salience, can be combined with focality preference. We use the hierarchy of grammatical complements as a givenness-based model of salience (Brennan *et al.*, 1987). In the model, focality preference is simply expressed by putting the focus proper in front of the list of forward-looking centers $C_f\ (U_{n-1})$ as illustrated in Figure 5.

FOCUS PROPER > SUBJECT > OBJECT/PrepOBJECT > OBJECT 2 > OTHER COMPLEMENTS > ADJUNCTS

Figure 5: Hierarchy of verbal complements with focality preference

The fact that focality preference overrides givenness in determining the degree of salience of elements in an utterance has also consequences for the assumption in Centering that continuing speaking about the same elements in discourse is perceived as more coherent than changing the centre of attention. As described in Section 2, interrupting the centering chain of reference in Centering results in a *shifting* transition state which being assumed to be less coherent than *continuing* and *retaining* transition states is assigned a lower rank than them. In our opinion, this is not correct. In the majority of cases, discourse is coherent and because speakers explicitly change the "focus of attention" using information structure related devices this shift is as coherent as centre continuation. Therefore, coherence in discourse is not only expressed by the fact that speakers continue speaking about the same centres for a while, but also by other phenomena such as information structure or relations holding between discourse units, which can be used to discover parallelism and other phenomena, as proposed in i.a. (Hobbs, 1979; Asher, 1993; Kehler, 2000). Explicitly focally marked elements in the focal-marking constructions which we have described, can easily be recognised in Danish. However, there are utterances where the main focus can only be identified by analysing the context of discourse. This analysis requires much more sophisticated processing techniques than those used in simple resolution approaches such as Centering-based ones.

6 Related Approaches

Hajičová *et al.*'s proposal that entities referred to in the focal part of utterances have the highest degree of salience is unique. However, the fact that elements in the focal part of an utterance can be very accessible in the addressee's discourse model is confirmed by psycholinguistic experiments conducted by Arnold (1998). Arnold tests the accessibility of focal nominals in clefts and compares their accessibility with the accessibility of subjects which in all linguistic theories are considered to be "very" *given*. Her experiments indicate that both the focus of clefts and the grammatical subject increase the accessibility of their referents. Arnold also investigates reference to foci of clefts and subjects in a corpus. The results of her analysis indicate that the referents of both are highly likely to be referred to again.

Some focal constructions have also been recognised as special in various resolution approaches, although these approaches do not explicitly refer to a general theory of focality preference.

Sidner (1983) suggests that the discourse focus, which in her resolution framework is the most prominent entity at that point of discourse, is explicitly indicated in cleft-, pseudocleft- and *there*-constructions.

Mitkov (1998) recognises topicalisation as one of the many factors influencing anaphora resolution. Furthermore, in his proposal, the objects in a number of verbal constructions in certain types of text also receive a high score as candidate pronominal antecedents.

Fraurud (1992) proposes a simple algorithm for resolving intersentential pronominal anaphora in Swedish. In her proposal, the highest ranked antecedent in a sentence is the subject, but she uses recency as a second ranking criterion. This means that rightmost nominals are preferred to leftmost ones. In Swedish, (Vallduví & Engdahl, 1995), as in Danish (Togeby, 1993; Paggio, 1997), focal elements tend to occur in the final part of the utterance while topic elements preferentially occur in the beginning of an utterance. Hence, Fraurud's algorithm, in some cases, ranks nominals in the focal part of the utterance higher than nominals in the topic part.

7 Summary and Concluding Remarks

The analysis of Danish discourse indicates that entities referred to in the focal part of an utterance in information structure terms are, in some specific cases, more salient than entities referred to in the topic part of the utterance. Nominals in the topic part of an utterance often correspond to the most given entities in the addressee's discourse model. They are preferred as antecedents in most models of salience because the models identify degree of salience with degree of givenness in the addressee's discourse model. In these models, antecedents representing newly introduced information, which correspond to focal information, are assigned the lowest degree of salience. Only few types of focal nominals are sometimes chosen as antecedents in some of the proposed models, not because they are focal, but because they are the leftmost candidate antecedents in the preceding utterance or because they are the subjects of the utterance.

The only model of salience explicitly assigning the highest degree of salience, and then of accessibility, to entities referred to in the focal part of the utterance is that proposed by Hajičová *et al.* However, in Hajičová *et al.*'s model all nominals in the focal part of an utterance are assigned the highest degree of salience. Furthermore, in their resolution system focal elements only compete with given (topic) elements as antecedents of intersentential pronominal anaphora.

Centering-based models identifying degree of salience with degree of givenness do not account for the high prominence of focal elements, but, on the other side, Hajičová's assumption that all focal elements have the highest degree of salience is not always true.

In this paper we proposed that accessibility by default is connected with *givenness* as assumed in Centering, but speakers can explicitly change the degree of accessibility of elements in discourse by marking them as salient with information structure related devices. Only when speakers explicitly mark nominals as the main focus of an utterance, these nominals have the highest degree of salience and can be chosen as antecedents of anaphors. In these cases, the shift of focus of attention is, in our opinion, as coherent as continuing speaking about the same elements, because it is pre-announced to the addressee. Also in this aspect, our proposal departs from the original formulation of Centering. A tentative list of constructions in which this explicit focus-marking occurs was also given in the paper together with a model of salience combining the Centering-based approach with focality preference.

Although in this paper we have focused on givenness and focality as indicators of the salience of entities in discourse, other aspects such as animacy (see i.a. (Fraurud, 1992)) influence salience. Furthermore, factors such as parallelism and world knowledge are stronger than salience-based preferences and should be applied after salience-based resolution in ambiguous cases, i.a.

(Sidner, 1983; Kameyama, 1998). However, because world knowledge and conventional presuppositions require a much more deep analysis of discourse than simple syntactic structure, they are less attractive than salience-based models in applied approaches.

In the last part of the paper, we shortly presented the results of psycholinguistic studies which confirm the hypothesis that some types of foci have high degree of accessibility and we listed resolution algorithms that give high accessibility ranking to some focal phenomena, i.a. (Sidner, 1983; Fraurud, 1992; Mitkov, 1998). Our proposal is new because it generalises the relation between givenness and focality preferences and relate them to parallelism.

We have only analysed Danish data. Our proposal should be verified on other languages.

References

Ariel, M. 1994. "Interpreting Anaphoric Expressions: A Cognitive Versus a Pragmatic Approach." *Journal of Linguistics*, 30(1).3-40.

Arnold, J. 1998. *Reference Form and Discourse Patterns*. Ph.D. thesis. Stanford University.

Asher, N. 1993. "Reference to Abstract Objects in Discourse." *Studies in Linguistics and Philosophy* 50. Dordrecht: Kluwer Academic Publishers.

Brennan, S., M. Friedman and C. Pollard 1987. "A Centering Approach to Pronouns." *Proc. of the 25th Annual Meeting of the ACL (ACL'87).* Stanford, CA. 155-162.

Daneš, F. 1968. "Some Thoughts on the Semantic Structure of the Sentence." *Lingua* 21.55-69.

Diderichsen, P. 1957[1946]. *Elementær Dansk Grammatik*. Copenhagen: Gyldendal.

Duncker, D. and J. Hermann 1996. "Patientord og lægeord - særord eller fællesord?" *Månedsskrift for Praktisk Lægegerning - Tidsskrift for Praktiserende Lægers Efteruddannelse*. 1019-1030.

Fraurud, K. 1992. *Processing Noun Phrases in Natural Discourse*. Department of Linguistics - Stockholm University.

Givón, T. 1976. "Topic, Pronoun and Grammatical Agreement." C. Li (ed.) *Subject and Topic*. New York: Academic Press. 149-188.

---------- 1979. *On Understanding Grammar*. New York: Academic Press.

---------- (ed.) 1983. *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam: John Benjamins.

Gregersen, F. and I. Pedersen (eds.) 1991. *The Copenhagen Study in Urban Sociolinguistics*. Reitzel.

Grosz, B. and C. Sidner 1986. "Attention, Intentions, and the Structure of Discourse." *Computational Linguistics* 12(3).175-204.

----------, A. Joshi and S. Weinstein 1995. "Centering: A Framework for Modeling the Local Coherence of Discourse." *Computational Linguistics* 21(2).203-225.

Gundel, J., N. Hedberg and R. Zacharski 1993. "Cognitive Status and the Form of Referring Expressions in Discourse." *Language* 69(2).274-307.

----------, M. Hegarty and K. Borthen 2001. "Information structure and pronominal reference to clausal introduced entities." *Proc. of the ESSLLI 2001 Workshop on Information Structure,* *Discourse Structure and Discourse Semantics,* Helsinki. http://www.coli.uni-sb.de/korbay/esslli01-wsh/index.html.

Hajičová, E. and J. Vrbová 1982. "On the Role of the Hierarchy of Activation in the Process of Natural Language Understanding." J. Horecký (ed.) *Proc. of the 9th Int. Conference on Computational Linguistics,* 107-113, Prague-Amsterdam.

----------, P. Kuboň and V. Kuboň 1990. "Hierarchy of Salience and Discourse Analysis and Production." H. Karlgren (ed.), *Proc. of the 13th Int. Conference on Computational Linguistics (COLING'90)*, volume III, 144-148, Helsinki.

Henrichsen, P. J. 1998. "Peeking Into the Danish Living Room." CST (ed.), *Nodalida '98 Proc. - The 11th Nordic Conference on Computational Linguistics*. Center for Sprogteknologi and Department of General and Applied Linguistics-University of Copenhagen.

Hobbs, J. 1979. "Coherence and Coreference." *Cognitive Science* 3(1).67-90.

Hoffmann, B. 1998. "Word Order, Information Structure and Centering in Turkish." M. Walker, A. Joshi and E.Prince (eds.), *Centering in Discourse*. Oxford: Oxford University Press.

Jensen, K. A. 1989. *Projekt invandrerdansk*. Technical report, Copenhagen University.

Kameyama, M. 1988. "Japanese zero pronominal binding, where syntax and discourse meet." William Poser (ed.), Papers from the 2nd Int. Workshop on Japanese Syntax. Stanford, CA: CSLI.

---------- 1996. "Indefeasible Semantics and Defeasible Pragmatics." M. Kanazawa, C. Piñon and H. de Stwart (eds.) *Quantifiers, Deduction and Context*, 111-138. Stanford, CA: CSLI.

---------- 1998. "Intrasentential centering: A case study." M. Walker, A. Joshi and E. Prince (eds.), *Centering Theory in Discourse*, 89-112. Oxford: Oxford University Press.

Kehler, A. 2000. "Coherence and the Resolution of Ellipsis." *Linguistics and Philosophy* 23.533-575.

Kruijff, G.-J. M. 2001. *A Categorial-Modal Architecture of Informativity: Dependency Grammar Logic and Information Structure*. Ph.D. thesis, Charles University, Prague.

Kuno, S. 1987. *Functional syntax: anaphora, discourse, and empathy.* Chicago: University of Chicago Press.

Mitkov, R. 1998. "Robust pronoun resolution with limited knowledge." *Proc. of the 36th Annual Meeting of the ACL and the 17th Int. Conference on Computational Linguistics (COLING-ACL'98)*, 869–875, Montreal.

Navarretta, C. 2002. *The Use and Resolution of Intersentential Pronominal Anaphora in Danish Discourse*. Ph.D. thesis, Center for Language Technology and Department of General and Applied Linguistics, Copenhagen University.

Paggio, P. 1997. *The Treatment of Information Structure in Machine Translation*. Ph.D. thesis, University of Copenhagen.

Prince, E. 1981. "Toward a taxonomy of given-new information." P. Cole (ed.), *Radical Pragmatics*, 223-255. New York: Academic Press.

---------- 1992. "The ZPG Letter: Subjects, Definiteness, and Information-status." W. Mann and S. A. Thompson (eds.), *Discourse Description. Diverse Linguistic Analyses of a Fund-Raising Text.* 295-325. Amesterdam: John Benjamins.

Quirk, R., S. Greenbaum, G. Leech and J. Svartvik 1985. *A Comprehensive Grammar of the English Language*. Longman.

Sgall, P., E. Hajičová and J. Panevová 1986. *The Meaning of the Sentence in its Semantic and Pragmatic Aspects*. Dordrecht: Reidel.

Sidner, C. 1983. “Focusing in the Comprehension of Definite Anaphora.” M. Brady and R. Berwick (eds.), *Computational Models of Discourse*, 267-330. Cambridge, MA: MIT Press.

Strube, M. and U. Hahn 1996. “Functional Centering.” *Proc. of the 34th Int. Conference on Computational Linguistics (ACL’96)*, 270–277, Santa Cruz, CA.

---------- 1998. “Never Look Back: An Alternative to Centering.” *Proc. of the 36th Annual Meeting of the ACL and the 17th Int. Conference on Computational Linguistics (COLING-ACL’98)*, vol. II, 1251-1257, Montreal, Quebec, Université de Montréal.

Tetreault, J. 2001. “A Corpus-Based Evaluation of Centering and Pronoun Resolution.” *Computational Linguistics* 27(4).507-520.

Togeby, O. 1993. *PRAXT - Pragmatisk tekstteori 1-2*. Aarhus Universitetsforlag 1993.

Vallduví, E. and E. Engdahl 1995. “The linguistic realization of information packaging.” *Linguistics* 34(3).459-519.

Woods A., P. Fletcher and A. Hughes 1986. “Statistics in Language Studies.” *Cambridge Textbooks in Linguistics*. Cambridge: Cambridge University Press.

Pronouns Without NP Antecedents: How do we Know when a Pronoun is Referential?

Jeanette Gundel[1], Nancy Hedberg[2] and Ron Zacharski[3]
[1]*University of Minnesota,* [2]*Simon Fraser University,* [3]*New Mexico State University*

Pronouns without explicit noun phrase antecedents pose a problem for any theory of reference resolution. We report here on an empirical study of such pronouns in the Santa Barbara Corpus of Spoken American English, a corpus of spontaneous conversation. Analysis of 2,006 third person personal pronouns in fourteen transcripts indicates that 330 (16%) lack NP antecedents. Of these, the following three types are most common: (1) pronouns whose referents are associated with a recently evoked entity ('inferrables'); (2) pronouns that refer to a fact, proposition, event, activity, situation, or reason evoked by a previous non-nominal constituent; and (3) pleonastic pronouns. We focus on some problematic cases that could be analysed as either referring to entities introduced by or implied in previous discourse, or as non-referential, including pleonastic. Such cases include certain non-specific uses of *they*, possible subjects of truncated cleft sentences, and possible subjects of truncated extraposition constructions.

1 Introduction

The referent of a prototypical pronoun has been recently introduced into the discourse by an explicit noun phrase 'antecedent', as in (1):

(1) a. My neighbor's Bull Mastiff bit a girl on a bike.
  b. **It**'s the same dog that bit Mary Ben last summer. (Gundel *et al.*, 1993)

These are the simplest cases for a theory or algorithm for pronoun resolution to account for. It is well known, however, that an NP antecedent is neither necessary nor sufficient for appropriate pronoun use. As we propose in (Gundel, *et al.*, 1993), a recent NP antecedent does not license use of the personal pronoun *it* if it does not bring the intended referent of the pronoun into focus of attention, the necessary 'cognitive status' for appropriate use of this form. This explains the contrast between (1b), where the referent of *it* is introduced in syntactically prominent subject position, with (2b) where it is introduced in a syntactically embedded, non-argument position that is much less likely to bring an entity into focus.[1]

(2) a. Sears delivered new siding to my neighbors with the Bull Mastiff.
  b. #**It**'s the same dog that bit Mary Ben last summer. (Gundel *et al.*, 1993)

[1] As Gundel *et al.* (1993) point out, syntax is a crucial factor in determining whether an entity will be merely activated or brought into focus of attention, but semantic and pragmatic factors play a role as well. See (Gundel *et al.*, 2003) for more detailed discussion of various factors that promote the salience of discourse entities.

Moreover, an NP antecedent, or an explicit antecedent of any sort, is not always necessary. The intended referent may be brought into focus in some other way or it may be easily accommodated, a fact which poses problems for any theory of reference resolution (see (Cornish, 1999; Gundel *et al.*, 2000; Byron, 2000), *inter alia*).

In the present paper, we report on an empirical study of personal pronouns without NP antecedents in the Santa Barbara Corpus of Spoken American English. Previous studies offer some support for the view that referential pronouns without NP antecedents occur more frequently in casual conversations than they do in other genres. For example, in (Gundel *et al.*, 2000) we report that such forms occurred more frequently in newsgroup conversations about eating disorders than in web pages about such disorders. Sanford *et al.* (1983) report that in an editing task subjects tended to replace referential pronouns without explicit antecedents with full noun phrases. Of the pronouns of interest in our current study, particular attention will be paid to cases where it isn't clear whether the pronoun is referential or not.

2 Methodology

We analysed roughly the first 10 pages, ranging from 9.5 to 25 minutes, in each of 14 transcripts from the Santa Barbara Corpus of Spoken American English Part-1 (DuBois *et al.*, 2000). Each transcript was analysed by one of two coders. One coder coded transcripts 1-7 and a second coder coded transcripts 8-14. Coder reliability was improved by having a random sampling of the transcripts analysed by a third coder. Any discrepancies were discussed among all coders to reach a consensus on coding practices. Transcripts were then reanalysed by the original coder using this coding standard. Classification of any remaining hard-to-classify cases was done by consensus.

The transcripts represent conversations between two to six speakers in a variety of settings. For example one conversation took place when the three participants were preparing a dinner, another took place at a birthday party, and a third was a classroom discussion. We coded each of the 2,006 third-person personal pronouns (excluding false starts) in the corpus for whether or not it had an explicit co-referring NP antecedent.[2] Pronouns without such antecedents were further classified as pleonastic (i.e. expletive and non-referential) or referential. Referential pronouns were classified as having a non-nominal antecedent that evoked a fact, proposition, activity, situation, etc. (e.g. *I'm not*

[2] For the sake of comparison, we also counted the demonstrative pronouns in the transcripts, finding 601, and classified them as to whether or not they had an NP antecedent. As discussed in Section 4, only 28% of the demonstrative pronouns had NP antecedents, contrasting sharply with the personal pronouns in Table 1.

*all bent out of shape about **it*** (5.1072), following a statement that chemical makeup largely determines who we are) or as an inferrable, if they had no explicit antecedent at all and therefore had to be inferred. (e.g. *I went over to a store, where **it** says, sewing alterations*. (11.453)).

The results of the primary classification are shown in Table 1.

	N	**%**
NP antecedent	1676	83.55
Inferrable	88	4.39
Non-nominal antecedent	110	5.48
Pleonastic	92	4.58
Other*	40	2.00
Total	2006	100.00

* The ‘Other’ category included pronouns referring to entities in the extralinguistic environment, idioms such as ‘God damn it’, and pronouns that we couldn’t interpret.

Table 1: 3rd Person Personal Pronouns in Santa Barbara Corpus

It can be seen that roughly 16% of the pronouns lacked a coreferring NP ‘antecedent.’ We now turn to a discussion of each of the major categories of these.

3 Inferrables

We classified as ‘inferrables’ those pronouns whose referents were not directly introduced into the discourse by a previous linguistic expression or by their presence in the extralinguistic context.

3.1 Inferrables with specific referents

In (3), the referent of the pronoun is an example of what Prince (1981) calls an ‘inferrable’, (cf. ‘associate anaphor’ (Hawkins, 1978, 1991), ‘indirect anaphor’ (Erkü & Gundel, 1987; Gundel *et al.*, 2000)). The addressee has to infer from mention of the kids across the street that *she* refers to their mother.

(3) [*Talking about how the kids across the street threw paint in their yard*.]
Those kids are just – And **she**’s pregnant with another one. (2.294)

All of the clearly referential inferrables in our data have what Cornish (1999) calls an ‘antecedent trigger’.[3] For example, in (3), the antecedent trigger for *she* is *the kids.* There are a variety of relations between the antecedent trigger and the inferrable pronoun. These include stereotypical connections such as children to mother, as in (3), or discussion of a class to *he* referring to the teacher. In (4), the antecedent trigger is the name of the musical group Oba Oba (or possibly the whole statement *We went to see Oba Oba*) and the referent

[3] Inferrables which lack a specific referent (‘vague inferrables’) will be discussed in 3.2.

of the pronoun *it*, the concert performed by the group, must be inferred from the statement that the speaker went to see the group.

(4) We went to see Oba Oba. You know. Ruben loved **it**. (6.157)

Other possible relations include generic kind to set of specific entities, individual to couple, or individual to group (cf. the 'poset' relations of Hirschberg (1991)). In (5), for example, the antecedent trigger is *Trish* and the inferrable makes reference to the couple, Trish and her husband.

(5) A: Was it Trish who told me she was pregnant?
B: She looked really good. Where are **they** going to church? (13.221)

In (Gundel *et al.*, 2000), we propose that such pronominal inferrables constitute minor violations of the restriction that the referent of a pronoun must be in the addressee's focus of attention, if unstressed, or at least activated (in working memory), if stressed (Gundel *et al.*, 1993). There is no reason to assume, for example, that reference to children, as in (3), will necessarily bring into focus, or even activate, a representation of the children's mother or that mention of Trish in (5) would automatically bring into focus a representation of both Trish and her husband. We would therefore also not expect a formal mechanism to activate a person's partner every time that person is mentioned. But such violations of the restriction that the referent of the pronoun must be already activated or in focus are easily accommodated by way of a bridging inference (Clark & Haviland, 1977) that links the referent to a recently activated entity. Gundel *et al.* (2000) further point out that pronominal inferrables are relatively infrequent and, for the most part, restricted to casual, spontaneous discourse, a fact which is consistent with the hypothesis that they are often grammatical violations and that accommodation is involved in their processing. These claims are also supported by our current data, where less than 5% of the total number of personal pronouns are inferrables.

We also included among inferrables examples like those in (6) where the pronoun refers to the class/kind that a recently activated entity is part of.

(6) A: Where is that salad spinner? Here it is.
B: And possibly the most spurious device ever created.
A: Oh I think **they**'re great. (3.155)

Such examples differ from inferrables like those in (3)-(5), however, since it could be argued that the generic kind is necessarily activated when a specific entity is processed.[4] In the case of linguistically introduced entities, this would

[4] This would also explain why all statuses on the Givenness Hierarchy (Gundel *et al.*, 1993) entail the status 'type identifiable,' i.e. the addressee can be assumed to have a representation of the class/kind that the referent belongs to.

follow from the fact that processing the referent of an NP always involves processing of lower levels of syntactic representation of that NP. Thus, we can assume that processing the phrase *that salad spinner* necessarily activates not only a representation of the particular salad spinner that this phrase refers to, but also a representation of the generic kind of salad spinners (the referent of *salad spinner*). In (6), moreover, the kind would be not only activated, but brought into focus since speaker B makes a generic statement about salad spinners.[5]

Whether they are in focus or not, these inferrables are referential, and a specific referent must be assigned in order for full interpretation to take place. In accommodation cases, the inferential process required for resolution relies on general background knowledge and the hearer's ability to access an appropriate referent without undue processing effort (Sperber & Wilson, 1986/95) and there is no obvious formal mechanism that activates the appropriate referent based on linguistic information alone. As noted, however, such examples seem to be found primarily in casual speech, and even here they are relatively infrequent. We now turn to a discussion of the possibility that some inferrable pronouns might not be referential at all.

3.2 Vague Inferrables

There are some inferrable pronouns that are only loosely referential but may be restricted in reference by a recently activated entity or other contextual factors. For example, in (7) and (8), *they* refers to people, and in (9), it is even more vague.

(7) And they say that if there's six years between children, there's not that much rivalry. (7.1247)

(8) And they probably didn't have to wash their salads back then, because they didn't know what was on them. (3.165)

(9) He said I didn't get done working until after nine. …Cause that five-car pile up they had between Hardin and Crow? (7.414)

Reference resolution in such cases is not only difficult, but typically unnecessary. Sentences like (7) can be replaced by agentless passives with no loss of information content, as in (10).

(10) It is said that if there's six years between children, there's not that much rivalry.

To sum up this section, we classified nearly 5% of the pronouns in our data as 'inferrables.' In one class of such cases, the referent of the pronoun must be inferred, linking it to an entity activated by a previous noun phrase. In another

---

[5] Note, however, that the personal pronoun in A's second contribution in (6) would be possible even without the generic statement by B.

class of cases, there is no clear referent. Rather, a non-specific *they* is used to evoke a meaning such as ‘people in general’ or ‘people in the principal’s office.’ We suggested that it is not necessary to resolve the reference in such vague-inferrable cases. Both types of inferrables violate the condition on personal pronouns that they have a referent that is already in the focus of the addressee’s attention, which may be why they are characteristic of unplanned, casual speech.

4 Pronouns With Non-Nominal Antecedents

Events and activities, as well as facts, propositions, situations and other ‘higher-order’ entities are often introduced into discourse by non-nominal constituents like clauses, sequences of clauses and verb phrases. Previous work has found that it is more common for such entities to be referenced by a demonstrative pronoun than by a personal pronoun. For example, Webber (1991) found that only 15 out of 96 (or 15.6%) pronominal references to clausally introduced material in written English were made using the personal pronoun *it* as opposed to a demonstrative pronominal *that* or *this*. Hegarty *et al.* (2001), in a study of both spoken and written texts, reached an almost identical result, finding that only 15 out of 95 references to entities introduced by expressions larger than a noun phrase were made with *it* as opposed to *this* or *that*.

The theory of referential expression choice of Gundel *et al.* (1993) can explain these figures if entities introduced by a verb phrase, clause or sequence of clauses are activated, but are much less likely to be brought into focus than entities introduced by noun phrases in syntactically prominent positions. Gundel *et al.* (2003) further suggest that one factor determining whether an entity introduced by a non-nominal is brought into focus, and is therefore available to immediately subsequent reference with *it* is its degree of ‘world immanence’ (roughly degree of abstractness), as discussed in Asher (1993). Specifically, the more world immanent the referent, the more likely it is to be brought into focus when first introduced. Gundel *et al.* (2003) propose that this is because clauses and verb phrases directly introduce only the events, states or activities they describe. Interpretation of pronouns that refer to entities such as facts and propositions that are inferred from these eventualities thus requires additional processing as such entities are not brought into focus of attention (and possibly not even automatically activated) by simply processing the non-nominal constituent. The examples in (11)-(14), all from (Hegarty *et al.*, 2001), illustrate these claims.

(11) John insulted the ambassador. **It/that** happened at noon.

(12) John insulted the ambassador. ??**It**/**that**/**this** was intolerable to the embassy.

(13) A: I read somewhere that the poodle is one of the most intelligent dogs around.
B: well uhm. . I definitely wouldn't dispute **that**. (Switchboard Corpus, Dialog 2019)
B':?? well uhm. . I definitely wouldn't dispute **it**.

(14) a. "We believe her, the court does not, and **that** resolves the matter," Mr. Montanarelli said today of Ms. Lewinsky's testimony that… (NY Times, 5/24/00)
b. "We believe her, the court does not, and **it** resolves the matter," Mr. Montanarelli said today of Ms. Lewinsky's testimony that…

In (11), where the pronoun refers to the event of John insulting the ambassador, the personal and demonstrative pronouns are equally acceptable. By contrast, the personal pronoun is less acceptable in (12) where it refers to the situation of John's insulting the ambassador. A personal pronoun is also less acceptable than a demonstrative in (13), where it refers to a proposition, and in (14), where it refers to a fact.[6] The contrast is especially striking in (14), where there is a competing 'in focus' entity, the court. The personal pronoun in (14b) would be taken as referring to the court rather than to the complex fact inferred from the content of the preceding two clauses.

Our data support the claim that referents of pronouns with non-nominal antecedents are not usually brought into focus, and are thus relatively inaccessible to subsequent reference with the personal pronoun *it*. Only 110, roughly 5%, of the total number of personal pronouns and 1/3 of the total number of personal pronouns without NP antecedents referred to higher-order entities associated with a non-nominal expression, i.e. facts, propositions, activities, events, situations, or reasons. Moreover, only 16 of these (15% of those with non-NP antecedents and 0.8% of the total number of personal pronouns) were coded as referring to facts and propositions, and few if any of those were clear cases.

By contrast, an analysis of demonstrative pronouns in our data reveals that 433 out of 601 of these (72%) lacked NP antecedents. Although we haven't yet done a classification of these demonstrative pronouns as to semantic type of the referent, perusal of the data reveals quite a few references to facts and propositions, for example (15) and (16):

(15) Pete: I stuck up for you today at that store.
Harold: **That**'s true. (2.169.79)

(16) This is a raging bureaucracy, . . and there's nothing I can do. . . I have found **that** out. (4.445.98)

[6] The semantic type of the entity is not always clear cut, but it can usually be determined from the semantics of the predicate. For example, *it* in *It happened at noon* refers to an event (cf. *The event happened at noon*, but *??The situation happened at noon*), but *it* refers to a situation in *It was intolerable* (cf. *The situation was intolerable*, but ??*The event was intolerable*). See (Asher, 1993; Hegarty *et al.*, 2001; Gundel *et al.*, 2003) for more discussion.

In (15), *that* refers to the proposition that Pete stuck up for Harold today at the store. In (16), *that* refers to the fact there is nothing the speaker can do. The pronoun *it* would not have been felicitous in either of these cases.

With regard to personal pronouns, it is at least consistent with the proposal in (Gundel *et al.*, 2003) that very few personal pronouns with non-nominal antecedents in our data refer to facts and propositions and, as noted above, those that were coded as referring to facts and propositions are not clear cut cases. Some of the clearest examples are given in (17)-(19).

In (17), *it* in *it's a double whammy* may be interpreted as referring to the fact, inferred from the previous three clauses, that Chicanos do not vote in great numbers (compared to other Americans) and that Americans do not vote in great numbers (compared to people in other countries).

(17) ... Chicanos do not vote in great numbers. And we don't participate in many organizations in great numbers.. . .I don't care if you're African-American, …I don't care if you're Asian-American, . . . and I don't care if you're Latino, or whatever. . . . Most Americans, . . . do not vote, . . . in great numbers. . . . So, . . **it**'s a double whammy. (12.1026)

In (18), the speaker is suggesting that they check out the proposition that the measuring cup is unbreakable. In (19), the speaker is saying that the proposition that they approve the loan request is moved and seconded.

(18) Wendy: . . . Yes. Microwavable, chef. Eight ounce measuring cup. Is virtually unbreakable.
Kevin: Virtually, let's find out.
Kendra: Let's check **it** out. (13.551)

(19) Joe: I am moving that we approve this loan request
(about three minutes of discussion)
Fred: I second then Joe . . .
Joe: So **it**'s moved and seconded . . to uh . . (14.415)

Note, however, that (17)-(19) are not counterexamples to the claim that a clause or verb phrase does not bring an associated fact or proposition into focus, since, in all three cases, the fact or proposition has been mentioned, overtly or covertly, in an intervening sentence. This could be what brings the fact or proposition into focus and licenses reference with a personal pronoun. In (19), moreover, in Joe's first contribution, the entity is introduced as a complement to a verb that takes a proposition as its argument. So it is likely to have been introduced as a proposition the first time.

The majority of pronominal references with non-nominal antecedents in our data were classified as activities (25%) or situations (57%). The relatively high number of personal pronouns referring to activities is consistent with the proposal that activities and states are more likely to be brought into focus because they are directly introduced by a non-nominal constituent. In (20), for example, the pronoun *it* refers to the activity of doing the translations, which was introduced in the infinitival clause.

(20) I'm going to do some translations for her and stuff? And um, you know, I have to make at least 50 dollars or so, to make **it** worth my time. (7.323)

In (21), the pronoun refers to the activity of tap dancing.

(21) He has to double **it** down to like one-fifth speed or something, before they can pick it up. (2.105)

It is less clear, however, why there should be so many situations referenced with pronominal *it*, as situations are intermediate in world immanence; but the distinction between situations, activities and states was not always clear and some of these may have been misclassified.

For example, in (22), the pronoun was coded as referring to a situation. However, it could also be interpreted as referring to the activity of watching the car thief showing how not to get your car broken into. The latter interpretation would be more consistent with the use of a personal pronoun here, given the claim that activities are directly introduced by the non-nominal constituent and thus more likely to be brought into focus.

(22) He's gonna show us, you know, how not to protect your car, not to get it, you know, ripped off man. Cause, you know, I -- . , yeah, I was into **it**. (6.31)

Finally, some of the examples of *it* coded as referring to situations were quite vague, for example (23).

(23) Mary: ...It is really hard living with another couple.
Alice: ... I mean, .. we -- If we set our -- .. if we sit down and set some rules, which we never did, .. it could work. (7.103.78)

One reason why we had difficulty classifying many of the examples discussed in this section is that they were vague or unclear in reference. It may be the case that the possibility of quasi-referentiality is a feature of personal pronouns *it* and *they* that distinguishes them from demonstrative pronouns *this* and *that*. The indeterminacy of reference characteristic of the examples in this section, like the use of personal pronouns for inferrable entities discussed in the last section, might well be a feature unique to casual, spontaneous speech.

5 Pleonastics

The main three types of pleonastic pronouns were cleft pronouns, extraposition pronouns and atmospheric pronouns, as exemplified in (24)-(26) respectively:

(24) Was **it** Trish who told me she was pregnant? (13.216)
(25) I just think **it**'s so damn weird we're here. (5.529)

(26) **It**'s so cold outside. (7.558)

Clear examples of pleonastic pronouns in non-elliptical constructions such as those in (24)-(26) can usually be identified as non-referential based on grammatical, i.e. syntactic or lexical, information alone. But there were some cases where it is not clear whether the pronoun is truly non-referential. There were two types of such pronouns in the corpus: truncated cleft pronouns and truncated extraposition pronouns.

5.1 Truncated Cleft Pronouns

Subjects of truncated cleft sentences, as in (27), where there is ellipsis of the cleft clause *who stole Hector's radio,* were also classified as pleonastic.

(27)A: It's obvious now that this guy w- -- This was the one who stole .. Hector's [radio]. I mean, .. nobody came out and told you, guess what, I confess.
B: Oh, we knew. .. We knew. .. We figured **it** had to be Michael. (2.70)

However, Hedberg (2000) presents a theory of cleft sentences that claims that cleft pronouns combine with the cleft clause to form a discontinuous definite description, i.e. a referring expression. The appropriate form of the pronoun is determined by the cognitive status of the intended referent of the description, according to the Givenness Hierarchy of *Gundel et al.* (1993). In (27), for example, the interpretation of the elided description (the one who stole Hector's radio) is in focus because it has just been introduced in a syntactically prominent focus position in (27A). Hedberg maintains that the cleft pronoun is not a meaningless, pleonastic element but rather is referential in the same sense that a determiner is in the case of full clefts, and in the sense that at least some pronouns are in the case of truncated clefts. To understand in what sense the subject pronoun of a truncated cleft is referential, compare (27) with Hedberg's example in (28).

(28) My heart beat fast, for I had thought that as the discoverer of the body I would be the first to be called; but to my surprise, **it** was Marcel.

Here, the truncated cleft could be replaced by a full cleft, *It was Marcel who was called.* Hedberg's claim is that the subject pronoun in the truncated cleft co-refers with *the first to be called.*

There are several examples of truncated clefts in the Santa Barbara corpus whose subjects we classified as pleonastics but which could equally well be classified as referring to a reason or a cause evoked in prior discourse. For example, (29) and (30):

(29)A: So that's why you're interested in death?
B: Maybe **it**'s because my parents were old? When I was young? (5.499)

(30)A: What do you think makes em look African?
B: Their mustaches?
A: Is **it** the way their little beard goes? (2.517)

Here again, the truncated cleft could be replaced by a full cleft: *Maybe it's because my parents were old that I am so interested in death*; *Is it the way their little beard goes that makes em look African?* Furthermore, the sentences could be paraphrased as pseudoclefts: *Maybe why I'm so interested in death is because my parents were old*; *Is what makes em look African the way their little beard goes?* However, the cleft pronoun in both (27) and (28) can also be taken as referring to a reason or cause evoked in the previous question. So (26) may be paraphrased as *Maybe the reason I'm so interested in death is because my parents were so old*; *The cause of them looking African is the way their little beard goes*.

In (31), the pronoun could be analysed as the pleonastic subject of a truncated cleft, but it could also refer to a cause, the existence of which can be inferred from the fact that the speedometer fell. The full cleft paraphrase here would be *I knew exactly what it was that caused it.* The noncleft paraphrase would be *I knew exactly what the cause was.*

(31) I saw my .. my speedometer just go Brr=. .. like that just dow=n,. You know, and I knew exactly what **it** was.

The subject pronouns in (27)-(31) can thus be analysed as 'referential.' Moreover, unlike subjects of full clefts, their non-referential nature can't be predicted on the basis of grammatical properties alone.

It is consistent with the data, however, to claim that truncated cleft pronouns are true pleonastic pronouns, and that what the hearer must do is to reconstruct the material elided from the cleft clause. What's important here is that for purposes of interpretation it doesn't matter which analysis is chosen. Either the referent of the pronoun must be resolved, or the logical form of the utterance must be enriched (in the relevance-theoretic sense of Sperber and Wilson (1986/95)) to provide the information in the elided cleft clause.

5.2 Truncated Extraposition Pronouns

We classified some pronouns as pleonastic because they could be analyzed as truncated extraposition pronouns, as in (32) and (33).

(32) And the second week they were just like (YELL), and so I had to scream at them, all week long. And **it** was really awful, cause I felt horrible about it. (4.83)

(33) You can't really tell when they blush. **It**'s very unusual. (4.298)

Both these sentences can be paraphrased as extraposition constructions: *And it was really awful that I had to scream at them, all week long; It's very* *unusual for them to blush*. They thus have a full paraphrase that is identical to the type that full extraposition clauses have, as in (34) and (35):

(34) But, for me **it**'s really difficult to pick up a book about death. (5.217)
(35) And finally **it** dawns on Lisabeth that she doesn't see Mom that much. (6.275)

Pronominal subjects of extraposition constructions are generally analysed as pleonastic, and we classified them as such. But the subject pronouns in the truncated cases could equally well be analysed as referring to entities evoked in previous discourse, such as a situation in the case of (36) or a generic event or process in the case of (37). And the pronoun can be replaced by a full definite phrase referring explicitly to such a referent:

(36) And **the situation** was really awful, cause I felt horrible about it.
(37) **The event of them blushing** is very unusual.

However, as with the truncated cleft examples, the non-referential, pleonastic nature of the pronoun cannot be determined solely on the basis of grammatical properties. And even if the pronoun is analysed as pleonastic, the ellipsis has to be reconstructed in order for full interpretation to take place.

Table 2 breaks down pleonastic pronouns into the subtypes discussed here. The table shows that close to half of the pronouns classified as pleonastic were subjects of truncated clefts or extraposition constructions which could therefore have not been identified as pleonastic/non-referential based on purely grammatical information.

	N	**%**
Atmospheric	8	8.70
Full cleft	10	10.87
Full extraposition	26	28.26
Truncated cleft	27	29.35
Truncated extraposition	14	15.22
Other pleonastic*	7	7.61
Total	92	100.00

* This category includes such examples as *It seems...* and *His rule has it that...*

Table 2: Pleonastic Pronouns

6 Conclusion

To sum up, 330 of the 2,006 third person personal pronouns in our corpus (16%) lacked NP antecedents. This figure can be compared to the percentages of pronouns without NP antecedents reported on in (Byron, 2002). As Byron notes, Eckert and Strube (2000) found that 55% of pronouns in a set of Switchboard dialogs lacked NP antecedents. Byron and Allen (1998) found that 50% of pronouns in the TRAINS corpus lacked NP antecedents; and Botley (1996) found that 20% of pronouns in a corpus of newswire documents, literary narrative, and parliamentary proceedings lacked NP antecedents. In at least some of these cases, demonstrative pronouns were included in the study, so the total number of pronouns lacking NP antecedents can be expected to be higher than the figure we found since, as discussed above, demonstrative rather than personal pronouns are often used to refer to propositions, facts, situations, and speech acts introduced by a non-nominal constituent. Byron and Allen (1998) compared demonstrative pronouns with the personal pronoun *it* and found that 79% of the former lacked NP antecedents while only 25% of the latter did so. These percentages compare quite closely with our 71.5% and 16.5% respectively.

A central finding of this paper is that it isn't always easy to determine whether a given personal pronoun is referential. We found several apparently inferrable pronouns that were vague in reference, and some references to higher-order entities were quite vague. Furthermore, quite a few pronouns were difficult to classify as either pleonastic or referential. The speech genre analysed in this paper was that of unplanned conversation, which was in most transcripts, quite casual. It seems likely that instances of semi-referentiality would be maximized in this genre. In future research, we will compare our results with a similar study of more formal, planned speech, even edited, written speech such as newspaper articles. We also plan to examine demonstrative pronouns more carefully, to see if semi-referentiality is also a feature of demonstrative pronouns. Finally, more work is needed on the classification of type of referent (e.g. proposition, fact, situation, etc.) for both personal and demonstrative pronouns.

7 References

Asher, N. 1993. *Reference to Abstract Objects in Discourse*. Dordrecht: Kluwer Academic Publishers.

Botley, S. 1996. "Comparing Demonstrative Features in Three Written English Genres." *Proc. of the Discourse Anaphora and Resolution Colloquium (DAARC-96)*, 86-105.

Byron, D. 2002. "Resolving Pronominal Reference to Abstract Entities." *Proc. of the 2002 Meeting of the ACL.*

---------- 2000. "Semantically Enhanced Pronouns." *Proc. of the 3rd Int. Conference on Discourse Anaphora and Anaphor Resolution (DAARC-2000)*, Lancaster University, UK.

---------- and J. Allen 1998. "Resolving Demonstrative Pronouns in the TRAINS93 Corpus." *Proc. of the 2nd Colloquium on Discourse Anaphora and Anaphor Resolution (DAARC-2)*, 68-81.

Clark, H. and S. D. Haviland 1977. "Comprehension and the Given-new Contract." R. Freedle (ed.) *Discourse Production and Comprehension*. Norwood: Ablex. 1-40.

Cornish, F. 1999. *Anaphora, Discourse, and Understanding. Evidence from English and French*. Oxford: Clarendon Press.

Dubois, J., W. Chafe, C. Meyer and S. Thompson 2000. *Santa Barbara Corpus of Spoken American English–Part 1*. Linguistic Data Consortium, University of Pennsylvania.

Eckert, M. and M. Strube 2000. "Dialogue Acts, Synchronising Units and Anaphora Resolution." *Journal of Semantics* 17.51-89.

Erkü, F. and J. Gundel 1987. "The Pragmatics of Indirect Anaphors." J. Verschueren and M. Bertuccelli-Papi (eds.) *The Pragmatic Perspective: Selected Papers from the 1985 Int. Pragmatics Conference*. Amsterdam: John Benjamin. 533-545.

Gundel, J., N. Hedberg and R. Zacharski 1993. "Cognitive Status and the Form of Referring Expression in Discourse." *Language* 69.274-307.

----------, ---------- and ---------- 2000. "Statut Cognitif et Forme des Anaphoriques Indirects." *Verbum* 22.79-102.

----------, M. Hegarty and K. Borthen 2003. "Cognitive Status, Information Structure, and Pronominal Reference to Clausally Introduced Entities." I. Kruijff Korbayova and M. Steedman (eds.) *Journal of Logic, Language and Information. Special Issue on Information Structure*, 12(3).

Hawkins, J. 1978. *Definiteness and Indefiniteness: A Study in Reference and Grammaticality Predictions*, London: Croom Helm.

---------- 1991. "On (In)definite Articles." *Journal of Linguistics* 27.405-42.

Hedberg, N. 2000. "The Referential Status of Clefts." *Language* 76.891-920.

Hegarty, M., J. Gundel and K. Borthein 2001. "Information Structure and the Accessibility of Clausally Introduced Referents." *Theoretical Linguistics* 27(2/3).163-186.

Hirschberg, J. 1991. *A Theory of Scalar Implicatures.* New York: Garland.

Prince, E. 1981. "Towards a Taxonomy of Given-new Information." Peter Cole (ed.) *Radical Pragmatics*. New York: Academic Press. 223-255.

Sanford, A., S. Garrod, A. Lucas and R. Henderson 1983. "Pronouns Without Explicit Antecedents." *Journal of Semantics* 2.303-318.

Sperber, D. and D. Wilson 1986/95. *Relevance: Communication and Cognition,* Oxford: Blackwell.

Webber, B. 1991. "Structure and Ostension in the Interpretation of Discourse Deixis." *Language and Cognitive Processes* 6.107-135.

Syntactic Form and Discourse Accessibility

Gregory Ward[1] and Andrew Kehler[2]
[1]Northwestern University, [2]University of California San Diego

One of the central issues in studies of reference is the relationship between morphosyntactic form and the accessibility of discourse referents. However, most of the work in this area has been concerned primarily with reference to entities; considerably less work has addressed the relationship between syntactic form and the discourse accessibility of events. In this paper, we consider forms of event reference that involve what Bolinger (1972) termed 'identifier *so*', including the *do so* construction. In particular, we consider those referring expressions whose antecedents are not (syntactically-matching) VPs. In so doing, we respond to and counter a recent criticism of our previous work by Fu *et al.* (2001) and discuss several factors that appear to affect the accessibility of events evoked by nominalizations. Our account is based on a corpus of naturally-occurring and felicitous examples with *do so* that are apparently disallowed under the Fu *et al.* account.

1 Introduction

One of the central issues in studies of reference – be they theoretical, computational, or psycholinguistic – is the relationship between morphosyntactic form and the accessibility of discourse entities. For instance, work in theoretical linguistics concerned with COGNITIVE or INFORMATION STATUS (Chafe, 1976; Prince, 1981; Ariel, 1990; Gundel *et al.*, 1993; Lambrecht, 1994; *inter alia*) has attempted to account for the felicity of a particular referential form in a given discourse context and, in particular, the factors that affect the accessibility of referents in that context. As is well known, one such factor is the syntactic position in which a referring expression occurs, e.g. the oft-noted claim that entities referenced in subject position are typically more salient than those mentioned in positions that are lower on the obliqueness hierarchy (objects, arguments, adjuncts, etc.). Similarly, computational linguists have sought to determine the factors that contribute to accessibility as a basis for developing algorithms for pronoun resolution (Brennan, 1987; Lappin & Leass, 1994; Strube, 1998; Mitkov, 2002; *inter alia*). Finally, psycholinguists have used a variety of experimental methods to tease apart the seemingly competing factors that determine how people assign referents to pronouns, often with contradictory results (Crawley *et al.*, 1990; Smyth, 1994; Stevenson *et al.*, 1994; Chambers & Smyth, 1998; *inter alia*).

It is safe to say that most of the work in this area has been concerned primarily with reference to entities; considerably less work has addressed the relationship between syntactic form and the discourse accessibility of events. These two areas of inquiry differ in an important respect. When considering the accessibility of entities evoked by nominals, the primary concern has been the syntactic position in which that nominal occurs. In the case of events, however, one must also consider the effect of the particular syntactic form used to evoke the event itself: whether it was evoked by an active voice clause, a passive voice clause, or any of a variety of more marked sentential constructions, including gerunds and even full-fledged NP nominalizations. As we will see, this choice of form may also affect the accessibility of the event in question.

A comprehensive account that considers the relationship between the full range of syntactic forms available for evoking events in discourse and the full set of referring expressions available for subsequent reference to those events would take us well beyond the scope of this paper. Instead, we will focus here on a particular class of referring expressions: those that involve what Bolinger (1972) termed 'identifier *so*'. Examples of identifier *so*, used preverbally and as part of the *do so* construction, are shown in (1) and (2):

(1) "And with complete premeditation [they] resolved that His Imperial Majesty Haile Selassie should be strangled because he was head of the feudal system." He was *so strangled* on Aug. 26, 1975, in his bed most cruelly. (*Chicago Tribune* 12/15/94)

(2) As an imperial statute the British North America Act could be amended only by the British Parliament, which *did so* on several occasions. (*Grolier Encyclopedia*)

Previous accounts of *so* anaphora have treated it as a form of SURFACE ANAPHORA (Hankamer & Sag, 1976; *inter alia*) which requires the presence of an appropriate syntactic VP antecedent for its interpretation. In such a model, the issue of morphosyntactic form and relative accessibility simply does not arise: either there is a suitable VP antecedent available or there is not. However, examples like (2) are problematic for this view, since an active voice occurrence of *did so* is used felicitously to refer to an event evoked from a passive clause. As such, the active voice VP required by a surface anaphoric account – *amend the British North America Act* – does not occur in the prior discourse and is thus unavailable as an antecedent. In what follows, we will primarily be interested in antecedents of *so* anaphora that involve such mismatches.

In fact, in our previous work (Kehler & Ward, 1995, 1999) we have argued that *do so* does not qualify as a surface anaphor in that it imposes no restrictions on the particular syntactic form of its antecedent. We have argued instead that, like other forms of anaphora, *do so* is interpreted in terms of purely semantic referents within the hearer's discourse model. However, in a recent paper, Fu *et al.* (2001) take issue with this account and maintain that *do so* is in fact a surface anaphor. Moreover, they claim that the felicity of an antecedent expression in conjunction with the anaphor *do so* is a reliable diagnostic for the presence of a VP in that expression, and indeed use this diagnostic to argue for a particular syntactic analysis of process nominalizations.

In this paper, we extend our previous analysis and respond to Fu *et al.*'s (2001) proposal by demonstrating that compatibility with *do so* anaphora under closer examination does not provide evidence for their syntactic analysis. Indeed, using another class of nominalizations not addressed by Fu *et al.* (2001) (in addition to the evidence cited in (Kehler & Ward, 1999)), we show that reference with *do so* is sensitive to the relative salience of the evoked event, with the morphosyntactic form of the antecedent being but one of many factors that determine the felicity of such reference. As such, the question of when an event associated with a nominalization is sufficiently accessible for subsequent reference with identifier *so* is considerably more complex than the state of affairs described by Fu *et al.* (2001).

In the next section, we briefly review the account of *so* anaphora from (Kehler & Ward, 1995, 1999), which aims to provide a unified and compositional account of both preverbal *so* and *do so* despite their curious and idiosyncratic anaphoric properties. In Section 3, we discuss the alternative account put forth by Fu *et al.* (2001), and respond to and counter their criticism of our earlier account. We follow up this analysis in Section 4 with a discussion of a corpus of naturally-occurring and felicitous examples of *do so* with nominalized antecedents that are disallowed under the Fu *et al.* (2001) analysis. These examples bear a striking resemblance to acceptable examples involving so-called ANAPHORIC ISLANDS at the nominal level per the analysis of Ward *et al.* (1991); in both cases felicity of reference cannot be accounted for by appeal to morphosyntactic considerations alone. Among the variety of factors that appear to affect the accessibility of events evoked by such nominalizations, we discuss three that stand out in our data: semantic transparency, modification, and genericity.

2 Properties of Identifier So

Identifier *so* is used to refer to a contextually salient event of the type denoted by the verb it modifies.[1] It may appear in either preverbal or postverbal

[1] As such, none of the following uses of *so* are identifier:

- preposed propositional *so*: So it seems. So you say. So it is.
- postverbal propositional *so*: I think/suppose/say/believe so.
- veridical *so*: Is that so?
- consequential *so*: A: He's a pig. B: So you're **not** going out with him after all?
- particle *so*: So, how long have you been at Northwestern?

position, as illustrated in (3a–b) respectively, or in postverbal position as part of the *do so* construction, illustrated in (3c).

(3) Secretary Powell spent two months lobbying the United Nations very hard on Iraq.
  a. By *so lobbying*, he was able to say that the U.S. had at least tried to get a war resolution passed.
  b. By *lobbying so*, he was able to say that the U.S. had at least tried to get a war resolution passed.
  c. By *doing so*, he was able to say that the U.S. had at least tried to get a war resolution passed.

As we have argued elsewhere (Kehler & Ward, 1999), the two positional variants of identifier *so* display quite different properties. For example, only postverbal *so* permits exophoric reference, as shown in (4a–b):

(4) [Andy is holding a newborn baby with one hand behind his head, and shows Gregory]
  a. Andy: By *holding him so*, you add support to his developing neck muscles.
  b. Andy: #By so holding him, you add support to his developing neck muscles.
  c. Andy: #By *doing so*, you add support to his developing neck muscles.
  d. Andy: By *holding him this way*, you add support to his developing neck muscles.

Moreover, although it appears postverbally, the *so* of *do so* has precisely the same semantic and pragmatic properties of preverbal identifier *so*. For instance, as shown by (4b–c), both preverbal *so* and *do so* require that the referent event be LINGUISTICALLY EVOKED, that is, explicitly introduced in the discourse via a linguistic expression. In contrast, postverbal *so* is simply a manner adverbial anaphor and, like other such anaphors, permits situational evocation (compare 4a and 4d).

Another distinction between preverbal *so* and *do so* on the one hand and postverbal *so* on the other is that only the latter specifically requires an evoked manner. Consider again example (1), repeated below as (5).

(5) "And with complete premeditation [they] resolved that His Imperial Majesty Haile Selassie should be strangled because he was head of the feudal system." He was *so strangled* on Aug. 26, 1975, in his bed most cruelly.

Replacing *so strangled* with *strangled so* in this passage results in a decidedly odd reference, in the same way as the manner adverbial *in that way* does:

- 'queer' *so*: 'I've come to the conclusion,' he told me, 'that I'm not really "so" at all. I much prefer girls.' At this date the cant word among homosexuals for their proclivities was 'so'. (OED)
- additive *so*: Bill gave a speech and so did Hillary.
- intensifier *so*: There are **so** many uses of 'so'!
- generation X *so*: I'm like, so going out with him.
- and so on and so forth...

(6) ?? "And with complete premeditation [they] resolved that His Imperial Majesty Haile Selassie should be strangled because he was head of the feudal system." He was *strangled so/in that way* on Aug. 26, 1975, in his bed most cruelly.

This oddness presumably results from the fact that there is no evoked manner associated with the strangling event; modifying the passage to include one explicitly (e.g., "with a rope") results in perfect felicity for both postverbal *so* and the adverbial *in that way*:

(7) "And with complete premeditation [they] resolved that His Imperial Majesty Haile Selassie should be strangled with a rope." He was *strangled so/in that way* on Aug. 26, 1975, in his bed most cruelly.

Crucially, however, both *do so* and *so doing* pattern with preverbal *so* in not requiring an evoked manner:

(8) "And with complete premeditation [they] resolved that His Imperial Majesty Haile Selassie should be strangled because he was head of the feudal system." And they in fact *did so*, on Aug. 26, 1975, in his bed most cruelly.

(9) "And with complete premeditation [they] strangled Imperial Majesty Haile Selassie because he was head of the feudal system." By *so doing*, they forever changed the course of Ethiopian history.

Based on this evidence, we take preverbal *so* and the *so* of *do so* to be positional variants of the same event anaphor.[2] We discuss these two forms further in the sections that follow.

2.1 Preverbal so

Examples of the preverbal *so* construction are provided in (10)–(12).

(10) "And with complete premeditation [they] resolved that His Imperial Majesty Haile Selassie should be strangled because he was head of the feudal system." He was *so strangled* on Aug. 26, 1975, in his bed most cruelly. (=1)

(11) In fact, in substantiating these fears, Judge Bork again essentially concedes that economic freedom is a component of the Constitution: "We already have clauses that could be used to protect economic freedom – and were *so used*." (*Wall Street Journal*)

(12) In fact, it is interesting that, in English, at least, there is virtually no marking of an NP with respect to the Discourse-status of the entity it represents. Of course, if an NP is indefinite and is thereby understood as evoking something Hearer-New, we can infer Discourse-New. However, if it is not *so marked*, then, with one exception, we cannot tell from its form whether it has occurred before in the discourse. (Prince, 1992:304)

---

[2] This is not to say that preverbal *so* and *do so* are interchangeable; the former is considerably more restricted than the latter. There appears to be a very general constraint at play in that the referent has to be more specific than the denotation of the verb modified by *so*, although this need not be a manner as with postverbal *so*. Hence, example (5) would be odd without the appearance of the because clause. In the case of *do so*, however, this specificity constraint is met automatically, since any event is more specific than a generic 'doing'.

The information status of the referent event associated with the preverbal *so* construction is constrained in a number of ways. First, the referent event associated with preverbal *so* must be SALIENT in the mental model of the hearer; compare (13) with (10):

(13) "And with complete premeditation [they] resolved that His Imperial Majesty Haile Selassie should be strangled because he was head of the feudal system." They also resolved to commit a variety of other violent acts, although those would come somewhat later. Obviously, these people were very prone to violence.
  a. #Selassie was *so strangled* on Aug. 26, 1975, in his bed most cruelly.
  b. Selassie was strangled on Aug. 26, 1975, in his bed most cruelly.

The substantial material that intervenes between the first mention of the strangling event and the subsequent reference to it in (13a) renders the reference infelicitous. Note that the salience criterion implies that the event must be HEARER-OLD in the sense of Prince (1992); that is, felicitous use of preverbal *so* requires that the speaker have evidence that the hearer is familiar with the event at the time of the utterance.

Second, as we have already argued, the referent must be linguistically evoked; that is, reference with identifier *so* cannot be exophoric, as shown by (14) (in addition to (4b–c)):

(14) [A and B together have just witnessed Haile Selassie being murdered by strangulation]
  a. A: #He was *so strangled* most cruelly.
  b. A: He was strangled most cruelly.

Finally, as with other anaphoric expressions, the referent event associated with preverbal *so* may be INFERRABLE in the sense of Prince (1981, 1992):

(15) Regarding a possible Elvis Presley stamp, Postmaster General Frank notes that anyone *so honored* must be "demonstrably dead" for 10 years. (*Wall Street Journal*)

Here, there is no honoring event that is explicitly introduced into the discourse. Instead, a hearer must reason from the mention of *a possible Elvis Presley stamp* that putting someone's picture on a stamp constitutes a kind of honoring event. Although the chain of reasoning required for this interpretation is quite complex, reference to the inferred event with *so* is fully felicitous.

2.2 Do so

As argued above, preverbal identifier *so* is related to the *do so* construction, illustrated in (16):

(16) Sam sold his stock on insider information, and Martha *did so* too.

Previous accounts of this intensively investigated construction have noted its seemingly idiosyncratic syntactic and anaphoric properties (Lakoff & Ross, 1966; Anderson, 1968; Bouton, 1970; Halliday & Hasan, 1976; Hankamer & Sag, 1976; Sag & Hankamer, 1984; Quirk *et al.*, 1985; Miller, 1990; Ward *et al.*, 1991; Cornish, 1992; Fu & Roeper, 1993; Dechaine, 1994; Fu *et al.*, 2001). Here we will only discuss those properties of the construction that bear directly on our analysis.

First, it is clear that the *do* of *do so* is main verb *do* and not auxiliary *do* (Hankamer & Sag, 1976; Sag, 1976; Quirk *et al.*, 1985; Miller, 1990; Dechaine, 1994; *inter alia*).

(17)a. Hillary did so.
b. *Did Hillary so?
c. Did Hillary do so?
d. Hillary did.
e. Did Hillary?

As (17a–e) show, the main verb *do* of *do so* does not undergo auxiliary inversion, unlike auxiliary *do* (Miller, 1990). Likewise, the main verb *do* permits *so*, unlike auxiliary *do*:

(18)a. Dubya filed a lawsuit, and Al did too. [auxiliary *do*]
b. Dubya has filed a lawsuit, and Al has too.
c. Dubya will file a lawsuit, and Al will too.

(19)a. Dubya filed a lawsuit, and Al did so too. [main verb *do*]
b. *Dubya has filed a lawsuit, and Al has so too.
c. *Dubya will file a lawsuit, and Al will so too.

Here, we see that it is the main verb *do*, and not the auxiliary form of *do*, that co-occurs with *so*. Furthermore, *do so* (and its participial variant *so doing*) is more restricted in its use than auxiliary *do* (Lakoff & Ross, 1966; Anderson, 1968; Bouton, 1970; Hankamer & Sag, 1976; Quirk *et al.*, 1985; Miller, 1990; Dechaine, 1994; *inter alia*).

(20)a. Al knows French, and Tipper does too. [auxiliary *do*]
b. Al knows French, and so does Tipper. [auxiliary *do*]
c. #Al knows French, and Tipper does so too. [main verb *do*]
d. #Al knows French, and in so doing, is not popular with Republicans. [main verb *do*]

In these examples, we see that a strongly stative verb like *know* permits auxiliary *do* with ellipsis, but disallows main verb *do*.[3]

As for the status of the *so* of *do so*, one might be tempted to categorize it as an NP given its superficial similarity to other event anaphors, such as *do it* and

[3] Hankamer and Sag (1976) suggest that the key distinction is between active (nonstative) antecedents (which are said to compatible with *do so*) and stative VP antecedents (which are not), but as pointed out by Quirk *et al.* (1985) and Dechaine (1994), a more fine-grained categorization is required. We thank an anonymous reviewer for helpful comments with respect to this issue.

do that. However, as others have noted (Bouton, 1970; Hankamer & Sag, 1976; Quirk *et al.*, 1985), there is strong distributional evidence that the *so* is categorially an adverb. First, note that it does not passivize like NPs do:

(21)a. *...and so was done by Hillary.
b. ...and it was done by Hillary.
c. ...and that was done by Hillary.

Second, unlike NPs, it does not cleft:

(22)a. It is that which Hillary did.
b. *It is so which Hillary did.
c. What Hillary did was that.
d. *What Hillary did was so.

Given these facts, we conclude that the *do* of *do so* is an intransitive main verb and the *so* is an adverbial modifier.

In the analysis presented in (Kehler & Ward, 1999), all of the previously discussed properties of *do so* are captured by a compositional account in which the *do* of *do so* denotes the most semantically general type of event, and the *so* marks the information status of that event as both salient and discourse-old, the latter of which excludes situationally-evoked referents. Seen in this way, *do so* (and its variant *so doing*) are simply forms of standard hyponymic reference (cf. (Miller, 1990)), as can be seen by considering the following progression of examples:

(23) The hit man dispensed with his mob boss by shooting him in broad daylight, with plenty of witnesses around.
a. By *so shooting him*, the hit man established himself as his victim's likely successor.
b. By *so murdering him*, the hit man established himself as his victim's likely successor.
c. By *doing so*, the hit man established himself as his victim's likely successor.

Continuations (23a–c) illustrate *so* anaphora using the same verb (*shooting*), a more general hyponym (*murdering*), and ultimately the most general hyponym (*doing*), respectively. This progression from specific to general event type directly parallels the situation for reference to entities with nominal anaphors, as illustrated by the different referential options given in (24).

(24) Chris took his poodle to the vet. The poodle / the dog / the animal was in a lot of pain.

Given this analysis, we can account for the fact that *do so* is restricted in its range of reference: stative events like *know* and *own* are simply not 'doings' as they do not involve agency (cf. 20a–d). Thus, as a form of hyponymic reference to a general type of doing event, use of *do so* is incompatible with such predicates. In this way, it parallels its nominal counterparts *do it* and *do that*, although for those referring expressions the constraint arises from semantic restrictions that transitive *do* places on its direct object.

Moreover, this analysis accounts for the fact that *do so* is not a surface anaphor in the sense of Hankamer and Sag (1976), i.e. that it does not require an antecedent of any particular syntactic form. Instead, *do so* again patterns referentially like its nominal counterparts *do it* and *do that* in that they are all used to refer to events in (the speaker's representation of) the hearer's mental model of the discourse.

Indeed, our account correctly predicts the existence of mismatches between the morphosyntactic form of the anaphor *do so* and that of its antecedent. In what follows, we briefly present several classes of naturally-occurring examples of anaphor-antecedent mismatches. The first class consists of mismatches between the voices used each clause; consider (25–26):

(25) Section 1 provides the examples to be derived by Gapping, and a formulation of Gapping capable of *doing so*. [= deriving the examples] (text of Neijt (1981))

(26) As an imperial statute the British North America Act could be amended only by the British Parliament, which *did so* on several occasions. [= amended an imperial statute] (=2)

In these examples, the passive voice of the antecedent sentence does not match the active voice of the anaphor; under a surface anaphoric account of *do so*, such mismatches are predicted to be ill-formed. Likewise, in our second class of examples (27–28), the antecedent expression has been nominalized, and hence (under most accounts; see Section 3) there is not the requisite VP antecedent available as required by a surface anaphoric account:

(27) The defection of the seven moderates, who knew they were incurring the wrath of many colleagues in *doing so*, signaled that it may be harder to sell the GOP message on the crime bill than it was on the stimulus package. [= defecting] (*Washington Post*)

(28) Even though an Israeli response is justified, I don't think it was in their best interests to *do so* right now. [= respond] (token provided by Dan Hardt)

Finally, in examples (29–30) the antecedents are 'split':

(29) Patients who view these discussions or who ask questions must do so at their own risk. [= view these discussions / ask questions][4]

(30) Fortunately, the first person to die in 1990 and the first couple to file for divorce in 1990 were allowed to *do so* anonymously. [= die / file for divorce] (text of Roeper (1990), cited by J. McCawley's "1990 Linguistic Flea Circus" and discussed by Dalrymple *et al.* (1991))

Again, we see that the conjoined VP required under a surface-anaphoric account of *do so* is not available here; a (presumably quite dubious) cut-and-

[4] www.hsforum.com/listdisclaim

paste operation would be necessary to combine two VPs from separate sentences and add the appropriate connective (i.e. *or* in (29) and *and* in (30)) to arrive at an antecedent of an appropriate form.

To summarize thus far, *do so* is a compositional anaphoric construction consisting of intransitive main verb *do*, denoting the most general of event types, and identifier *so* marking the information status of that event. As our corpus of naturally-occurring data reveals, these constructions are not restricted to antecedents of a particular syntactic form.

3 A Recent Syntactic Proposal

Despite the existence of such mismatches, Fu *et al.* (2001), henceforth FRB, attempt to resurrect the notion that *do so* is a surface anaphor. While aware of the problems for such an analysis raised by our earlier work (Kehler & Ward, 1995), they maintain that *do so* categorically requires a syntactic VP antecedent, and use this assumption to argue that PROCESS NOMINALIZATIONS such as those in (27–28) contain a VP in their syntactic representations.[5] While space does not allow us to present their analysis in detail, we will briefly discuss three problems that we see with their argumentation and the conclusions they draw from it.

3.1 Comparison with 'Deep Anaphora'

FRB cite the contrast between (31–32) as evidence that *do so* requires a underlying syntactic VP as an antecedent:

(31) His removal of the garbage in the morning and Sam's *doing so* in the afternoon were surprising. (= FRB's 42b)

(32) *Kim's accident in the morning and Sue's *doing so* in the evening were not coincidences. (= FRB's 43b)

Both *removal* in (31) and *accident* in (32) are event-denoting, but only (31) is acceptable with *doing so*. FRB conclude that the difference must therefore be syntactic: the nominalization *removal* must contain a VP in syntax to license *do so*, whereas *accident*, which is not a nominalization, does not.

If this is the reason for the contrast, however, then the contrast should disappear when *do so* is replaced by an indisputably non-surface anaphor such as *do it*, which imposes no requirement for a syntactic VP antecedent. However, the judgments in fact stay the same with this replacement:

[5] Process nominalizations are deverbal nouns that denote an event of the type associated with the nominalized verb. For instance, *defection* in (27) denotes a 'defect' event and likewise *response* in (28) denotes a 'respond' event. Process nominalizations are thus distinguished from RESULT nominalizations (e.g., *invention*, used to refer to an object itself and not the act of inventing it), and ROLE nominalizations (e.g., *murderer* denotes the agent of a murder event and *employee* denotes the theme of an employ event; see below).

(33) His removal of the garbage in the morning and Sam's *doing it* in the afternoon were surprising.
(34) *Kim's accident in the morning and Sue's *doing it* in the evening were not coincidences.

Thus, while there is a contrast in event accessibility between nominalizations and event-denoting nouns, this contrast cannot be a direct result of a putative requirement that there be a VP antecedent for *do so*. If that were the case, deep anaphors like *do it* would not show precisely the same contrast. Thus, the distinction between (31) and (32) offers no evidence (one way or the other) bearing on the issue of whether nominalizations contain a VP in syntax.

3.2 Non-Process Nominalizations

By design, FRB's analysis applies to process nominalizations, which presumably excludes other types such as role nominalizations. But a corpus search reveals many examples of felicitous, naturally-occurring examples of precisely this type:

(35) One study suggests that almost half of young female smokers *do so* in order to lose weight. [= smoke][6]
(36) The majority of horse riders *do so* purely for leisure and pleasure. [= ride horses][7]
(37) AmericaNet.Com, its officers, directors or employees are not responsible for the content or integrity of any ad. Sellers/buyers/subscribers/investors *do so* at their own risk. [= sell/buy/subscribe/invest][8]
(38) Data from the Retirement Survey reveals that 5% of early retirees *do so* because of the ill health of others. [= retire early][9]

In response, FRB could conceivably argue that role nominalizations also have an underlying VP structure. But for many such cases, such an analysis is simply untenable:

(39) # My computer *does so* faster than yours. [= compute]
(40) # The boat's propeller failed to *do so*, and now we're stuck. [= propel]

The problem with this aspect of FRB's analysis is that their distinction is a categorical one: a VP is either present in the syntax or it is not. Moreover, we know of no independent evidence to the effect that some role nominals incorporate VP syntax and others do not, nor do we believe that such evidence exists. However, what we do find is that role nominalizations display a clear gradience with respect to compositionality, and we will argue in Section 4 that

[6] www.ustrek.org/odyssey/semester1/111800/111800madwomen.html
[7] www.league.uk.com/news/media_briefings/2002/may_2002/17_may_02_a_livery_yard.htm
[8] www.americanet.com/Classified/sendad.html
[9] www.npi.org.uk/reports/Active_Ageing.pdf

this compositionality in part determines the accessibility of the referenced event.

3.3 Other Syntactic Mismatches

As we observed earlier in Section 2.2, the data that challenge the notion that *do so* requires a syntactic antecedent are not limited to cases involving nominalized antecedents. For instance, several previously cited examples – (25) and (26) – involve syntactic mismatches in which there is no plausible VP antecedent. Consider (2), repeated below as (41):

(41) As an imperial statute the British North America Act could be amended only by the British Parliament, which *did so* on several occasions. (=2)

About this particular example, FRB say:

> it is not clear that it seriously jeopardizes the claim that *do so* requires a VP/V′ antecedent...That in the first conjunct the direct object is occupied by a trace, rather than a full NP may very well turn out to be immaterial for the licensing of the anaphor *do so*. (2001:572–573)

However, FRB do not provide the details necessary to adequately evaluate this possibility and its ramifications.[10] Further, they never make explicit their assumptions about how *do so* is interpreted in the first place. The question requires attention because in a standard surface anaphoric theory (Hankamer & Sag, 1976), the requirement for a matching syntactic antecedent results from a deletion process that applies during production (or, alternatively a reconstruction process that applies during interpretation). As such, it needs to be explained why a form like *do so* would impose a syntactic requirement in light of the fact that it is not associated with an ellipsis site.

The only argument FRB provide in support of their response to the well-formedness of example (41) is an alleged contrast with adjectival antecedents, which, they claim, are presumably worse because they do not involve a trace. Consider (42):

(42) ?? This act turned out to be amendable, and the British Parliament *did so* in its last session.

But a considerably more acceptable variant is readily constructed:

[10] Their phrase "may very well turn out" suggests that these details have not been worked out by them, nor will we attempt to do so here. In any case, it is certainly not obvious to us how such an analysis could be made to be consistent with both FRB's goals and the relevant facts in the literature on ellipsis and event reference, including constructions such as antecedent-contained deletion that require that constituents be reconstructed with traces left intact.

(43) After the British Parliament found out that the act was amendable, they elected to *do so* at their very first opportunity.

Thus, although more details about their syntactic account will be necessary before it can be fully evaluated, we find it doubtful that such account can be made to handle the relevant set of facts through independently-motivated mechanisms.[11]

FRB also do not address cases in which *do so* is felicitous with a split antecedent – such as examples (29), (30), and (37) – in which a suitable antecedent is not available:

(44) Fortunately, the first person to die in 1990 and the first couple to file for divorce in 1990 were allowed to *do so* anonymously. (=30)

Again, it is hard to see how such examples could be accommodated in a purely syntactic treatment. On the other hand, our analysis predicts the patterning of such reference with pronominal reference, which is likewise compatible with split antecedents:

(45) The first person of the year to die is usually listed in the newspaper, and so is the first couple to file for divorce. In a rare show of respect, this year *their* names were kept private.

As it is clear that pronouns are not surface anaphors, FRB's analysis fails to capture the analogous behaviour of the two forms.

Lastly, FRB criticize us for failing to note an alleged parallel between *do so* anaphora and adverbial modification, based on examples such as the following:

(46)a. Kim's explanation of the problem to the tenants thoroughly (did not prevent a riot). (= FRB's 1a)
b. The occurrence of the accident suddenly (disqualified her). (= FRB's 2a)

However, the vast majority of our informants reject these sentences out of hand, on the relevant readings in which the adverb modifies the nominalization and not the matrix verb. Therefore, we consider the fact that our account fails to establish such a parallel to be a feature of the analysis, and not a drawback. On the other hand, we find that FRB's analysis fails to capture a different generalization: that the *so* in *do so* is the same (identifier) *so* found in the productive *so*+V construction, which, as we pointed out earlier, clearly does not require a syntactically-matching antecedent (cf. 15).

---

[11] FRB cite other examples in which deverbal adjectives are unacceptable as antecedents of *do so*, concluding that "the slight improvement of the –*able* cases [i.e. (42)] may have to do with whether the verbal meaning is preserved" (573, fn. 24). We agree, and take such data to provide evidence for our analysis (see Section 4, where we make similar arguments about felicitous cases involving role nominalizations), whereas such gradations in the data appear to call into question the categorical predictions made by FRB's analysis.

4 Nominalizations, Accessibility, and Anaphoric Islands

One of the conclusions we reached in the last section is that the sharp line drawn by FRB's account between process and nonprocess nominalizations – the former which are claimed to be possible antecedents for *do so* and the latter which are not – does not reflect the distribution of felicitous reference one actually finds in the data. Instead, what we have found through our corpus study is that there are examples of both types that range from perfectly felicitous to totally unacceptable. An adequate account will therefore have to allow either type of nominalization to serve as an antecedent *in principle* – as our analysis does – and will accordingly have to explain the gradient felicity of the data through other means. In our discourse-based analysis, the felicity of *do so* with a nominalized antecedent boils down to the question of whether the nominalization renders its underlying event sufficiently salient, and not whether a syntactic VP antecedent exists in the context.

The debate between syntactically-based and discourse-based analyses of *do so* is strikingly reminiscent of a previous debate in the literature concerning the existence of so-called ANAPHORIC ISLANDS as they pertain to nominal-level reference. Arguing against the existence of a purely structural constraint prohibiting reference to entities evoked from "word-internal" positions (Postal, 1969), Ward *et al.* (1991) found that reference to such entities is indeed possible under the right pragmatic conditions, providing numerous naturally-occurring examples such as (47) and (48).

(47) Although casual *cocaine* use is down, the number of people using *it* routinely has increased. (= Ward *et al.*'s (1991) (22a))

(48) Patty is a definite *Kal Kan* cat. Every day she waits for *it*. (= Ward *et al.*'s 20b)

An analysis of a corpus of naturally-occurring uses of *do so* with role-nominalized antecedents (including examples for over 25 different verbs) shows that such reference patterns directly with the anaphoric island data discussed by Ward *et al.* (1991). While various pragmatic factors may conspire to render a particular use of either an entity anaphor (e.g. pronouns) or an event anaphor (e.g. *do so*) infelicitous, those same factors in another context can also permit such usages, and thus in neither case can their occurrence be ruled out by syntactic considerations alone.[12]

The most striking commonality between anaphoric island violations at the nominal level and reference to role-nominalized events with *do so* is the central

---

[12] Ward *et al.* (1991) ultimately drew a different conclusion about *do so* anaphora, stating that "it follows that no discourse context will render *do so* anaphora felicitous with non-VP antecedents" (p. 462), a conclusion not supported by the current corpus-based study.

role of the SEMANTIC TRANSPARENCY (Ward *et al.*, 1991) or ANALYZABILITY (Langacker, 2000) of the antecedent. Langacker defines analyzability as "the extent to which speakers are cognizant of the presence and the semantic contribution of component symbolic elements" (2000:127). To use his example, if we were to coin the term *flinger*, the hearer must use the meanings of its morphemes to derive the word meaning as "something that flings"; thus it is fully analyzable. On the other hand, the present-day meanings of *computer*, *freezer*, and *propeller* go well beyond the meanings "something that computes/freezes/propels", to the point where the corresponding underlying events almost certainly receive considerably less activation upon mention. Indeed, the underlying events for *ruler*, *pliers*, and *plumber* probably do not get activated at all by the mere mention of these words. As noted by various authors (Aronoff, 1976; Bauer, 1983; Langacker, 2000), there is a long-term tendency for words to lose their analyzability as they gain conventionalized meanings.[13] We would therefore expect role nominalizations to activate their underlying event representations to varying degrees – very much so for *flinger* and perhaps not at all for *ruler* – and thus the acceptability of using *do so* to refer to such underlying events would vary accordingly.[14]

This pattern appears to be borne out by our corpus. Examples (49)–(52) demonstrate felicitous uses of *do so* with highly transparent role nominalizations:

(49) Although most collectors *do so* for the sheer fun of the hobby, the question "What's the value of my collection?" does arise. [= collect][15]

(50) Most antler hunters *do so* recreationally, says Mr. Hovinga. [= hunt antlers][16]

(51) Residents should include contact information so that respondents may do so directly. [= respond][17]

(52) Users of information from any Applied Discovery web site *do so* at their own risk. [= use information from any Applied Discovery web site][18]

That is, *collectors* are people who collect; *users of information from any Applied Discovery web site* are people who use information from any Applied Discovery web site. On the other hand, our searches yielded no comparable examples with the nominalizations *computers, freezers, propellers*, or others

[13] As Langacker notes, the effect is similar to the fading of metaphors, in which ultimately speakers are unaware of the metaphorical basis for a word or phrase.

[14] Note, therefore, that the felicity of *do so* anaphora for a given nominalization might actually change over time, if that nominalization is evolving toward a conventionalized meaning.

[15] www.leuchtturm.com/en/prod/en_overview.htm

[16] www.sublette.com/scj/v4n26/v4n26s2.htm

[17] www.artists-in-residence.com/parlor/messages/9/9.html?0

[18] www.applieddiscovery.com/termsConditions/default.asp

that are similar in terms of their (low) degree of semantic transparency. That is, we found no cases like (53) and (54) despite our best efforts.

(53) #Most computers *do so* quickly these days.
(54) #Today's boat propellers *do so* with great force.

In sum, the transparency of a nominalization affects the extent to which the representation of the underlying event is activated, which in turn determines the extent to which this representation can be accessed with an event referential form such as *do so*.

If accessibility is indeed the operative notion for governing the felicity of *so* anaphora, then we would expect such felicity to be influenced by other factors that affect accessibility, perhaps in more subtle ways. One such factor is the effect of modifiers on a role nominalization. Adjectival modifiers of role nominalizations, for instance, can describe properties of either the entity denoted by the nominalization or the underlying event. Consider the following example from our corpus, in particular the last sentence:

(55) Gulden ignored a race official and jumped the tape marking the finish area to shake hands with his runners. He was the only coach in the shutes, a coach whose instincts have always told him this moment is important. After 25 years, he is remarkably more tenacious than ever. Other coaches show up at meets in jacket and tie, assigning their assistants to points on the course. Gulden shows up in sweats and puts in a few thousand meters himself, running from point to point. The greatest teachers *do so* by example.[19]

Here, the adjective is used to modify the underlying teaching event: *greatest teachers* are people who teach the greatest. On the other hand, a *tall teacher* describes a teacher who is tall, and not someone who teaches in a tall way. As such, we see a distinction in the accessibility of the event depending on whether the accompanying adjective modifies it or not:

(56)a. The greatest teachers *do so* by example.
b. ?The tallest teachers *do so* by example.

It would appear that the adjective in (56a) increases the accessibility of the teaching event enough to support subsequent reference with *do so*, whereas *tall* in (56b) causes attention to be placed on the teacher as an entity, thereby reducing the accessibility of the underlying event.[20]

Finally, an examination of our corpus suggests that another factor affecting the accessibility of events is the genericity of the event in question. That is, we

[19] www.departments.bucknell.edu/pr/BucknellWorld/1995-1/gulden.html
[20] It is worth noting that even though *teachers* in (56a) is a suitable antecedent for *do so*, it clearly does not allow the type of adverbial modification that FRB would predict if their analysis were extended to role nominalizations: **The teacher of the students greatly (is tall).*

have found that generic (or quantified) role nominalizations evoke events that are more accessible than those evoked by non-generic (or non-quantified) nominalizations. We hypothesize that the reason for this preference is because generics are typically used in situations in which the underlying event plays a key role in the main assertion of the sentence, hence highlighting the event. For instance, the pragmatic force of a sentence like (57a) centers on the fact that the people John has to dine with smoke. While non-generic nominalizations can also be used this way (57b), they also commonly serve other purposes in which the event is less central, such as to merely single out a referent of an NP amongst alternatives (57c).

(57)a. John often has to dine with smokers.
b. John had to dine with a smoker yesterday – poor John.
c. John dined with that smoker over there yesterday.

As such, a non-generic role nominalization may not create the same expectation for the centrality of the event that a generic does. While our searches cannot be considered exhaustive, our preliminary analysis suggests that non-generic examples are fewer, although importantly they do exist:

(58) While it is certainly sad that he died early, it has no lasting eternal consequences. This is quite the opposite for the killer, however. Assuming that the killer *did so* maliciously and not accidentally, there are serious eternal consequences for his act. [= kill][21]

What these three factors – semantic transparency, modification, and genericity – have in common is that they all affect the extent to which the event that underlies the use of a role nominalization becomes activated and accessible during semantic interpretation. This, in turn, affects the degree to which that event is available for subsequent reference with an event anaphor. We suspect that there are other such factors as well for which we would make the same prediction: the more accessible the underlying event, the more felicitous subsequent reference to it using *do so* will be. These findings, taken together with the results of Ward *et al.*'s (1991) analysis of anaphoric islands, indicate that both entity and event reference are governed by essentially pragmatic – and not morphosyntactic – factors.

5 Conclusions and Future Research

We have shown that referential forms that involve identifier *so* – including the *do so* construction – do not impose any purely syntactic restrictions upon their antecedents. Instead, like other event-referential expressions, they are used to refer to events that have been evoked in the hearer's discourse model. The

[21] www.frontpage2000.nmia.com/~nahualli/LDStopics/Theology/4school.htm

broad range of syntactic forms that can license identifier *so* anaphora – including role nominalizations – seriously calls into question the claim of FRB that *do so* is a diagnostic for the existence of an underlying verb phrase in the syntactic representation of process nominalizations.

On the other hand, there is a connection between syntactic form and discourse accessibility and, in turn, between discourse accessibility and felicity of reference with *do so*. We have provided a (non-exhaustive) set of factors that affect the accessibility of events underlying a class of antecedents that are disallowed by FRB's approach – role nominalizations – which in certain circumstances allow subsequent reference with *do so*. However, many important questions remain concerning the role of these and other (as yet unidentified) factors in determining the salience of events evoked by various linguistic forms. Whatever those factors ultimately turn out to be, we hope to have demonstrated the crucial role of discourse factors in the interpretation of event anaphora.

References

Anderson, S. 1968. *Pro-sentential Forms and Their Implications for English Sentence Structure*. Technical Report NSF-20, Aiken Computation Laboratory, Harvard University.

Ariel, M. 1990. *Accessing Noun Phrase Antecedents*. London: Routledge, Croom Helm.

Aronoff, M. 1976. *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.

Bauer, L. 1983. *English Word-Formation*. Cambridge: Cambridge University Press.

Bolinger, D. 1972. *Degree Words*. The Hague and Paris: Mouton.

Bouton, L. 1970. "Do So: Do + Adverb." *Studies Presented to R. B. Lees by his Students*. J. Sadock (ed.), 17-38. Edmonton: Linguistics Research.

Brennan, S., M. Friedman and C. Pollard 1987. "A Centering Approach to Pronouns." *Proc. of the 25th Meeting of the ACL (ACL'87)*. Stanford, CA. 155-162.

Chafe, W. 1976. "Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View." *Subject and Topic*. C. Li (ed.), 25–55. New York: Academic Press.

Chambers, C. and R. Smyth 1998. "Structural Parallelism and Discourse Coherence: A Test of Centering Theory." *Journal of Memory and Language* 39.593-608.

Cornish, F. 1992. "*So Be It*: The Discourse-Semantic Roles of *So* and *It*." *Journal of Semantics* 9.163-178.

Crawley, R., R. Stevenson and D. Kleinman 1990. "The Use of Heuristic Strategies in the Interpretation of Pronouns." *Journal of Psycholinguistic Research* 19.245-264.

Dalrymple, M., S. Shieber and F. Pereira 1991. "Ellipsis and Higher-Order Unification." *Linguistics and Philosophy* 14.399-452.

Dechaine, R.-M. 1994. "Ellipsis and the Position of Subjects." *Proc. of NELS 24*, 47-63.

Fu, J. and T. Roeper 1993. "The VP within Nominalizations: Evidence from Adverbs and the VP Anaphor *Do-So*." Paper presented at the Syntax-Morphology Connection, MIT.

----------, ---------- and H. Borer 2001. "The VP within Process Nominals: Evidence from Adverbs and the VP Anaphor *Do-So*." *Natural Language and Linguistic Theory* 19.549-582.

Gundel, J., N. Hedberg and R. Zacharski 1993. "Cognitive Status and the Form of Referring Expressions in Discourse." *Language* 69.274-307.

Halliday, M. and R. Hasan 1976. *Cohesion in English*. English Language Series, Title No. 9. London: Longman.

Hankamer, J. and I. Sag 1976. "Deep and Surface Anaphora." *Linguistic Inquiry* 7.391-426.

Kehler, A. and G. Ward 1995. "On the Anaphoric Status of *Do So*." Paper presented at the Annual Meeting of the Linguistic Society of America, New Orleans.

---------- 1999. "On the Semantics and Pragmatics of 'Identifier *So'*." *The Semantics/Pragmatics Interface from Different Points of View (Current Research in the Semantics/Pragmatics Interface Series, Volume I)* ed. Ken Turner, 233-256. Amsterdam: Elsevier.

Lakoff, G. and J. Ross 1966. "Criterion for Verb Phrase Constituency." Technical Report NSF-17, Aiken Computation Laboratory, Harvard University.

Lambrecht, K. 1994. *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.

Langacker, R. 2000. *Grammar and Conceptualization*. Berlin: Mouton De Gruyter.

Lappin, S. and H. Leass 1994. "An Algorithm for Pronominal Anaphora Resolution." *Computational Linguistics* 20.535-561.

Miller, P. 1990. "Pseudogapping and *Do So* Substitution." *Papers from the 26th Regional Meeting of the Chicago Linguistic Society*, 293-305.

Mitkov, R. 2002. *Anaphora Resolution*. London: Longman.

Neijt, A. 1981. "Gaps and Remnants – Sentence Grammar Aspects of Gapping." *Linguistic Analysis* 8.69-93.

Postal, P. 1969. "Anaphoric Islands." *Papers from the 5th Regional Meeting of the Chicago Linguistic Society*, 205-239.

Prince, E. 1981. "Toward a Taxonomy of Given-New Information." *Radical Pragmatics*. P. Cole (ed.), 223-255. New York: Academic Press.

---------- 1992. "The ZPG Letter: Subjects, Definiteness, and Information-Status." *Discourse Description: Diverse Analyses of a Fundraising Text*. S. Thompson and W. Mann (eds.), 295-325. Amsterdam & Philadelphia: John Benjamins.

Quirk, R., S. Greenbaum, G. Leech and J. Svartvik 1985. *A Comprehensive Grammar of the English Language*. New York: Longman.

Sag, I. and J. Hankamer 1984. "Toward a Theory of Anaphoric Processing." *Linguistics and Philosophy* 7.325-345.

Smyth, R. 1994. "Grammatical Determinants of Ambiguous Pronoun Resolution." *Journal of Psycholinguistic Research* 23.197-229.

Stevenson, R., R. Crawley and D. Kleinman 1994. "Thematic Roles, Focus and the Representation of Events." *Language and Cognitive Processes* 9.519-548.

Strube, M. 1998. "Never Look Back: An Alternative to Centering." *Proc. of the 17th Int. Conference on Computational Linguistics and the 36th Annual Meeting of the ACL* (Coling-ACL '98), Montreal, Canada.

Coreference and Anaphoric Relations of Demonstrative Noun Phrases in Multilingual Corpus

Renata Vieira[1], Susanne Salmon-Alt[2] and Caroline Gasperin[1]
[1]UNISINOS São Leopoldo, [2]ATILF – CNRS Nancy

We present a corpus study concerning the use of demonstrative noun phrases in Portuguese and French. The motivation for this study is to verify specific features related to the coreferential and anaphoric role of such expressions in written texts. These features serve as background knowledge for the development of a multilingual tool for coreference and anaphoric resolution.

1 Introduction

Recent work on anaphora resolution is pointing to the fact that different types of referring expressions (pronouns, definite descriptions, demonstratives) are based on different features or require different knowledge for reference resolution (Strube *et al.*, 2002; Sant'Anna & Lima, 2002; Salmon-Alt & Vieira, 2002; Poesio *et al.*, 2002).

In this work, motivated by rising background knowledge for the design of a multilingual tool for anaphora resolution, we analyse in detail syntactic, discourse and semantic features specifically related to the use of demonstrative noun phrases. As primary data, we use Portuguese and French corpora of written texts.

Section 2 defines the main concepts (*coreference*, *anaphora* and *demonstrative noun phrases*) used in this study. Section 3 provides a detailed overview of the features we investigated. Section 4 describes the annotation task, the corpora and the annotation tool. A discussion of the results is presented in Section 5, and Section 6 discusses conclusions and future work.

2 Coreference and anaphoric relations of demonstrative noun phrases

According to related work on demonstratives in the area of descriptive linguistics (Corblin, 1987), demonstrative noun phrases are considered to be interpreted on the basis of the salience of the referent. A referent can be salient, for example, because of a pointing gesture or a previous mention. The fact that salience based on pointing gestures is excluded in our corpus study of written discourse implies that the interpretation of demonstratives tend to be more closely related to previous text, as this is the only source of salience.

Bearing this in mind, we designed a corpus study focusing on coreference and anaphoric relations of demonstrative noun phrases. Coreference has been defined by van Deemter and Kibble (2000) as the relation holding between linguistic expressions that refer to the same extra-linguistic entity. A slightly different discourse relation is anaphora. In an anaphoric relation, the interpretation of an expression is dependent on previous expressions within the same discourse, but the anaphor and its antecedent may refer to different referents. Therefore, an anaphoric relation may be coreferential or not, and as it is known, a particularly difficult question is to determine the relation holding between the anaphor and its antecedent. (Strand, 1996; Vieira & Teufel, 1997; Poesio & Vieira, 1998).

An expression may be anaphoric in the strict sense that its interpretation is only possible on the basis of the antecedent, as it is in general the case of pronouns in written discourse. On the other hand, it might be coreferential without being anaphoric, in the sense that the entity has been mentioned before in the text, as it is the case of subsequent mentions of self explaining expressions such as *the champion of the world cup – the team that won the world cup championship*.

In this work, we are interested in both coreferential and anaphoric relations. The analyses have been made regarding (i) several features of the textual antecedents of given expressions, such as verifying whether the antecedent is coreferential or not; (ii) its syntactic structure; and (iii) certain semantic properties.

In this study, we consider demonstrative noun phrases (NPs) in Portuguese and French. These are noun phrases starting with a demonstrative determiner (Table 1) and having a head noun, such as (*cette région*, *esta região, this region*). In both French and Portuguese, demonstrative determiners vary in gender and number. We are not considering demonstrative pronouns being full nominal constituents such as *este, esta, isto, aquele* (Portuguese) or *celui-ci, ceux de gauche* (French).

	Masculine		**Feminine**	
	Portuguese	**French**	**Portuguese**	**French**
Singular	este esse aquele	ce(t)	esta essa aquela	cette
Plural	estes esses aqueles	ces	estas essas aquelas	ces

Table 1: Demonstrative determiners

3 Criteria for the corpus analysis

3.1 Types of coreferential and anaphoric uses

One goal of our classification experiments was to investigate coreferential and anaphoric demonstratives. Relations between a demonstrative description *d* and its textual antecedent *a* (if any) were, therefore, classified depending on different categories of use.

Direct coreference: *d* corefers with a previous nominal expression *a*; *d* and *a* have the same nominal head:

(1) a. *as autoridades gregas (the greek authorities)*
d. *essas autoridades (these authorities)*

Indirect coreference: *d* corefers with a previous nominal expression *a*; *d* and *a* have different nominal heads:

(2) a. *a Albânia (Albania)*
d. *este país (this country)*

Other anaphora: the antecedent is not a nominal expression or the relation between demonstrative and its antecedent is not a coreference relation:

(3) a. *adoptar medidas de âmbito nacional (to adopt measures)*
d. *essa adopção (this adoption)*

These classes, based on previous work on computational processing of definite descriptions (Vieira & Poesio, 2000), enable us to evaluate the proportion of coreferential relations and of noun phrase antecedents for demonstrative noun phrases. The reason for isolating nominal antecedents from other expressions such as verb phrases, sentences or paragraphs is to evaluate how well a system for anaphora resolution of demonstratives can perform on the basis of nominal expression relations only. An assumption that seems to be reasonable within the context of the current state of the art of automatic anaphora resolution (Mitkov, 2002). The distinction between *same nominal head* and *different nominal head* allows us to observe the frequency of semantic bridging between a demonstrative and its antecedent, and gives therefore an idea about the need of additional lexical knowledge sources.

The *other anaphora* class represents the uses of demonstratives that require special techniques to identify antecedents that are not noun phrases (sentences, paragraphs or sets of those) and antecedents that do not refer to the same entity as the anaphoric demonstrative.

3.2 Syntactic structure of demonstrative noun phrases

French and Portuguese demonstrative noun phrases have been classified according to the presence or not of adjectival, prepositional and relative-clause modifiers. Each demonstrative NP belongs to one of the following classes, growing in terms of complexity:

Noun phrases containing only a head noun without modifiers (DET N), also including a few cases of Portuguese or French elliptical noun phrases such as *ce dernier – esse último* (*this latter one*):

(4) *cette région – esta região (this region)*

Noun phrases with adjectival modifiers (DET (ADJ N | N ADJ)):

(5) *ces pratiques abusives – estas práticas abusivas (these abusive practices)*

Noun phrases with prepositional phrases introduced by *de (of)* and perhaps adjectival modifiers (DET (N | ADJ N | N ADJ) OF (N | N ADJ | ADJ N)):

(6) *ces usages vulnérables de la route (these vulnerable uses of the road)*
(7) *esta ajuda de emergência (this help of emergency/emergency help)*

Nouns phrases with relative clauses and perhaps adjectival modifiers (DET (N | ADJ N | N ADJ) REL_PRO):

(8) *ces oiseaux que la loi protège (these birds that the law protects)*
(9) *este grave problema social que sofrem os cidadãos (this serious social problem that suffer the citizens)*

The reason for analysing this criteria was to explore a possible relation between complexity of syntactic structures and discourse roles of demonstrative NPs, traditionally considered as being predominantly coreferential or anaphoric (Corblin, 1987). Our underlying hypothesis is that demonstratives, whose interpretation is mainly context dependent, are preferably realized through simple noun phrase structures. In other terms, following (Löbner, 1985), the arguments for their semantic function are provided mainly by textual antecedents and not through noun phrase complements.

3.3 Size of antecedents

Also important for resolving anaphora is knowledge about certain characteristics of the antecedents. In preliminary analyses of the corpus, we noticed that demonstrative expressions tend to refer to ideas expressed throughout the texts (cases such as *this problem, this situation, these facts*). These abstract concepts have as antecedents not just clearly defined entities such as those referred to by noun phrases, but whole sentences or paragraphs as well as disjoint parts of texts.

To check the frequency of these cases in our corpus, we divided the antecedents into four categories (in the examples, “a” is for antecedents, “d” is for demonstratives):

Antecedents that were NPs (for which a single head noun can be clearly identified):

(10)a. *a substituição da fuligem por um produto menos nocivo (the substitution of the soot by another less harmful product)*
d. *este problema (this problem)*

Antecedents identified as being part of a sentence (longer than an NP but not a complete sentence):

(11)a. *estas taxas são aumentadas periodicamente (these taxes are increased periodically)*
d. *este procedimento do Governo italiano (this procedure of the Italian government)*

Antecedents that were full sentences:

(12)a. *A Comissão das Comunidades Europeias declarou pretender investir no transporte ferroviário de mercadorias, principalmente para distâncias de pelo menos 500 quilómetros e, se possível, superiores a 1 000 quilómetros. (The European Community Comission declared its intention of investing on rail transport for goods, mainly for distance greater than 500 km and, if possible, greater than 1000 km.)*
d. *esta posição (this position)*

Antecedents that were larger than one sentence (or not clearly identifiable by only one linguistic expression).

As systems for anaphor resolution usually consider only relations holding between noun phrases, our analysis will shed some light on how this assumption may influence the performance of such systems.

3.4 Semantic Analysis

Finally, certain basic semantic features (*concreteness* vs. *abstractness* and well-defined lexical relations) were analysed for the head nouns of demonstrative NPs and their antecedents.

First, the head nouns of both demonstratives and their antecedents were classified manually as abstract or concrete nouns according to distinctions presented in (Cegalla, 1996; Cunha & Cintra, 1985):

Concrete nouns refer to real existing beings (names of people, places, institutions, species), or else, things that imagination considers like that (fairy).

Abstract nouns refer to notions, actions, states and qualities. They are nouns referring to things that do not exist in the world by themselves; they depend on other beings to exist: beauty, love, trip, life.

This enabled us to compare the matching between concrete and abstract features of demonstrative and their antecedents. We also verified the syntactic structure of the antecedents for concrete demonstratives to test our hypothesis that concrete demonstratives have a tendency to have noun phrases as antecedents instead of more complex structures such as sentences or paragraphs.

Second, we analysed the semantic relation holding for those cases classified as indirect coreference, that is

Hypernymy:

(13)a. *Angola (Angola)*
d. *esse país (this country)*

Synonymy:

(14)a. *o período de 1991/1995 (the period of 1991/1995)*
d. *essa altura (this time)*

Discourse deictic (anaphora that rely on particular positions within the text, as in *este último* (*this last one*), analysed in (Corblin, 1999)):

(15)a. *o Conselho de Estado grego (the Greek State Council)*
d. *este último (this latter)*

Other semantic relations (less well defined relations):

(16)a. *a proteção das aves (the birds protection)*
d. *neste domínio (in this domain)*

As these semantic relations were observed within the context, pairs such as *obras cinematográficas – aquele tipo de criação artística / cinematographic works – that kind of artistic creation* were considered as synonymy. Also, the analysis was mainly made regarding the semantic relations holding between the head nouns of the two noun phrases (exceptions are special cases such as the previous examples *that kind of*). Therefore, while the relation holding between *1989* and *that time* was considered as hypernymy, the one holding between *the period of 1991/1995* and *that time* was considered as synonymy.

4 Corpus annotation

4.1 Corpus

The corpus of our study consists of French and Portuguese texts from the MLCC corpus. This multilingual parallel corpus contains written questions asked by members of the European Parliament and corresponding answers from the European Commission, published in the Official Journal of the European Commission, C Series, Written Questions 1993.

In order to have about 250 demonstratives for each language, we had to select a corpus of approximately 50,000 words, corresponding to 90 question-answer pairs. Table 2 presents a description of the resources we used. Although the texts are parallel texts, the French version has a greater number of demonstratives (291) than the Portuguese version (243).

Corpus	Language	Nb words	Demonstratives
MLCC	French	~ 50,000	291
	Portuguese		243

Table 2: Corpus for the study of demonstrative NPs

4.2 Annotation tool

MMAX[1] is a tool for corpus annotation (Müller & Strube, 2001), supporting annotation of electronic corpora, providing an interface for creating markables, annotating relations between markables, and browsing the annotation. It allows the specification of user-definable attributes for the markables and computes the Kappa reliability measure for different annotations. All data is represented in XML format. To annotate the corpus with the MMAX tool, we first transformed the corpus from its original SGML TEI standard to XML MMAX format, generating MMAX words and text files.

```
<words>
  <word id="word_49">milhares</word>
  <word id="word_50">de</word>
  <word id="word_51">refugiados</word>
</words>
```

Figure 1: Words basic file

```
<markables>
  <markable classification="indirect"
   id="markable_3" pointer="markable_8"
   np_form="demNP" span="word_135..word_136"/>
</markables>
```

Figure 2: Markables output file

The basic input format contains *word* elements as shown in Figure 1. The output of the annotation process is an XML file, containing a list of markables and their attributes as shown in Figure 2.

4.3 Annotation task

The annotation procedure was divided into three phases: selecting the markables, assigning the antecedents, and classifying the uses. We separated the task of selecting an antecedent from that of classifying types of use, according to previous experience (Vieira *et al.*, 2002) suggesting that low inter-annotator agreement was at least partly due to the complexity of the task. We considered that a native speaker identifies an antecedent in a more intuitive way if the task does not include classification at the same time. Phase 1 was done by one annotator for each language and the annotations of phases 2 and 3 were done by two subjects for each language.

Phase 1 – Selection of markables: In this phase, one annotator uses MMAX to mark the demonstrative descriptions in the corpus. Each demonstrative NP corresponds to a markable to be analysed in the following phases.

[1] http://www.eml.org/english/Research/NLP/Downloads

Phase 2 – Identification of textual antecedents: Two annotators (native speakers) mark the antecedents of the previously selected demonstratives.[2]

Phase 3 – Classification of the coreference and anaphoric relations: In the third phase of the annotation, the relationship between demonstratives and their textual antecedents were classified, according to the uses defined in Section 3.1. Additionally, we checked the values for the syntactic and semantic features also introduced in the previous section.

5 Results

Here we show the resulting analysis of the features described in Section 3: general distribution of coreferential and anaphoric use of demonstrative NPs (5.1), their syntactic structure (5.2), the type of antecedents for demonstrative anaphora (5.3) and some basic semantic characteristics of demonstrative NPs head nouns (5.4). In Section 5.5, we correlate some of these properties.

5.1 Types of coreferential or anaphoric uses

Since demonstratives are likely to identify their referent on the basis of salience, and given our material (written texts), we expected them to be necessarily related to previous discourse, and preferentially in a coreferential way. Our classification results do support these hypotheses for both French and Portuguese corpora.

Category %	French	Portuguese
Direct coreference	32	34
Indirect coreference	21	18
Other anaphora	47	48
Total	100	100

Table 3: Classification of French and Portuguese demonstratives

The results in Table 3 show that demonstratives are context dependent, with more than half of them being coreferential with previous NPs. The other half are either coreferential with antecedents which are not NP or not coreferential.

Demonstratives whose antecedents were not explicitly marked are also included in the *other anaphora* class. The fact that we observed a high number of abstract head nouns for demonstratives of this group (*manner, range, problem, reason, purpose, situation, case, decision, context,* …) led us to investigate further correlations between concreteness/abstractness of head nouns and type of anaphoric use (Section 5.5).

[2] Antecedents greater than one sentence as well as antecedents not clearly identifiable by a single text chunk were not marked due to practical reasons related to the tool (the selection of such long markables would prevent the visual distinction of markables and antecedents in the texts).

5.2 Syntactic structure

Table 4 presents the distribution of French and Portuguese demonstratives over the classes for syntactic structures, as defined in Section 3.2. In French as well as in Portuguese, demonstratives present few modified structures: only 20 % in both languages are subject to adjectival, prepositional or relative clause modification.

Syntactic structure %	Demonstrative NPs		Definite NPs	
	French	Portuguese	French	Portuguese
DET N	80.4	80.2	35.4	40.8
DET(ADJ N | N ADJ)	10.3	7.6	22.6	22.7
DET (N |ADJ N | N ADJ) OF N	7.2	7.3	30.0	28.7
DET (N | ADJ N | N ADJ) REL_PRO	1.1	0.8	2.3	2.3
Other	1.0	4.1	9.7	5.5
Total	100	100	100	100

Table 4: Syntactic structure of demonstratives, compared to definites

When compared to the structure of definite descriptions investigated in previous work (Vieira *et al.*, 2002), we noticed the difference between definites and demonstratives regarding the proportion of noun phrases belonging to class 1 (head noun without modifiers). This proportion is about 37% for definites in the two languages, whereas for demonstratives this structure is verified for about 80% of the cases. One possibility is that definite descriptions are more often interpreted on the basis of semantic information, but not necessarily anaphorically to entities introduced within the previous discourse, as first observed in (Poesio & Vieira, 1998). If one considers that the quantity of semantic information increases with the adjunction of modifiers, then the fact that they belong mainly to complex classes would confirm this hypothesis. Moreover, one can suppose that the more semantic information is given within the definite noun phrase itself, the less important is the interpretational dependency on information provided by previous discourse.

Regarding demonstratives, in French as well as in Portuguese, we have few modified demonstrative NPs (only about 20%). As opposite to the explanation for definites, this small proportion can be seen as a confirmation of the interpretational property of demonstratives of referring to something already salient through previous discourse. Indeed, the lack of modifiers and, therefore, less semantic information about the referent increases the need of supplying this information by the discourse context and might be seen as a confirmation for considering demonstratives as mainly anaphoric expressions rather than discourse new, according to the Giveness Hierarchy model (Prince, 1981, 1992; Gundel *et al.*, 1993).

5.3 Size of antecedents

Type of the antecedent %	French Ann.1	French Ann.2	Portuguese Ann.1	Portuguese Ann.2
NP	81	68	62	65
< Sentence	9	7	22	9
Sentence	6	10	4	1
Not marked	4	15	12	25
Total	100	100	100	100

Table 5: Type of antecedent for demonstrative anaphora

The results in Table 5 show that the antecedents for demonstrative NPs were noun phrase structures at least in 62% for all annotators. In the remaining cases, the antecedents were identified as one single sentence, part of a sentence or paragraphs (which accounts for most cases of antecedents not marked). This gives us an idea of the limitation of systems that work on anaphor resolution based on NP structures only. Such a system is likely to fail on about 30% of the cases on the basis of this assumption.

From the results shown in Section 5.1 (Table 3), we could see that nearly 50% of the demonstratives were coreferential with previous NPs. However, the number of NP antecedents identified by the annotators (Table 5) sum up to 81% of the cases. Therefore, at least 30% of the demonstratives stand in other kind of anaphoric relation with previous NPs. An example is:

(17)a. *l‘ installation, dans la forêt pétrifiée, de neuf aérogénérateurs (the installation, in the petrified forest, of nine wind generators)*

d. *cette atteinte portée à un monument d‘ histoire naturelle d‘ importance considérable (this considerable attack to a monument of natural history)*

Examples of demonstrative NP head nouns, for which antecedents were not marked, are *point*, *interpretation*, *efforts* or *sense*. Again, we have mainly abstract nouns, for which a specific textual antecedent is hard to identify in the text. Therefore, the correlation between the semantics of the demonstrative head noun and the size or type of the antecedent was investigated, as presented in Section 5.5.

5.4 Semantic analysis

Concrete vs. abstract demonstratives and antecedents

Semantic classification %	French	Portuguese
Concrete	21	22
Abstract	79	78
Total	100	100

Table 6: Demonstrative NP head nouns

Table 6 shows the results regarding the semantic analyses of demonstrative head nouns, according to the abstract and concrete distinction (Section 3.4). As for their distribution, the results confirm our hypothesis: there is a clear predominance of abstract head nouns in demonstrative noun phrases (near 80%). Another positive point is the equal distribution of concrete and abstract head nouns in French and Portuguese since the classification was done manually by different annotators. Table 7 shows the semantic classification of the antecedent head nouns, for each annotator and for both languages. Whereas demonstrative noun phrases were predominantly abstract for both languages, the classification of the antecedents was found to be less consistent. In Portuguese, the antecedents were mainly concrete (57%) and for French, mainly abstract (67%).

Semantic Classification %	French			Portuguese		
	Ann. 1	Ann. 2	Averag e	Ann. 1	Ann. 2	Averag e
Concrete	32	33	33	66	49	57
Abstract	68	66	67	34	51	43
Total	100	100	100	100	100	100

Table 7: Semantic classification of antecedent head nouns

Given the classification results for the demonstrative NPs (Table 6), this means also that demonstrative anaphora are sometimes used to re-classify the entity referred to by the antecedent by a more abstract noun, this observation being consistent with previous linguistic analyses of discourse roles of demonstrative NPs (Corblin, 1987). An example for such a case is:

(18)a. *une essence super à teneur en octane plus élevée (a super benzine with higher octane)*
d. *cette dernière qualité (this latter quality)*

Furthermore, we also investigated the correlation between concrete and abstract demonstratives and their antecedents as well as the relation between concrete and abstract demonstratives with the size of the antecedents. The results are reported in Section 5.5.

Semantic relations

Another semantic feature we analysed was the semantic relation holding between indirect coreferential demonstratives and their antecedents. Table 8 shows the distribution over the semantic relations presented in Section 3.4. Concerning well-defined semantic relations, there is a clear predominance of hypernymy. However, other frequent type of relation is the *other semantic relations* class, referring to cases often based on general semantic inference, which do not correspond to a precise lexical semantic relation.

Semantic relation %	Portuguese		French	
	Ann. 1	Ann. 2	Ann. 1	Ann. 2
Hypernymy	41	65	33	40
Synonymy	5	4	7	10
Discourse deictic	3	4	15	19
Other semantic relations	51	27	45	31
Total	100	100	100	100

Table 8: Semantic relations for demonstratives (indirect coreference)

5.5 Cross feature analyses

Concreteness/abstractness and anaphoric relations

Semantic classification %	French		Portuguese	
	Concrete	Abstract	Concrete	Abstract
Direct coreference	50	28	64	24
Indirect coreference	34	11	31	23
Other anaphora	16	61	5	53
Total	100	100	100	100

Table 9: Semantic of head nouns vs. anaphoric relation

The observation of many abstract head nouns for non coreferential demonstratives (Section 5.1) raises the question of whether the semantic features of demonstrative head nouns (i.e. abstract or concrete) allow predictions about the type of the anaphoric relation between the demonstrative NP and its antecedent.

Table 9 shows this relation for French and Portuguese demonstratives. They confirm our intuition by showing that more than 80% of demonstratives with a concrete head noun enter in a coreference relation with their antecedent, whereas this is the case for only 40% of demonstratives with an abstract head noun. This observation could be used as a baseline for evaluating demonstrative anaphora resolution separately for concrete and abstract head nouns.

Concreteness/abstractness of demonstratives and antecedents

Dem NP	Antecedents %			Total
	Concrete	Abstract	not NP	
Concrete	94	2	4	100
Abstract	30	25	45	100

Table 10: Semantics of demonstratives and antecedents (Portuguese)

Dem NP	Antecedents %			Total
	Concrete	Abstract	not NP	
Concrete	92	8	0	100
Abstract	7	67	26	100

Table 11: Semantics of demonstratives and antecedents (French)

In Section 5.4 we presented the classification into *concrete* or *abstract* for the head nouns of demonstrative NPs and antecedents. Here, we analyse the interconnection between these features. Tables 10 and 11 show the percentage of concrete and abstracts antecedents, depending on concreteness or abstractness of the demonstratives, according to one annotator for each language. Demonstratives were considered to be either concrete or abstract, but antecedents are sometimes not expressed as NPs.

For concrete head noun demonstratives, the antecedent head noun is concrete as well most of the times (over 90 % for both languages). This observation could be important for anaphor resolution heuristics, since it allows excluding less plausible antecedent candidates for concrete demonstratives, provided a suitable lexicon containing the needed semantic information. An example follows:

(19)a. *associations ecologists (ecologist associations)*
d. *ces associations (these associations)*

Cases where concrete demonstratives are anaphoric to abstract head noun antecedents are rare in both languages. We found here cases of metonymy (20) and process-result polysemy (21). In both cases, the relation could not be said coreferential in a strict sense.

(20)a. *le vol Air Lingus EA 643 (the flight Air Lingus EA 643)*
d. *cet avion (this plane)*
(21)a. *une demande d'information (a request for information)*
d. *cette letter (this letter)*

For demonstratives with abstract head nouns, things are less straightforward. It seems however that the probability that they refer to entities introduced previously by concrete head nouns is low (between 0.07 and 0.3, depending on the language), although it is still higher than the inverse case (abstract antecedent for concrete demonstrative). This could be explained by the fact that in addition to result-process polysemy (*informatics, activity*), this configuration includes also generic anaphora (classes referred to by expressions like *this genre*, *this species*), as shown in the following examples:

(22)a. *des entrepises informatiques (informatics companies)*
d. *cette activité industrielle (this industrial activity)*
(23)a. *les rares chèvres sauvages (the rare wild goats)*
d. *cette espèce (this species)*

Finally, we present an example of a demonstrative NP with abstract head noun whose antecedent has also an abstract head noun. However, this is a combination that cannot be predicted, since the antecedents of abstract demonstratives were non-NPs in up to 50% of the cases.

(24)a. *l'exode de milliers d'Albanais (the outflow of millions of Albanians)*
d. *cet afflux massif de réfugiés auxquels elles doivent fournir une assistance humanitaire (this massive influx of refugees to whom they should provide humanitarian assistance)*

Semantics of demonstratives and syntactic structure of antecedents

Finally, we correlated semantics (*concrete* vs. *abstract*) of demonstratives with different syntactic structures of the antecedents (NP and non-NP), investigating whether the semantic feature of a head noun makes it possible to predict the preferred syntactic structure of the antecedent. The results for one annotator per language are presented in Table 12.

Demonstrative head noun	Antecedents %			
	French		Portuguese	
	NP	non-NP	NP	non-NP
Concrete	100	0	94	6
Abstract	74	26	53	47

Table 12: Semantics of demonstratives and type of antecedents

As a result, concrete demonstratives were related to NP antecedents for the majority of the cases for both languages (94 to 100%). Again, for abstract head nouns, it is difficult to draw conclusions, since they seem to be generally distributed over NP and non-NP antecedents.

6 Agreement issues

We verified the inter-annotator agreement on classifications as well as on the identification of antecedents for each language. In order to evaluate the inter-annotator agreement on the classification task, we calculated Kappa (Carletta, 1996) for each experiment. This measure establishes K = 0.8 as good agreement. We calculated Kappa for the three classes (direct coreference, indirect coreference, other). We found K = 0.79 for French and K = 0.65 for Portuguese demonstratives. These results show better agreement than those of previous experiments related to four different classes of definite descriptions (Vieira *et al.*, 2002). The improvement might be related to the reduced number of classes as well as to the fact that we isolated in this experiment the identification of the antecedent from the classification task. Informal feedback from the annotators also suggests that the annotation task was easier for demonstratives than for definites. We have also compared the choice of antecedents for the two annotators of each language.

The results are presented in Tables 13 and 14. For annotators 1 and 2 in each language, these tables show cases where the antecedent was the same or not (A1=A2, A1≠A2) in correlation with the type of antecedent chosen (*direct*, *indirect*, *other* as well as those cases in which the antecedent was not marked ∅, because it was greater than a sentence). There was total agreement on the antecedents for 51% of the cases in Portuguese and 69.8% for French. Most cases of disagreement for Portuguese were related to cases where the antecedent was not marked. In some cases (around 4% in Portuguese and 9% in French), the antecedents chosen by the annotators are not the same but they are coreferential expressions themselves (Coreference(A1,A2)) which can be considered as partial agreement.

Agreement on antecedents		#	%
A1 = A2	**Direct**	61	25,1
	Indirect	31	12,7
	Other	20	8,2
	A1 = A2 = ∅	12	4,9
	Total agreement	124	51
A1 ≠ A2	**(A1 or A2) = ∅**	62	25,5
	Coreference (A1, A2)	10	4,1
	¬ Coreference (A1, A2)	47	19,3
	Total disagreement	119	49

Table 13: Agreement on antecedents in Portuguese corpus

Agreement on antecedents		#	%
A1 = A2	**A1 = A2 = ∅**	11	3,8
	Direct	76	26,1
	Indirect	43	14,8
	Other	73	25,1
	Total agreement	203	69,8
A1 ≠ A2	**(A1 or A2) = ∅**	29	10,0
	Coreference (A1, A2)	27	9,3
	¬ Coreference (A1, A2)	32	11,0
	Total disagreement	88	30,2

Table 14: Agreement on antecedents in French corpus

7 Conclusions and future work

This study investigated anaphoric and coreferential properties of demonstrative noun phrases in French and Portuguese. Having in mind the overall objective of designing a tool for definite and demonstrative noun phrase reference resolution, the main conclusions of this work are the following:

As suggested by linguistic description (Corblin, 1987) and as opposed to definite descriptions (Poesio & Vieira, 1998; Vieira *et al.*, 2002), the interpretation of demonstrative noun phrases is mainly context dependant, in the sense that human annotators are able to find, for more than 80% of them, textual chunks as antecedents. Moreover, this hypothesis seems to be reinforced by the finding that over 80% of demonstrative NPs are noun phrases without any additional modifier, suggesting that this type of anaphora is less informative by itself and relies heavily on textual context.

However, the demonstrative NPs were identified as coreferential with previous NPs in about 50% of the cases only. This observation gives raise to two comments.

First, for all the cases where the antecedent is a non nominal text chunk, i.e. for more than 40% of demonstrative NPs in our corpus, it is difficult to select a precise portion of the text as an antecedent: the limits between verbal phrases, sentences and even paragraphs for presenting an idea recovered with abstract nouns such as *this manner, this situation* or *this point of view* are not easy to analyse.

Secondly, when the relation of a demonstrative and its antecedent is not a coreferential one, the amount of world knowledge and reasoning required for the resolution is very large. As for other types of nominal anaphora (Poesio *et al.*, 2000), less than half of the cases enter in a well-defined lexical relation and could therefore be resolved on the base of lexical resources such as WordNet. An additional problem is here the lack of a well-developed WordNets for other languages than English.

However, as challenging as these problems may be seen, we raised several cross-language features specifically related to the discourse role of demonstrative expressions: they are not only mainly textual dependent for their interpretation (either coreferential or anaphoric), but in more than half of the cases, the antecedent is also an NP. Furthermore, classification experiments on basic semantic features of the head nouns involved in demonstrative anaphora and the related antecedents (abstract vs. concrete entity) have shown that concrete demonstratives have high tendency to take concrete NPs as antecedents (over 90%). Abstract demonstratives rely in a less strong way on antecedent NPs (between 50% and 70%, depending on annotators and languages).

As an overall conclusion, one might keep in mind two important points: on the one hand, most of the properties we investigated seems to be valid across languages, since the results are similar in French and in Portuguese; on the other hand, the specific distribution of the syntactic and semantic features for demonstrative NPs seems to justify a specific treatment of this kind of anaphora as opposed to other anaphoric expressions, such as pronouns or definite descriptions. Further work is needed for the analysis of coreferent demonstrative with non-NP antecedents as well as for non-coreferent anaphoric demonstratives.

Acknowledgments

This work was developed with financial support of CNPq, INRIA and FAPERGS. We also would like to thank Gabriel Ávila Othero and our annotators Cassiano Haag, Jean-Luc Benoit, Emmanuel Schang, and Margarete Silva.

References

Carletta, J. 1996. "Assessing Agreement on Classification Tasks: The Kappa Statistic." *Computational Linguistics* 22.249-254.

Cegalla, D. 1996. *Novíssima Gramática da Língua Portuguesa*. São Paulo: Nacional.

Corblin, F. 1987. *Indéfini, Défini et Démonstratif*. Droz, Genève.

---------- 1999. "Les références mentionnelles : le premier, le dernier, celui-ci." In Mettouch A., Quinyin H.: *La référence (2). Statut et processus*. Travaux linguistiques du CERLICO, P.U. Rennes.

Cunha, C. and L. Cintra 1985. *Nova Gramática do Português Contemporâneo*. Rio de Janeiro: Nova Fronteira.

Gundel, J., N. Hedberg and R. Zacharski 1993. "Cognitive Status and the Form of Referring Expressions in Discourse." *Language* 69.274-307.

Löbner, S. 1985. "Definites." *Journal of Semantics* 4.279-326.

Mitkov, R. 2002. *Anaphora Resolution (Studies in Language and Linguistics)*. London: Longman.

Müller, C. and M. Strube 2001. "MMAX: A Tool for the Annotation of Multi-modal Corpora." *Proc. of the 2nd IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*. Seattle, 45-50.

Poesio, M. and R. Vieira 1998. "A Corpus-based Investigation of Definite Description Use." *Computational Linguistics* 24.183-216.

----------, T. Ishikawa, S. Walde and R. Vieira 2002. "Acquiring Lexical Knowledge for Anaphora Resolution." *Proc. Language Resources and Evaluation Conference* (LREC 2002), Las Palmas, 1220-1224.

Prince, E. 1981. "Toward a Taxonomy of Given-new Information." P. Cole (ed.), *Radical Pragmatics*. New York: Academic Press, 223-256.

---------- 1992. "The {ZPG} Letter: Subjects, Definiteness, and Information Status." S. Thompson and W. Mann (eds.) *Discourse Description: Diverse Analyses of a Fund-raising Text*. Amesterdam: J. Benjamins, 295-325.

Salmon-Alt, S. and R. Vieira 2002. "Nominal Expressions in Multilingual Corpora: Definites and Demonstratives." *Language Resources and Evaluation Conference* (LREC 2002), Las Palmas, 1627-1634.

Sant'Anna, V. and V. Lima 2002. "Resolution of demonstrative anaphoric references in portuguese written texts." *Proc. of Portugal for Natural language Processing PorTAL 2002*, Faro, 87-90.

Strand, K. 1996. "A Taxonomy of Linking Relations." *IndiAna Workshop*, Lancaster.

Strube, M., S. Rapp and C. Müller 2002 "The influence of minimum edit distance on reference resolution." *The 2002 Conference on Empirical Methods in Natural Language Processing*. Philadelphia, 312-319.

van Deemter, K. and R. Kibble 2000. "On Coreferring: Coreference in MUC and related annotation schemes." *Computational Linguistics* 26(4).629-637.

Vieira, R. and S. Teufel 1997. "Towards Resolution of Bridging Descriptions." *35th Int. Joint Conference on Computational Linguistics*, Madrid, 522-524.

Anaphoric Demonstratives: Dealing with the Hard Cases

Marco Rocha
Universidade Federal de Santa Catarina

This paper presents a corpus-based study of anaphoric demonstratives in dialogues in English and Portuguese, with a view to an implementation as part of an anaphor resolution module in a dialogue interface. The resolution of anaphoric demonstratives is often seen as particularly hard because of implicit and discourse-chunk antecedents. In an attempt to overcome such difficulties, antecedents of anaphoric demonstratives found in dialogues in the London Lund Corpus, for English, and in the CDC-RJ, for Portuguese, were analysed according to realisation and phrase structure. Anaphoric demonstratives in collocations were identified in a list so as to establish patterns which were then associated individually with a resolution strategy. An approach relying on information structure, as defined by topic tracking, and other data derived from corpus analysis is proposed to identify adequate processing strategies for the resolution of cases which could not have their resolution paths associated to collocations.

1 Introduction

The study on anaphoric demonstratives presented here is part of a larger investigation which aims at building an integrated model of anaphoric phenomena in dialogues. The model was conceived in the belief that one important aspect of a successful approach to anaphora resolution is the ability to deal ultimately with all forms of anaphoric phenomena in an integrated classification model. Having in mind this integrated model, the annotation scheme outlined in (Rocha, 1997) and fully described in (Rocha, 1998) is the analytical tool used to classify cases of anaphora in two dialogue corpora, namely, the spoken language component of the London-Lund Corpus, for the English data, and the Corpus de Diálogos Clínicos do Rio de Janeiro (CDC-RJ), for Portuguese data.

Therefore, the annotation classifies each case of anaphora according to four properties seen as conducive to the successful completion of the integrated classification model of anaphoric phenomena. The first property, in an analytical ordering, is the **type of anaphor**, which includes anaphoric demonstratives, the central issue in the present study, as one of its possible values. The full range of possible values for this property will not be discussed here (Rocha, 1998). Regarding implementation, it seems reasonable to expect that the mapping of values for the type of anaphor should be relatively straightforward on the basis of the output of a POS tagger and a parser. That is the reason why the property is seen as the first stage in the process of analysing anaphoric references for purposes of both annotation and, in the future, automatic annotation and resolution.

An anaphoric demonstrative is defined in the study as tokens of *that*, *this*, *these* and *those*, in English; and *este*, *esse*, *aquele*, *isto*, *isso*, *aquilo* and *o*, plus their gender and number inflections, in Portuguese, whenever used as pronouns. As known, English demonstratives may also appear as determiners and, in the case of *that*, as a relative pronoun and a conjunction. It is also true that Portuguese demonstratives *este*, *esse*, *aquele* and their inflections may occur as determiners; and so may *o* and its inflections. However, the latter are classified as articles whenever appearing in the function of a determiner, whereas the former are classified as determinative demonstratives. The identification of the anaphor in question is thus essentially a matter of POS tagging, so that usage as a pronoun is distinguished from other possible usages.

The identification of the antecedent for anaphoric demonstratives is surely a far more complex matter. The three remaining properties in the classification scheme are expected to provide analytical information leading to a training process which would ultimately render anaphor resolution approachable by means of machine learning techniques. These properties include a **processing strategy**, which classifies the resolution path towards identification of the antecedent; a **type of antecedent**, which is a classification of the antecedent, as perceived by the analyst, based on the explicit/implicit dichotomy; and a **topical role of the antecedent**, which attempts to classify the antecedent regarding phrase structure and also according to roles in terms of topicality, that is, discourse saliency. Phrase structure is included in the topical role classification because the topical role of a discourse chunk, as compared to a noun phrase, must be dealt with differently in terms of topicality.[1]

The notion of topic is notoriously difficult to use, although intuitively clear to human speakers. One particularly intractable aspect of the idea of topic in computational terms is that, linguistically, the concept is highly volatile. A number of approaches to topic tracking have been proposed in works within the area of text linguistics, but there is no agreement among linguists on a definition of what a topic is and how to specify a topic, given a piece of text. At the intuitive level, native speakers of a language, if asked, will differ in their opinions as to the ‘topic’ of a passage, talk or dialogue. It is not surprising, thus, that designers of computer systems have problems with the notion.

Nonetheless, under various forms and names, the idea of using information structure, in the sense of what is being actually said in a text or dialogue, to improve the performance of computational systems processing human

[1] This will be made clearer later in Section 3.5.

languages has been around recurrently since the seminal paper by Grosz (1986). However, the implementation of automated discourse segmentation on the basis of topic tracking has proved to be a particularly hard task. To a certain extent, this may be precisely the cause of the difficulty in processing anaphoric demonstratives. This type of anaphor very often requires the selection of discourse chunks as antecedents. In order to determine the appropriate antecedent, the use of information regarding the current topic, under various possible forms, seems to be the only way to go.

One understandable consequence of processing failure for a given task is a gradual discontinuation of attempts to implement it. Results are an integral part of research evaluation, and anaphora resolution is no exception. It is therefore better to have neat results, described in terms of precision and recall, than to insist in approaches that yield very limited success rates. This is certainly true, but it is also important to bear in mind that there are tasks which seem to require a certain type of information which is not as readily available. It is thus acceptable, to say the least, that some research effort is made towards thinking about solutions which cannot be implemented or do not produce immediate results.

One way forward, adopted in the study described here, is to collect information through corpus observation, and then try to codify this information by means of an annotation. Corpus analysis gathers information about a given phenomenon. In the long run, if the quality of the codification is systematically improved by testing, one may reasonably expect the results that could not be easily obtained in the beginning to be finally achieved. Two properties – namely, processing strategy and topical role of the antecedent – included in the classification model used in the study are difficult to annotate because possible values cannot be expressed in terms of surface data straightforwardly. The analyst must decide which category to use in order to classify a case with a substantial degree of uncertainty, and then, by repeatedly facing similar choices, come to an eventual point of equilibrium in which decisions become more and more stable. This process is far more difficult if several annotators are at work, which was not the case in the present study.[2]

It is hoped that the challenging complexity inherent to this classification model will eventually pay off by bringing into play information that will contribute to anaphor resolution in human language technology systems. The present study is meant as an attempt to show how this detailed classification

---

[2] It would be of course desirable to have as many annotators as possible working in the analysis of the corpora, regardless of interannotator discrepancy. The single-annotator option was not a matter of choice, but of necessity.

model may actually result in gain by offering inroads towards the solution of particularly hard problems of anaphoric reference. The remainder of the paper is organised as follows: the next Section briefly reviews a small number of related works; the third Section details the classification model as it applies to anaphoric demonstratives; the fourth Section points to possible solutions for implementation difficulties using the model; the fifth Section briefly describes possible developments regarding machine translation; the last Section concludes with further discussion of issues concerning the trade-off between analytical complexity and ease of implementation.

2 Related Work

As pointed out in (Byron & Allen, 1998), anaphoric demonstratives have received little attention in the literature about anaphora resolution. Although it might be said that some improvement has been observed since then, it is still a particularly difficult aspect of anaphoric reference which has often been neglected. In what regards work on anaphoric demonstratives in computational linguistics, the authors point to (Webber, 1988, 1991) and (Passonneau, 1993) as “notable exceptions” to this shortage of investigations.

Webber's discussion assumes that the problem of defining discourse segments will be solved by some means. It is of course particularly difficult to determine when a given discourse segment starts and a preceding one ends. The distinction between discourse entities that are discourse segments and those introduced by noun phrases remains important as a starting point for investigations on anaphoric reference by means of demonstratives. Also important in Webber's work is the distinction between what demonstratives point to and what they refer to “by virtue of pointing”. In her discussion on terminology, the author uses the term “discourse deixis” to name the sort of phenomenon she is concerned with. She also sees discourse segments as the entities demonstratives point to in these cases of discourse deixis.

Byron and Allen (1998) use an annotation scheme to classify occurrences of anaphoric pronouns in the TRAINS spoken dialogue corpus. The study thus bears some similarities to the one presented here. Also the ideas behind the annotation scheme reflect some of the facts which the corpus annotation used in the present study tried to capture. Most importantly, the scheme includes a referent category which, together with a classification of the antecedent, when identified by the analyst, seeks to include some of the complex discourse aspects also aimed at by the present study.

Interestingly as well, some of the facts included in the classification of a processing strategy for each case of anaphora have also been pointed out in Byron and Allen's treatment of anaphoric demonstratives, such as the influence of structural parallelism and the presence of statistical patterns which can be used for future improvement of approaches to this particularly difficult problem. The approach differs from the one presented here, however, in the use of the centering framework as part of the information included in the analysis of the anaphors, which is not attempted here.

Although not focused on demonstratives, the investigation described in (Paul *et al.*, 1999) shows major similarities with the approach used in the research described here. It is also a corpus-based approach which attempts to combine statistical information and a machine learning method, which is the same proposed here, that is, a decision tree trained on an annotated corpus. The decision tree learns coreference relations from a corpus on the basis of training attributes which include lexical word attributes applied to the anaphor, antecedent candidate and clause predicate. The output of the decision tree classifies each anaphor-candidate pair as either belonging to the reference or non-reference class. Thus, the decision tree acts as a filter to eliminate irrelevant candidates. It is therefore quite similar to the recognition procedure for the identification of the correct processing strategy in the present study. The subsequent selection of the best candidate on the basis of preference strategies also bears strong similarity to the resolution procedures associated to possible processing strategies for each type of anaphor.

Soon *et al.* (2001) describe a system to establish coreference relations between noun phrases which is also based on decision trees. The induction of the decision trees uses only twelve surface-level features and achieves good performance on this knowledge-poor basis when applied to two MUC data sets. Ng and Cardie (2002) extend the work of Soon *et al.* (2001), describing improved results by adding modifications to the machine learning framework and increasing a great deal the number of linguistic features. Results with the full feature set are described as significantly inferior, particularly for common noun resolution. Manual feature selection achieves better results. Overall, however, the task of establishing coreference relations automatically is still very difficult. Anaphoric demonstratives may be said to pose an even harder challenge due to complexities previously mentioned in the nature of possible antecedents.

3 The classification model

Annotated cases which provide the empirical foundation for this investigation amount to 329 anaphoric demonstratives in a corpus of six dialogues in English, which add up to 22,915 words, thus yielding a ratio of 69.65. Comparatively, 167 anaphoric demonstratives were found and annotated in a corpus of six dialogues in Portuguese, which add up to 20,059 words, a ratio of

120.11. This large difference in the proportion of anaphoric demonstratives in each language is due to omitted subjects in Portuguese, which change the classification for the type of anaphor. A number of English anaphoric demonstratives have their equivalents in Portuguese realised by verbs without subjects, thus reducing the frequency rate of demonstratives. Contrastive cross-linguistic analysis is thus likely to uncover fruitful information for machine translation.

3.1 The processing strategy

As said in the introduction, the processing strategy is an attempt to classify the resolution path towards the identification of the antecedent or, in broader terms, a classification of the knowledge needed to resolve a given case of anaphora, and relate it to the type of anaphor – as mapped from the output of a tagger and parser – and its immediate context of occurrence, so that training of an anaphora interpreter for a dialogue interface could be carried out using machine learning techniques. There are fifteen possible values for the processing strategy in the general classification model for anaphoric phenomena used for the annotation of both the English and Portuguese dialogues.

The basic approach of the annotator is to check whether a plain first-candidate search backward would identify the correct antecedent, having, as the prototypical antecedent, a noun phrase explicitly introduced previously in the dialogue. In case it does, the value **first-candidate search** is assigned as the processing strategy used for the resolution of the anaphor under analysis. If the search leads to the identification of another anaphor as the antecedent, the value assigned is **first-candidate chain**, as the previously processed resolution of this anaphor should lead to the common antecedent.

Regarding anaphoric demonstratives in English, only 28.3% of the cases are resolved by means of these two similar strategies, as compared to personal pronouns, which were classified as resolved by means of these strategies in 70.75% of the cases. The situation is not different in the Portuguese corpus, where 24% of the anaphoric demonstratives were classified as resolved by means of the first-candidate strategies, as compared to a very high 81.95% of the personal pronouns. It is thus clear that an unsophisticated approach to the resolution of anaphoric demonstratives would not achieve good results.

However, if three other possible categories are added to the classification of English data for anaphoric demonstratives, namely **discourse knowledge**, **collocational knowledge** and **deixis**, 99.1%of the cases are classified. That leaves out 0.9 % of the cases, which, in absolute numbers, amount to three cases only, as shown in Table 1 below.

	English	Portuguese

Discourse knowledge	128 (38.9%)	45 (26.9%)
Collocation knowledge	61 (18.5%)	54 (32.3%)
First-candidate search	49 (14.9%)	18 (10.8%)
First-candidate chain	44 (13.4%)	22 (13.2%)
Deixis	44 (13.4%)	12 (7.2%)
Parallel	2 (0.6%)	1 (0.6%)
Set Creation	1 (0.3%)	2 (1.2%)
Others		13 (7.8%)
Total	329 (100%)	167 (100%)

Table 1: Distribution of anaphora cases per processing strategy in English and Portuguese

Moreover, close to 60% of all cases were classified as resolved by means of either discourse knowledge or collocational knowledge in both languages, although proportions are inverted, as the former is the most frequent in English, whereas the latter dominates the picture – to a lesser extent though – in Portuguese.

The distribution of cases per processing strategy is less concentrated in Portuguese. For the present paper, it has been decided that those categories which appeared only in the classification of Portuguese cases would not be discussed. They have been grouped in Table 1 under the name of **Others**. Categories used only for other types of anaphor that are not demonstratives will not be discussed as well. Nonetheless, the approach used for all cases is the same, that is, deciding on the appropriate processing strategy is seen as the essential step towards efficient anaphora resolution, and this is particularly true of hard cases which cannot be identified correctly on a first-candidate basis, such as anaphoric demonstratives have, in large proportion, shown to be. The discussion will focus then on the recognition of one of those three strategies as the adequate one to resolve a given case, as they, in combination with first-candidate procedures, would solve the great majority of cases of anaphoric demonstratives in both languages.

The recognition process is based on features of the immediate context, as observed in the training corpus. In this sense, the approach avoids very complex decisions, in an attempt to use surface information as much as possible. Some of these immediate contexts are suitable for treatment as collocations or patterns of co-occurrence. Therefore, cases classified as resolved by means of **collocation knowledge** are those in which an anaphoric demonstrative appears in a collocation that, in turn, is associated to a defined resolution path in a recurrent way. A list of fixed immediate contexts such as these described above was collected on the basis of corpus analysis and used in manual tests. A few examples of entries in this list, which is too large to discuss in full in this paper, are analysed in the following section.

3.2 The collocation list

Entries in the collocation list are made up of a header with the specific collocation or pattern, followed by information regarding the antecedent and the possible resolution path.

anything/something/nothing like that

- anaphoric
- antecedent = discourse chunk
- previous or same move
- connecting *or* frequent

The entry describes a pattern of occurrence for the demonstrative *that* reproduced on the top. Words separated by slashes are interchangeable. According to corpus data, these tokens of anaphoric demonstratives are never cataphoric and always have discourse-chunk antecedents. The chunk in question may be the previous move or a preceding part of the same move,[3] often connected by *or*. Thus, once this pattern is recognised, there is a resolution path associated to it included in the collocation list entry. Another example is given below.

that's (not) true

- anaphoric
- antecedent = discourse chunk
- previous move *the fact that*
- if coordinated clauses precede
- it may be the second one only

This entry defines a pattern that contains an optional *not*, codified by the round brackets, and a slightly more complicated resolution path. It is necessary to add *the fact that...* to discourse chunk antecedents, typically the previous move, in order to have them make sense in substitution tests. It is also pointed out that, in case the previous move is composed of coordinated clauses, the antecedent may be the last one only, but a measure of variation is to be found for this pattern. Collocation knowledge as a processing strategy is further illustrated by a third example.

that's it

nonreferential

This pattern was analysed as containing a nonreferential demonstrative, that is, there is no antecedent to speak of. This sort of information is also useful in order to avoid unnecessary searches for antecedents which are so vague as to

---

[3] In the sense used in (Sinclair, 1992).

play no role in the overall processing. They may be interpreted as purely interactional, amounting to a statement of support, such as *you're right*. A final example shows a particularly complex resolution path with segmentation information brought into play.

that X-be the reason for X

- anaphoric
- discourse implicit antecedent
- previous subsegment(s)
- agreement may be imperfect
- more than one *reason* as antecedent

The combinations of *that* with any form of the verb *to be*, codified as *X-be*, the phrase *the reason for*, and some undefined preposition object may have antecedents which are implicit in the preceding discourse, according to items of the entry. This means that the actual reason is not clearly stated and has to be inferred from a description which typically spreads over one segment or one or more subsegments. Moreover, the fact that a third-person singular verb and noun phrase forms are used does not guarantee that only one reason is being referred to. Segmentation plays a crucial role in the resolution of such hard cases. In order to clarify how segmentation works and helps anaphora resolution within the approach, the discussion regarding the entry above will be deferred to the next subsection, where an example extracted from the corpus is used to explain segmentation and advocate its use.

3.3 Segmentation based on topicality

The approach to segmentation used in the model is based on topic tracking. It means that the dialogues were analysed for topic continuity and divided in portions of text, called either segments or subsegments, on this basis. This was done because the information structure data is believed to be of help for anaphora resolution. Subsegments are defined as such portions of text which have topics that are subordinate to a higher-level segment topic in the sense that they depend on those segment topics to be properly interpreted semantically and discoursively. This poses the problem of how far the embedding should go, since it is perfectly possible to segment dialogues using a more complex hierarchy of subsubsegments and beyond. The approach presented here uses only segments and subsegments, without any further embedding. The choice is based on the perception that two hierarchical levels seem enough embedding for segmentation meant to support anaphora resolution. No assumptions are made regarding the actual processes involved in the mind of language users. Procedures shown in the development of this subsection should help making the decision process clearer. The discussion of the collocation list entry above will be resumed promptly.

In example (1) below, a token involving the pattern represented in the entry is presented with the segmentation coding added. The token appears at the very end of the passage. The segmentation coding used may be succinctly summarised as follows: the assigned topic for each segment or subsegment is shown between single quotes, following the *s* mark and a number for the segment which identifies it in the dialogue. Subsegments are also numbered within the segment they belong to, but they are marked *ss*. The *ss*-and-number coding is followed by the code for the segment they belong to with a slash to separate them. Segmentation coding is placed in a separate line which precedes the discourse unit and is marked by a star in the first column. The example is taken from the LLC with minor editing.

(1) * s6 ‘enlarged xerox’
A: as I say this is an enlarged xerox of print
this is very easy to do
because we have an enlarger xerox machine
you see
* ss1/s6 ‘chunk’
A: so we just take the er
on our own xerox machine
with scissors and paste
we get the the chunk that we want for a page
then we take it to the enlarger
and blow it up
er such that it then is sufficiently big
and sufficiently to we can er sort of er type onto
the opposite page of the of the Xerox
the adaptations concerned
* ss2/s6 `expense and difficulty of introducing error'
A: this means that we dispense with the erm
difficulty er the expense and difficulty
of introducing error
erm through typing these things out
* ss3/s6 ‘publishing expense’
A: and we also even more importantly
cut down on the publishing expense
cos a printer finds it a bloody sight easier
to set up particularly peculiars like Anglo-Saxon
in terms of letter press than er typescript
* ss4/s6 ‘reason for enlarging’
A: so er that was the reason for doing that

The collocation with the anaphoric demonstrative referred to appears at the end of the passage. The antecedent, that is, the reasons mentioned in the pattern, include *dispensing with the expense and difficulty of introducing error* and *cutting down on publishing expense*. Both have to be inferred from the preceding text, although morphologically similar words act as clues. It is clear, nonetheless, that there is more than one reason for doing the operation referred to as *enlarged xerox*, annotated as the segment topic, and thus the grammatical agreement does not help in the resolution. The fact is registered in the collocation entry, although the unhelpful agreement may be simply a result of a pattern-independent spoken language feature of informality.

The antecedents, that is, the reasons for doing the *enlarged xerox*, are the subsegment topics of the two preceding subsegments, respectively. These subsegment topics are nominalised forms of the discourse chunks classified as the preceding subsegments. Once the dialogue is thus segmented, the subsegments could significantly help in the process of delimiting discourse-chunk antecedents which contained implicit antecedents or were explicit antecedents themselves. Moreover, this segmentation information could help overcome unhelpful agreement and other difficulties resulting from spoken language features.

It is only fair to ask, nonetheless, how accurate or "objective" these topic assignments may claim to be. As mentioned in the introduction to this paper, the notion of topic is notoriously difficult to use, but it is also a recurrent feature of linguistic analysis at the level of discourse. People will have different opinions concerning the topic of a given dialogue or putative segment of dialogue. Some will prefer single noun phrases with a degree of conciseness, or which are literally part of the text; others will suggest discourse chunks which do not conform to phrase structure boundaries, or made-up phrases which are not part of the text but may be better definitions of the topic.

In order to annotate the corpus tokens of anaphora according to the scheme devised, it was essential to find a way of specifying topics which could be reproduced by other people. Although the annotation work was carried out by the author, thus eliminating interannotator disagreement, the scheme is expected to be useful to other people analysing anaphora or trying to implement an anaphora interpreter in a dialogue interface. Therefore, simple procedures were thought out as a means of identifying topics and segment dialogues. They will be presented further below.

Segmentation work showed that specifying global discourse effects, as well as segments, would capture important broader-than-the-segment textual effects. This also seems useful for the resolution of long-distance anaphora and references to recurrent discourse elements. The segment-and-subsegment-topic segmentation scheme was then expanded to include also a global **discourse topic** and **discourse thematic elements**. Procedures were created to specify these discourse-level roles as well. Analytical work has eventually demonstrated that specifying discourse-level roles before local-level segmentation saved time, as the general layout of topic continuity and resumption was obviously best seen from a bird's-eye view of the full dialogue. This made the identification of the discourse topic the starting point of the segmentation coding. The first procedure shown is thus the one used for the specification of the discourse topic.

1. Run a word frequency count for the dialogue
2. Select the five most frequent suitable items, discarding:
   - grammatical words such as pronouns, articles, prepositions and conjunctions
   - noun phrases of unspecific semantic content, such as *thing*, *sort* and *fact*
3. Check the distribution of these items throughout the dialogue, selecting the most evenly distributed for a working hypothesis
4. Lemmatise and reanalyse frequency and distribution
5. Check the position of the first appearance and prefer the candidate closest to the beginning of the dialogue.
6. The discourse topic must be a lexical item explicitly appearing in the dialogue.
7. Objects are preferred over people.[4]

This procedure guaranteed a relatively stable way to identify the discourse topic and the discourse thematic elements. The results of the analysis carried out for the identification of the discourse topic can be used as a starting point for the segmentation (step 1) by means of local-level topic specification, making the precedence of the analytical work meant to specify the discourse topic useful. The procedure for segment and subsegment topic specification and associated segmentation is shown below:

1. Analyse the patterns of lexical items in the dialogue, studying frequency, distribution and concentration in ranges in order to spot potential topics for large segments
2. Analyse the frequencies and distributions in 40-line chunks and integrate results to information from the previous step
3. Request frequency counts for shorter chunks if necessary
4. Check manually by analysing the exchanges in terms of prospection and encapsulation mechanisms in order to spot boundary moves.
5. Analyse the boundary moves in order to establish whether they introduce:

[4] This decision is based on the fact that people are referred to by specific pronouns, reducing the set of possible candidates substantially, whereas pronouns such as *it* and the anaphoric demonstratives have antecedents which are more difficult to identify. The benefits of segmentation, as thought out here, are thus focused on the resolution of harder cases.

- a segment: introduced topic is related to discourse topic but does not develop a current or previous segment topic, being a new local topic
- a subsegment: introduced topic develops current or previous segment topic, being clearly subsumed in this segment topic

6. Prefer a new segment to a subsegment of doubtful subsumption

Looking back to example (1), the segmentation coding presented marks that utterances as I say this is an enlarged xerox of print, so we just take the er…, this means that we dispense with the erm…, and we also even more importantly…and so er that was the reason for doing that were spotted, among others, as possible boundary moves, following initial quantitative procedures to narrow down choices. The full specification of how prospection and encapsulation mechanisms were adapted to a procedure like the ones above is beyond the scope of this paper (but see (Rocha, 1998)). Notice, however, that, in this particular 24-line portion of text, cue words or phrases (as I say, so, this means, and plus so again) definitively play a role. This means that subordinate sentences introduced by because or by non-finite verb forms have been discarded as likely boundary moves. However, two of the clue words or phrases present problems for the segmentation, because *and* is not always a cue word to signal boundary moves, and in fact this use may very well be the marked use rather than the unmarked. The phrase *this means* would require further investigation, although it might prove to be a frequent subsegment boundary move marker.

Other problems involve candidate boundary moves with similar clue words that are not chosen. In this passage, *then we take it to the enlarger* is a possible boundary move. No subsegment boundary was specified at this point because no new discourse entities are included in this particular utterance. It seems counterintuitive to set up a new subsegment with no candidate for topic. However, this requires a degree of processing regarding referring expressions, especially *enlarger,* which must be linked to *enlarger xerox machine,* or else it might be seen as a prospective new subsegment topic. Dangerously, this may result in circularity, as segmentation is carried out in order to help anaphora resolution and should not depend on resolution of any referring expressions. On the other hand, quantitative procedures would have to rely on an assumption of coreference, or else no actual counting could be done.

These challenges appear regularly as the segmentation work proceeds. A number of them have been resolved by using intuition. Others seemed to be more tractable on the basis of the procedures specified. It is certainly true that further work must be done in order to overcome such difficulties satisfactorily. Nonetheless, the procedures above proved very useful as a reference to segmentation work.

The passage ends with one more token of anaphoric demonstrative in a collocation, *for doing that*, a second token of anaphoric demonstrative which also requires complex inference for antecedent identification (in this case, the segment topic, *enlarged xerox*). However, combinations of *do* forms and anaphoric demonstratives have not been classified as collocations, because of the strong variations in resolution paths. Thus, the collocation list does not include every possible collocation that contains an anaphoric demonstrative, but only those which present stable resolution paths to the extent observed in the annotated data.

Summing up, it has been found that certain tokens of phrases containing anaphoric demonstratives were usefully listed as collocations or patterns of co-occurrence, since the path for achieving resolution seemed sufficiently stable to be organised as a resolution path. This usefulness was confirmed in manual tests carried out on a previously annotated test corpus (see Section 4). Concerning automation, the recognition of collocations such as those listed in 3.2 does not seem particularly difficult, as this would amount essentially to pattern matching, although this has not been tested. Nonetheless, this does not mean that resolution paths will be invariably easy, nor that the analysis of larger samples of specific collocations in the list could not disproof the regularities included in the entries so far listed. It does seem to indicate though that the association of co-occurrence patterns for anaphors to resolution paths may yield good results.

As mentioned before, the variety of resolution paths found for *do-that* anaphors could not be summarised in a collocation entry. For the purposes of the collocation list collected during annotation, therefore, it is not classified as a collocation. The second important processing strategy for the resolution of anaphoric demonstratives, **discourse knowledge**, is the form found, within the approach used to systematise information resulting from the analysis of the corpus, to organise the type of resolution path needed for correct antecedent identification in tokens which could not be resolved by means of first-candidate searches or resolution paths associated to identifiable collocations. Both collocation knowledge and discourse knowledge make use of segmentation information to specify resolution paths for a number of patterns specified in their distinct recognition procedures.

3.4 Discourse knowledge

The category **discourse knowledge** was created to classify cases that cannot be solved by means of first-candidate procedures, appear in various contexts that are not suitable for treatment as collocations and, by and large, pose difficult challenges to anaphora interpreters. The proportion of explicit and implicit antecedents is virtually the same in both languages for this processing strategy – 84.7% of the antecedents in English are explicit, compared to 84.4% in Portuguese – but the share of discourse-chunk antecedents is higher in English, reaching 65.5% of the cases, whereas only 48.8% of the cases are discourse chunks in Portuguese.

Annotated corpus data were used to specify a sequence of checks meant to recognise the appropriateness of the discourse knowledge strategy for the resolution of a given anaphor. Once this is achieved, it should be possible to define the resolution path towards the identification of the antecedent on the basis of information associated to the processing strategy. These checks were organised in steps to check on features which corpus analysis has highlighted as relevant. Concerning the collocation knowledge strategy, the checks are plain pattern matching against a previously specified list of collocation entries. If this is unsuccessful, checks for the possible definition of discourse knowledge as the appropriate strategy, such as the syntactic function features shown below, are used:

- anaphor is subject of non-copular verb
- anaphor is subject of link verb
- anaphor is object of verb
- anaphor is complement of copular verb
- anaphor is object of preposition

Once this first level of decision is resolved, similar checks are listed at the subsequent level, such as those used when the anaphor is the subject of a non-copular verb, the first possibility above:

if verb is *explain*-type
  - select previous move or turn as candidate
  - typical verbs
  - *show*; *account*; *explain*

if verb is not *explain*-type
  - identify pronoun

The series of steps above indicate, therefore, that the analysis of corpus data revealed that anaphoric demonstratives in general have the previous move as antecedents whenever the anaphor is the subject of *explain*-type verbs. Verbs typically found in the corpus were added to the information collected in order to help classifying unseen cases. However, if the anaphor is the subject of a verb which is not of the *explain*-type, it is necessary to identify the individual anaphor, as the pronoun *this* and *these* presented a different behaviour from *that* and *those* in the annotated corpus.

By gradually checking features of the context of occurrence against those predicted in the processing strategy checking instructions derived from corpus data, the appropriate discourse strategy should be eventually defined. This is obviously more likely to be successful if based on a larger number of tokens than the one examined here. If the contexts previously defined for a given strategy do not seem to fit the one observed for the anaphor to be resolved, the checking process restarts for the next most likely processing strategy according to corpus data related to the type of anaphor to which the token under scrutiny belongs. Once a given strategy is recognised as the adequate one, the resolution path, also specified on the basis of corpus data, is used to identify the antecedent. This resolution path often – and particularly so in the case of anaphoric demonstratives – uses information structure to identify the correct antecedent.

The processing strategy named **deixis** was classified as the correct one in 13.2% of the cases of anaphoric demonstratives in the English sample, as well as 7.2% of those in Portuguese. Those cases rely on situational information for the identification of an antecedent, which is typically an object, person or fact not explicitly mentioned in the text, nor in any way inferable from textual information. Thus, these are not, strictly speaking, cases of anaphora, but they have been included in the classification for the sake of thoroughness. One would envisage computational systems that would combine visual and textual information to interpret such cases, but the sort of processing involved would lead to the analysis of issues which wander too far away from those discussed in this paper. The next subsection examines the two properties in the model which classify the antecedent.

3.5 The topical role of the antecedent

The properties related to the classification of the antecedent are meant to offer further support for the resolution of anaphora. As said before, type of antecedent refers to the implicit/explicit dichotomy, with nonreferential as a possible third value. The topical role of the antecedent classifies antecedents recognised by corpus analysis on the basis of their roles within the topicality-based segmentation scheme. The annotation of specific cases of anaphora is thus linked to the segmentation coding. As seen, three levels of topicality were defined for the proposed analysis of information structure in dialogues: the discourse level; the local level; and the sublocal level. Each one of these levels was assigned a topic, so that antecedents found in corpus data were classified according to their topical role, that is, as a **discourse topic**, a **segment topic**, or a **subsegment topic**. If the sequence of checks started above were to be continued, and, as a result, the anaphoric demonstrative was identified as *this* or *these*, the next instruction would be to select the local level topic as preferential candidate antecedent and use lexical clues to check whether this noun phrase would be a good fit.

These lexical clues include agreement, selectional restrictions and association history within the dialogue. If there wasn't a good fit, the global level topic would be selected, and, if this was not a good fit either, the sublocal topic would be checked. This instruction is also a result of systematisation out of corpus data, which showed that, differently from most anaphoric demonstratives, subjects of noncopular verbs not of the *explain*-type were more frequently noun phrases, rather than discourse chunks, with these topical roles. Example (2) presents two fully annotated case of anaphoric demonstrative, together with segmentation coding for the segment as a whole.

(2) * s1 `senate committees'
B: I don't know you're on this uh senate committee of course aren't you
* ss1/s1 `combination of subjects'
B: the uh this sort of well for the combination of subjects this (De; ex_2; sst; FtC) is not going to come to anything
nobody really wants it they think it's ideal in one way but it's um - if it's going to fold up
A: mm mm mm
* s2 `letter in the Observer'
A: there was a very nice letter in the Observer on Sunday I don't know whether you noticed
B: I didn't see that (De; ex_4; st; FtC) no

The topic for segment one is *senate committee*; this is marked at its beginning by means of the previously described segmentation coding. The segment is subdivided into one subsegment, which specifies that *combination of subjects* is the issue for the *senate committee* in question. Next, a new local topic is brought up, and thus segment 2 begins with *letter in the Observer* as segment topic. The first anaphoric demonstrative is annotated as being a demonstrative (*De*, the type of anaphor), with an explicit antecedent (*ex*) given number 4 in the sequence of discourse entities, which allows coreference linking. The topical role of the antecedent is of subsegment topic (*sst*) and a first-candidate search would work as processing strategy. The same strategy also works for the second case. Since the resolution of these particular cases is achieved by means of the first-candidate strategy (*FtC*), the topical role information for the antecedent (subsegment topic (*sst*) in the first case and segment topic (*st*) in the second) is not important for resolution, although it may be used as confirmation. However, in cases such as the one in example (3) below, the topical role information is essential, as first-candidate searches do not yield adequate results for the correct identification of antecedents.

(3) * ss2/s6 `structure of faculty boards and boards of study'
A: erm we've got to er decide what the structure
of faculty boards and boards of studies are
going to be in the future
B: is that (ex_46; p_sst; DK;) the actual terms of reference ?

If this occurrence of anaphoric demonstrative was to be resolved, the collocation list would be checked first, as it seems best to start with those processing strategies that offer the fastest way to the definition of a resolution path. As this would fail, the next processing strategy to be checked would be discourse knowledge. The option would be subject of a linking verb, which typically are anaphoric demonstratives with discourse-chunk antecedents related to the local topic, when occurring within a segment, or to the sublocal topic, when occurring within a subsegment. This would confirm the previous move as the best candidate, with minor adaptations, given that the subsegment topic would have been previously defined, which would allow its recognition within the move.

The next section suggests ways of turning the ideas described above into an actual classifier based on information related to processing strategies, antecedent explicitness and topicality collected by annotating a training corpus.

4 Testing the model

Two dialogues – one in the English sample and another in the Portuguese sample – have been annotated but not included in the set of data used to produce the model, as they were set apart for testing purposes from the outset. These two dialogues are therefore a fair test of the predictive powers of the model. Both of them belong to the same corpora from which the databases of anaphora cases used to build the classification model for each language were collected. Testing with dialogues of a different origin is a desirable extension of the project which has not been as yet undertaken.

The testing of the model was carried out essentially manually, using the previously annotated dialogues set aside. In other words, none of the instructions, guidelines or any other aspect of the sequence of checks were automated for the purpose of testing. The analyst simply browsed through the annotated dialogue and stopped whenever an anaphora case was found. The appropriate entry for the type of anaphor in the model was then used to identify the antecedent and produce the resulting annotation. This annotation was then compared to the pre-existing one in order to assess the accuracy of the annotation produced on the basis of the model.

There were 804 cases of anaphora in the test corpus for English. Only six cases led to errors in the recognition of the appropriate processing strategy. This means that, in 99.3% of the cases, the guidelines in the model recognised the processing strategy to be used for the resolution correctly, predicting also the correct type of antecedent. However, in nine of these cases, the resolution path did not include information which could successfully handle the anaphoric reference and identify the correct antecedent, counting thus as secondary errors. Therefore, the correct identification of the antecedent was not possible in 15 cases, lowering the accuracy score to 98.1% of the cases if secondary errors are included.

The results are evidently satisfactory, but they should be seen with due caution. Firstly, the assumptions of segmentation and topical role assignment substantially reduce the difficulty of the task. It must be said, nevertheless, that this provides strong support for the inclusion of topicality as a crucial element in the resolution of anaphoric references. Secondly, these assumptions amount to a very significant boost in the chances of felicitous anaphor resolution, as compared to real-life processing situations, not only because it considers that these topical roles were efficiently assigned, but also because it assumes a full analysis of the entire dialogue prior to the resolution. Thus, the way topicality is used for anaphor resolution “on the fly”, that is, during a real conversation, still has to be accounted for.

Finally, it must be pointed out that the testing was carried out manually by the analyst, who used the information compiled in the model to resolve the anaphor tokens in the test corpus. As much as the analyst strove to be honest and strict in the decisions the testing involved, the encouraging results cannot be taken at face value. Ideally, the model should be transformed into an automatic procedure which would then be applied to a previously annotated dialogue for testing. This would at least eliminate any intuitions about language, biases or excessively tolerant decisions that may have influenced the highly favourable error percentages. The problem with this form of testing is of course that transforming the model into a program is no trivial matter. Alternatively, another analyst might use the guidelines in the model to carry out the testing. However, this solution would not eliminate biases and intuitions altogether, as the decisions would still be made by a human, although it would have the advantage of being someone other than the analyst who thought out the model.

In a balanced appraisal of the results above, it seems safe to assert that such high levels of accuracy cannot be due to chance or plainly to the analyst's leniency. Therefore, one might conclude that the model is at least a step in the right direction as a corpus-based approach to anaphora resolution based on a database of annotated cases. It is probably also true that the level of detail included in the annotation, particularly the topicality-related information, is likely to exert a positive influence over the efficiency of an anaphora interpreter in an NLP system.

Regarding the type of anaphor discussed in this study, there was an unusually high number of anaphoric demonstratives in the test dialogues, which were also resolved more often than expected by means of discourse knowledge. In the sample used to build the model, demonstratives were 10.7 % of the whole sample of anaphors. However, the percentage of anaphoric demonstratives in the test dialogue was 15.01%, and, among these, 67.76% were resolved by means of discourse knowledge, contrasting strongly with the 37% in the sample data. The detailed account which the discourse knowledge subtypes allowed certainly made a difference, as it covered virtually the whole variety of resolution paths needed.

There are a number of other factors in the test dialogue which caused several other significant departures from the probabilities obtained by analysing the sample data. The exact nature of these fluctuations need not be detailed in the current discussion. What is crucial to realise is that variation is unavoidable and unpredictable. The probabilities assigned to the combinations of categories across the variables are a powerful way to predict the patterns of anaphoric reference in a given dialogue, but the only way to achieve robust processing is to record recognition and resolution patterns for as many sub-types as possible, so as to deal with the ever-changing complexities of anaphoric reference in spoken language.

Simplicity is invariably a desirable feature of a theory. Notwithstanding, the model, as it stands, attains an equally desirable comprehensiveness which seems necessary to accomplish the task of describing patterns of anaphoric reference. Further testing may eventually demonstrate that certain instructions or sub-instructions are seldom put to any use or can be subsumed under others. It is in fact likely that actual implementation would impose restrictions on the range of possibilities covered. This study chose to include as many patterns as it could find, avoiding a-priori judgments on what is circumstantial and what is fundamental. Later developments may or may not streamline the model.

5 Suggested approaches to implementation

The treatment of anaphor resolution as a classification problem stems naturally from the corpus-based approach used to analyse anaphoric demonstratives in this study. So far, a few tests have been carried out with decision trees built using the C4.5 program (Quinlan, 1993). The classifiers thus generated seem to be a promising way of turning the sequence of checks above mentioned into an actual anaphora interpreter in a dialogue interface. Since several other machine learning techniques, such as neural nets, can also be used to build classifiers, other possibilities may prove more successful in the future. Unfortunately, testing for actual implementation of such an anaphora interpreter is proceeding at a very slow pace.

The thorniest problem with the approach is probably the automatic recognition of the attributes found to be important by a classifier, regardless of the technique used. Thus, it is possible, although not necessarily easy, for a human analyst to make decisions about topicality which are known to be very complex for automated treatment. Other ancillary techniques, such as checking the adequacy of selected candidate antecedents, may also prove to be far more difficult than thought initially. On the other hand, the potential usefulness of these operations seems undeniable on the basis of corpus analysis.

It does not seem to be so unlikely, however, that a record is kept of all discourse entities introduced during a given human-machine interaction by means of a dialogue interface. Using straightforward statistical tracking of potential elements, typically noun phrases, for the three saliency levels mentioned in the previous section, it may be possible to provide classifiers with information which is accurate enough to replicate the decisions made by the human analyst in what regards the recognition of a processing strategy for a given case of anaphora, together with a choice of an appropriate resolution path associated to features detected in the context of occurrence.

One interesting approach is described in (Karen, 1992). A connectionist network is used to model topic entity selection, using surface features and a bare minimum of semantic information. The model is designed to process sequences of simple clauses in written narrative work, such as stories, and concentrates on the noun phrases in these clauses. These noun phrases are tagged according to their position in the clause, as NP1, NP2, and so forth; sets of units are used to classify each noun phrase according to definiteness/indefiniteness; nine possible semantic classes, such as *human*, *food*, *location*, etc., each one further subdivided into possible different words within the class which are in the lexicon. Thus, *human* can be *boy*, *girl*, *man* or *woman*; and pronouns are classified as *male*, *female* or *neuter*.

The network architecture consists of the context units and topic units. The previous context and previous topic units store information about the preceding text by means of the noun phrases analysed. It combines this information with the new input clause to feed the present context unit, creating a new pattern of activation. The resulting output is represented by two topic units, the current topic and the expected topic units. These topics may or may not be the same, depending on the construction of the clause, which may signal the introduction of a new important entity. A current topic is not generated until the entity occurs at least twice. The network was trained in order to be able to generalise for stories it had not seen before. There are a number of other details that need not be discussed here.

The training and the tests are carried out on simple stories which are not a realistic example of written language and are a far cry from the messiness of spoken discourse. Also the factors defined as playing a role in the selection of topic entities are questionable as good predictors of topicality, and a number of other issues might be raised concerning the experiment with the connectionist network model. Nevertheless, the argument that a connectionist network could be successfully used to handle problems of topicality, combining features at the different linguistic levels to perform the task of topic tracking, seems to be worth pursuing.

Avoiding the complexity of such decisions does not seem to improve in any obvious way the prospects of resolving hard cases, such as those posed by anaphoric demonstratives. The relative shortage of approaches to this type of anaphor in the literature suggests an unwillingness to deal with cases that require a degree of speculation and perhaps unconventional solutions. The exploration of those still rarely investigated approaches may prove its practical fruitfulness not only for anaphoric demonstratives, but also for verb-phrase ellipses and other hard cases of anaphora.

6 Use of the model in machine translation

This study focused mainly on the analysis of anaphoric phenomena in real-life spoken language and does not include, at present, an attempt to implement the proposed model. Nonetheless, the study is concerned with the practicality of future developments and is based on the belief that the approach used will eventually prove useful for the actual resolution of anaphors in real systems. One aspect which stems naturally from the bilingual aspect of the study is the possible application of the information collected by means of the annotation in machine translation. For obvious reasons, most of the discussion related to machine translation concerns written language. Nevertheless, interest in spoken language and possible machine translation applications for this form of discourse have become gradually more frequent in recent research work.

If annotated for anaphoric links, an aligned corpus could support machine translation more effectively, as the translation examples would also provide information about equivalent forms of anaphora across languages. Thus, some difficulties likely to arise from a strictly example-based approach to translation might be minimised. The classification of a processing strategy for anaphoric reference could identify, for instance, collocations which used different types of anaphor in different languages. If the case of English and Portuguese is considered, for instance, a number of collocations which use pronouns in English should be replaced by plain verbs with omitted subjects in Portuguese, as the ratio of anaphoric demonstratives per total number of words has signalled.

For instance, if an entry in the collocation list, such as *that's (not) true* had to be translated, an example-based approach relying on an aligned corpus would be likely to have renditions such as *(não) é verdade* e *isso (não) é verdade*, the former with the omission of the subject. On the other hand, phrases such as *that's why* are almost invariably translated into Portuguese as *é por isso que* or *por isso que*, in which the demonstrative *isso* is never omitted. Annotation for anaphora could be used to distinguish cases of translation in which omission is acceptable, that is, in which an anaphoric verb form is acceptable or even the norm; and those which never accept omission.

This could be accomplished on the basis of a combined analysis using examples of translation and collocation lists containing anaphors for both languages, as the examples would suggest renditions using different kinds of anaphor in each one of the languages. It would also be possible to improve general results of an automatic translator by checking whether renditions refer to discourse entities as expected, taking the source text as a standard. One might think, ideally, of a capability to verify the accuracy of the full referring picture in a given translation by checking whether anaphoric devices used conform to patterns specified previously for the purposes of automatic annotation.

7 Conclusion

The paper presented a model for the classification of anaphoric phenomena in dialogues in English and Portuguese. The model is substantiated in an analytical tool, an annotation scheme, which attempts to cover all possible forms of anaphoric linkage, understood in as broad a way as possible. Four properties were thus specified as a useful way of codifying linguistic facts observed in manually analysed dialogue corpora. Notably, these properties include a processing strategy, which is an attempt to classify the type of knowledge involved in the resolution of a given case; and a topical role of antecedent, which is an attempt to relate each antecedent to a segmentation coding which implements a topicality-based approach to hard cases of anaphora. As reported, the model has only been tested manually.

Both properties require decisions by the analyst which are often made on the basis of intuition, although the repeated analysis of cases seems to result in an equilibrium in which common properties are used to group cases according to categories used as values within these properties. Regularities observed have been reliable enough, nonetheless, to offer grounds for the systematisation of procedures, in the case of topicality-based segmentation. Once topics for the segments and subsegments have been assigned, the antecedents can have their topical roles classified promptly.

The process of recognising the adequacy of a given processing strategy was also made into a series of steps based on annotated corpus data. Each processing strategy was associated to a resolution path on the same basis. Bearing in mind corpus size, it is easy to perceive that some patterns, at least, were established on the grounds of a very small number of cases. Nonetheless, the detailed specification of procedures for the recognition and resolution of appropriate processing strategies proved successful in manual tests. Usual restrictions to manual testing and small sample size apply, and further work is needed, particularly in what regards the reliability of the observed patterns and actual testing in computers.

The sketchy discussion of possible developments and applications for the annotation scheme is, thus, speculative in the sense that none of the suggestions brought up have been tested or implemented in real systems. It is believed, nevertheless, as a result of the experience in carrying out the study, that these possibilities should be at least broadly outlined. If nothing else, the suggestions serve the purpose of conveying the conviction that the present study is a first step in the development of a methodology to deal with anaphoric phenomena in spoken language, in both linguistic research and natural language processing.

References

Byron, D. and J. Allen 1998. “Resolving Demonstrative Anaphora in Trains93 Corpus.” *Proc. of the 2nd Discourse Anaphora and Anaphora Resolution Colloquium* (DAARC2).

Grosz, B. and C. Sidner 1986. “Attentions, Intentions and the Structure of Discourse.” *Computational Linguistics* 12(3).175-204.

Karen, L. 1992. “Identification of Topical Entities in Discourse: A Connectionist Approach to Attentional Mechanisms in Language.” *Connectionist Natural Language Processing*. N. Sharkey (ed.), Ch. 13, 274-95. Oxford: Intellect.

Ng, V. and C. Cardie 2002. “Improving Machine Learning Approaches to Coreference Resolution. *Proc. of the 40th Annual Meeting of the ACL* (ACL-02), 104-11.

Passonneau, R. 1993. “Getting and Keeping the Center of Attention.” *Challenges in Natural Language Processing.* M. Bates and R. M. Weischedel (eds.), 179-226. Cambridge: Cambridge University Press.

Paul, M., K. Yamamoto and E. Sumita 1999. “Corpus-based Anaphora Resolution: Towards Antecedent Preference. Coreference and its Applications.” *Proc. of the ACL Workshop.* A. Bagga, B. Baldwin and S. Sheldon (eds.). ACL.

Quinlan, J. 1993. *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann.

Rocha, M. 1997. “Supporting Anaphor Resolution in Dialogues with a Corpus-based Probabilistic Model. Operational Factors in Practical, Robust Anaphora Resolution in Unrestricted Texts.” *Proc. of the ACL Workshop*. R. Mitkov and B. Boguraev (eds.), 54-61. ACL.

Rocha, M. 1998. “A Corpus-based Study of Anaphora in Dialogues in English and Portuguese.” PhD Thesis. Falmer, Brighton: School of Cognitive and Computing Sciences, University of Sussex.

Sinclair, J. 1992. “Priorities in Discourse Analysis.” *Advances in Spoken Discourse Analysis*. Malcolm Coulthard (ed.). London: Routledge.

Soon, W., H. Ng and D. Lim 2001. “A Machine Learning Approach to Coreference Resolution of Noun Phrases.” *Computational Linguistics* 27.521-544.

Focus, Activation, and *This*-Noun Phrases: An Empirical Study

Massimo Poesio[1] and Natalia N. Modjeska[2]
[1]*University of Essex, Computer Science and Cognitive Science*
[2]*University of Edinburgh, Informatics*

We report on an analysis of the use of THIS-NPs: noun phrases with *this* or *these* as determiners, and the demonstrative pronouns *this* and *these*. We test the THIS-NP hypothesis, a refined and clarified summary of earlier proposals, such as (Linde, 1979; Gundel *et al.*, 1993; Passonneau, 1993), by way of a systematic analysis of the uses of these NPs in two different genres. In order to carry out the analysis, we used the parametric techniques for focus tracking from (Poesio *et al.,* 2004), and devised a reliable annotation scheme for classifying THIS-NPs in our corpus as *active* or not. 92% of THIS-NPs in our corpus were classified as referring to entities which are active in this sense. We tested three formalizations of the THIS-NP hypothesis. The version that received most empirical support is the following: THIS-NPs are used to refer to entities which are *active* but not the *backward-looking centre of the previous utterance*.

1 Motivation

Noun phrases with *this* or *these* as determiners and the demonstrative pronouns *this* and *these* (THIS-NPs henceforth)[1] are known to have at least three different functions. In formal semantics and pragmatics, they have mostly been studied for their deictic function, i.e., their ability to refer to objects in the visual situation, and particularly to objects the speaker is pointing at (Kaplan, 1979; Jarvella & Klein, 1982; André *et al.*, 1999).

(1) A [salesman on a car lot, nodding in the direction of a nearby cluster of trucks:]
These vehicles have four-wheel drive. (Roberts, 1999)

It is, however, well known that THIS-NPs can be used in other ways as well; and indeed, preliminary analyses of the corpus used in this study suggested that only about 39% of THIS-NPs were cases of visual deixis (Poesio, 2000). A second function of 'demonstrative' NPs was studied by Linde (1979), Gundel *et al.* (1993), and Passonneau (1993), among others. These authors pointed out that *pronominal* THIS-NPs in particular[2] are often used to refer to a discourse entity that is discourse-old, but not 'in focus'. In the following example, Martin Carlin is clearly established as 'focus' by the time utterance (2d) is

[1] We will mostly avoid the use of the term 'demonstrative' as the starting point of this research is the realization that not all these uses are 'demonstrative' in Kaplan's sense (Kaplan, 1979). We are concentrating on THIS-NPs because our corpus contains very few cases of *that* noun phrases.
[2] Passonneau studied the use of *that* rather than *this*.

encountered, but the antecedent of the THIS-NP *this area*, the Faubourg Saint-Antoine, is clearly not the main topic of conversation at this point.

(2) a. In spite of his French name, **Martin Carlin** was born in Germany and emigrated to Paris to become an ébéniste.
b. **He** settled there with other German and Flemish craftsmen and took employment in the workshop of Jean-François Oeben, whose sister **he** married.
c. Inventories made after **Carlin**'s death show that **the ébéniste** and **his** wife lived modestly in a five-room apartment in THE FAUBOURG SAINT-ANTOINE, an unfashionable quarter of Paris, with simple furniture, a few pastel portraits, and a black lacquer clock.
d. Few of **Carlin**'s wealthy clientele would have cared to venture into THIS AREA.

Finally, it is also known (Webber, 1991; Asher, 1993) that THIS-NPs can be used to refer to abstract objects such as propositions or plans (Webber used the term DISCOURSE DEIXIS for these cases) as in the following example:

(3) For example, binocular stereo fusion is known to take place in a specific area of the cortex near the back of the head. Patients with damage to this area of the cortex have visual handicaps but they show no obvious impairment in their ability to think. This suggests that stereo fusion is not necessary for thought. (Webber, 1991)

What the cases of discourse deixis and visual deixis, and the cases studied by Linde and Passonneau, have in common is that the THIS-NP is used to refer to an entity which, while salient, is not the current 'topic' or 'discourse focus' (we are deliberately using these terms in a vague way here). This intuition is at the heart of Passonneau's account of the anaphoric use of demonstrative pronouns, as well as of Gundel *et al.*'s account of the use of THIS-NPs (1993). We concentrate on Gundel *et al.*'s proposal in what follows.

Gundel *et al.*'s theory of the conditions under which referring expressions are used assumes that two factors interact in determining the choice of referring expression. The first of these factors is the ACTIVATION HIERARCHY: a speaker's choice of expression depends in part on assumptions about the 'cognitive status' of the referent in the hearer's information state. Gundel *et al.*'s 'activation levels' range from TYPE IDENTIFIABILITY for indefinite NPs, to IN FOCUS for pronouns.

In focus	>	Activated	>	Familiar	>	Uniquely identifiable	>	Referential	>	Type identifiable
it		*that, this, this N*		*that N*		*the N*		indefinite *this N*		*a N*

The second factor playing a role in Gundel *et al.*'s account are Grice's maxims of quantity:

Q1 Make your contribution as informative as possible

Q2 Do not make your contribution more informative than necessary

These maxims prevent the use of referring expressions associated with higher activation levels to refer to entities with a lower status, as we will see.

Gundel *et al.* propose that the use of THIS-NPS, as well as of pronoun *that*[3] requires the referent to be ACTIVATED, which status they characterize as "being represented in current short-term memory."[4] This condition would also license the use of THIS-NPS to refer to entities in focus; what prevents this, according to Gundel *et al.*, is Grice's Q1. Because a more specific referring form exists, the use of a demonstrative for entities in focus would implicate a lower activation level. We will merge these claims into a single hypothesis as follows:

> **The THIS-NP Hypothesis**: THIS-NPS are used to refer to entities which are ACTIVATED. However, pronouns should be preferred to THIS-NPS for entities IN FOCUS.

We believe the class of proposals exemplified by Gundel *et al.* can be made at the same time more broad in their coverage and more precise by (i) specifying which entities are supposed to be 'in focus' and (ii) being more explicit about the types of entities that can be 'in short term memory' without being 'in focus'. We carried out a corpus investigation aimed at refining, clarifying and testing these claims.

2 The GNOME corpus

In recent years, there has been an increasing interest in corpora as a means to explore linguistic generalizations, and increased sophistication in the methods used. Improvements include better techniques for storing and annotating language corpora, based on annotation standards such as XML. They also include techniques for measuring the RELIABILITY of a given annotation scheme (Passonneau & Litman, 1993; Carletta, 1996).

One of the main motivations for this work was that we felt that we could improve upon previous analyses of the uses of THIS-NPS by building on the results of our previous corpus-based studies of the uses of referring expressions and of salience (Poesio *et al.*, 2004; Poesio, 2000). One of the useful outcomes of this work is the GNOME corpus, in which NPS, the anaphoric relations between them, and their visual deixis status, have been marked in a reliable way (Poesio, 2000). We discuss the corpus in this section. A second resource that we could use are the methods for tracking the BACKWARD-LOOKING CENTRE, or CB (Grosz *et al.*, 1995; Walker *et al.*, 1998) – a well-known formalization of the notion of 'local focus' – automatically, instead of relying on

[3] But not of full *that* NPS, which only require the referent to have the lower 'familiar' status.

[4] In fact, for THIS-NPS, Gundel *et al.* claim that the referent has to be speaker-active, i.e., introduced by the speaker.

hand-identification, proposed in (Poesio *et al.*, 2004). These methods also allowed Poesio *et al.* (2004) to test several definitions proposed in the literature, among which they were able to find the 'best' (i.e. those which resulted in fewer violations of the claims of Centering theory). We summarize this study in the next Section. These two previous pieces of work allowed us a more systematic exploration of the conditions under which a THIS-NP was licensed.

2.1 The GNOME corpus: markup scheme

The annotation of the gnome corpus followed a systematic manual, available from the GNOME project's home page at `http://www.hcrc.ed.ac.uk/~gnome`; here, we discuss the most important details of the scheme. All units of text in the GNOME corpus that might be identified with utterances (in the Centering sense) are marked as `<unit>` elements; the attributes of such elements identify finite and non finite clauses, and various other properties. Each NP is marked with a `<ne>` tag and with a variety of attributes capturing syntactic and semantic properties. Important attributes for our purposes are cat (specifying the type of an NP), gf specifying its grammatical function, deix (whether the object is a visual deictic reference or not) and generic (whether the NP denotes generically or not).

A separate `<ante>` element is used in the GNOME scheme to mark anaphoric relations; the `<ante>` element itself specifies the index of the anaphoric expression and the type of semantic relation (e.g. identity), whereas one or more embedded `<anchor>` elements indicate possible antecedents (the presence of more than one `<anchor>` element indicates that the anaphoric expression is ambiguous):

(4)
```
<unit finite='finite-yes' id='u227'>
    <ne id='ne546' gf='subj'> The drawing of
            <ne id='ne547' gf='np-compl'>the corner cupboard, </ne>
    </ne>
    <unit finite='no-finite' id='u228'>or more probably
        <ne id='ne548' gf='no-gf'>an engraving of
                <ne id='ne549' gf='np-compl'> it </ne></ne>
    </unit>,
...
</unit>
<ante current="ne549" rel="ident"><anchor ID="ne547"></ante>
```

2.2 The GNOME corpus: texts

The GNOME corpus currently includes texts from three domains; texts from two domains were used in this study. The MUSEUM SUBCORPUS consists of descriptions of museum objects and brief texts about the artists that produced them.[5] The PHARMACEUTICAL SUBCORPUS is a selection of leaflets providing the patients with mandatory information about their medicine.[6] Each subcorpus contains about 6,000 NPs; in this study, we used texts from the first two domains, for a total of about 3,000 NPs, including 112 THIS-NPs. As for utterances, the corpus includes about 500 sentences, and 900 finite clauses; the actual number of utterances used in the study is one of the parameters that we varied, as discussed below.

3 Methods

For this study, we relied on some of the existing annotation of the GNOME corpus, as well as extending it. Our approach can be summarized as follows:

1. Provide a characterization of 'activated' that can be reliably annotated, and mark all 'activated' entities in the corpus, utterance by utterance.
2. Compute all entities which are 'in focus' at a given time, using notions from Centering theory (Grosz *et al.*, 1995) like CB and CP to come up with several alternative definitions of the notion of 'in focus,' and previously developed methods for automatically computing the CB and CP in the annotated corpus according to several ways of defining the notions of 'utterance' and the ranking function, using the existing annotation.
3. For each THIS-NP, check whether it refers to an entity which is activated, and whether the entity is in focus.

3.1 Specifying 'In Focus'

Centering Theory The notion of 'topic' or 'discourse focus' is notoriously difficult to formalize. We used as the basis for our investigation of this notion the terminology and ideas introduced in Centering Theory by Grosz *et al.* (1995) and Walker *et al.* (1998), in particular the notions of Backward-Looking Centre (CB) and Preferred Centre (CP). In the 'mainstream' version of Centering by Grosz *et al.* (1995), it is assumed that each UTTERANCE introduces new discourse entities (or Forward-Looking Centres) into the discourse, and in so doing, updates the 'local focus'. It is further assumed that the discourse entities introduced (or better, REALIZED) by an utterance are ranked; the most highly ranked entity in an utterance is called the CP. The CB is Centering's equivalent of the notion of 'topic' or 'focus', and is defined as follows:

---

[5] The museum subcorpus extends the corpus collected to support the ILEX and SOLE projects at the University of Edinburgh. ILEX generates Web pages describing museum objects on the basis of the perceived status of its user's knowledge and of the objects she previously looked at (Oberlander *et al.*, 1998). The SOLE project extended ILEX with concept-to-speech abilities, using linguistic information to control intonation (Hitzeman *et al.*, 1998).

[6] The leaflets in the pharmaceutical subcorpus are a subset of the collection of all patient leaflets in the UK which was digitized to support the ICONOCLAST project at the University of Brighton, developing tools to support multilingual generation (Scott *et al.*, 1998).

CB CB(U_i), the BACKWARD-LOOKING CENTRE of utterance U_i, is the highest ranked element of CF(U_{i-1}) that is realized in U_i.

It should be noted that Centering provides no definition of the notions of 'ranking', 'utterance' and 'realization'; researchers using the theory have to specify their own. For her comparative study of *it* vs. *that*, Passonneau developed her own notion of CB, that she called 'local centre'. Poesio *et al.* did a comparative analysis of several ways of 'filling in' Centering Theory's parameters (Poesio *et al.*, 2004). One of the results was that Passonneau's notion of local centre, while predicting pronominalisation well (virtually all realizations of discourse entities that were 'local centre' in Passonneau's sense were pronominalised), is very restrictive: only about 20 utterances out of 500 in the corpus have a 'local centre' in Passonneau's sense. Two instantiations of the notion of CB gave the best (and pretty much equivalent) results. Both of these involved identifying utterances with sentences, and allowing for indirect realization of the CB. However they differed in the ranking function: in one case, grammatical function (subjects rank more highly than objects that rank more highly than adjuncts) augmented with a linear disambiguation factor; in the other, Strube and Hahn's (1999) ranking function based on 'information status' (according to which hearer-old entities are ranked more highly than inferrables, which in turn are ranked more highly than hearer-new entities (Prince, 1992)).

Using Centering Theory to specify the notion of 'in focus' There are three natural ways of using Centering theory to formalize Gundel *et al.*'s idea that speakers adhering to Grice's maxims would prefer not to use THIS-NPS to refer to entities 'in focus':

1. THIS-NPS are preferentially used to refer to entities other than CB(U_i), the CB of the utterance containing the THIS-NP.
2. They are used to refer to entities other than CB(U_{i-1}), the CB of the *previous* utterance.
3. They are used to refer to entities other than CP(U_{i-1}), the most highly-ranked entity of the *previous* utterance.

We considered all three of these possibilities, under all of the 'best' ways of specifying the parameters of Centering Theory identified by Poesio *et al.*: identifying utterances with either sentences or finite clauses; considering both ranking based on grammatical function and ranking based on 'information status'; and allowing for both direct and indirect realization.

3.2 Clarifying 'Activated'

Types of Activated Entities The second aspect of the THIS-NP hypothesis that needs clarification is what it means for an entity to be 'activated'. According to

Gundel *et al.*, an entity is 'activated' if it is in 'current short term memory'. But 'short-term memory' could be taken to include all discourse-old entities, or all and only entities in the global focus (Grosz & Sidner, 1986), or all and only entities in the local focus. It's also not clear how discourse-deictic entities could be taken to be part of short-term memory.

Our starting point in attempting to define the notion of 'activated' more precisely was the hypothesis that an entity is 'activated' if it is in the GLOBAL FOCUS and is 'sufficiently salient'. The notion of global focus was introduced by Grosz and Sidner (1986) to characterize the entire set of entities which are in some sense part of the attentional state of the participants of a discourse. Grosz and Sidner do not provide a fully explicit specification of what's in the global focus, but they do introduce a few key ideas useful to make the idea of activation more specific. These ideas were further developed in subsequent research; our proposals here are mostly based on the formalization proposed in (Poesio, 1993, 1994; Poesio & Traum, 1997). One contribution of this later work that matters in the context of demonstratives, which often have a deictic interpretation, is the partial clarification of the relationship between linguistic information and visual information in the global focus. Poesio argued that Grosz and Sidner's hypothesis that the entire global focus can be formalized in terms of a single structure, a stack, is probably not right, and it is best to hypothesize separate structures for the 'linguistic' component of the global focus and for its 'visual' component, the former having a stack-like structure and containing every discourse entity introduced by a construction algorithm similar to that of DRT,[7] the latter having a situation-based structure and containing every entity in the visual scene. We will make similar assumptions here, assuming that anaphoric and discourse deictic THIS-NPS take their interpretation from the stack, whereas (visually) deictic THIS-NP get their interpretation from the visual component of the global focus.[8]

One of the key aspects of Grosz and Sidner's model, from our perspective, is the idea that the global focus – both in its 'linguistic' and in its 'visual' component –implicitly contains all sort of entities beyond those explicitly mentioned or currently in the visual focus of attention; these entities are added to the global focus either as a result of immediate inference, or as a result of search procedures activated by the use of an anaphoric expressions (Haviland & Clark, 1974). Grosz (1977) and Sanford and Garrod (1981) introduced the term

[7] Similar proposals are also made in SDRT (Asher, 1993; Lascarides & Asher, 1993).
[8] Walker (1998) suggests that the global focus has a cache structure. We will not be concerned with this issue here, except to notice that a cache model doesn't automatically give us a definition of 'activated': we still need to explain how entities in the visual situation enter in the cache.

IMPLICIT FOCUS to refer to this aspect of the global focus. One of the central functions of THIS-NPs is to introduce into the discourse objects previously part of the implicit focus. The problem is that we don't have a fully worked out theory of which entities are in the implicit focus; the best we can do at the moment is an analysis by cases, as done, e.g., by Byron (2002).

Kamp and Reyle (1993) discuss in some detail one type of entity that can be 'in the implicit focus' in this way, plural entities such as *they* in the following:

(5) John met Mary at the movies. <u>They</u> had both gone to see an old French film.

Webber (1991) and Asher (1993) analysed in detail a second type of entity that can enter 'implicit focus,' propositions, as seen in example (3). A third type of entity that can enter implicit focus, and not previously discussed (to our knowledge) in connection with uses of THIS-NPs, are *types*, in the broad sense, i.e. references to concepts whose instantiations are explicitly mentioned in the discourse. References to types take a variety of forms, the simplest among which is simply a full demonstrative like *this type* or *this kind*:

(6) A great refinement among armorial signets was to reproduce not only the coat-of-arms but the correct tinctures; they were repeated in colour on the reverse side and the crystal would then be set in the gold bezel.
Although the engraved surface could be used for impressions, the colours would not wear away.
The signet-ring of Mary, Queen of Scots (beheaded in 1587) is probably the most interesting example of <u>this type</u>; ...

More complex references to types refer to concepts introduced only very implicitly in the text; in this case, the process of adding the antecedent to the implicit focus, to the extent that it actually takes place (Poesio & Reyle, 2001), appears to be driven entirely by the use of the demonstrative:

(7) The craftsmen also bent carefully over cheaper metals or glass to create the jewelry that would adorn the arm of the humble servant girl, or the ordinary, insignificant woman, and would accompany her to her final resting place. <u>This yearning for embellishment, this special relationship between a woman and her jewelry</u> emerges quite clearly here ...

Certain types of ellipsis can also be considered as references to a type:

(8) The cutouts of the first are decorated with griffins set in rectangular panels, and <u>those of the second</u> with large buttons ...

On the basis of these considerations, we came up with a list of cases in which entities can be considered as being 'activated' precise enough that can be reliably annotated. To avoid confusions with Gundel *et al.*'s more general formulation, we introduce a new term, ACTIVE, for our own characterization. An entity is ACTIVE if that entity

1) is in the visual situation; or
2) is a CF of the previous utterance; or
3) is part of the implicit linguistic focus. We only consider as part of the implicit linguistic focus those entities that can be CONSTRUCTED out of *the previous utterance*. An entity can be constructed out of an utterance if:
   a) it is a plural object whose elements or subsets have been explicitly mentioned in that utterance; or
   b) it is an abstract entity introduced by that utterance. We consider two types of abstract entities:
      i) propositions
      ii) types

The Markup Scheme for Active Entities

We tested our hypothesis by classifying the THIS-NPs in our corpus as active or not. In this section, we discuss the markup scheme derived from the definition of 'active' above.

The annotation scheme developed in (Poesio, 2000), together with the focus tracking methods developed in (Poesio *et al.*, 2004), allowed us to classify two of the uses of THIS-NPs discussed in the literature: 'focus-shifting' uses and visual deixis. The existing annotation also already identified plural references to entities in the implicit focus. What was missing was a way to identify references to abstract entities in the implicit focus. The problem we had to face was that while developing the GNOME scheme we had found – as others did (Eckert & Strube, 2001; Navarretta, 2000) – that identifying the antecedents of 'discourse deictic' expressions in the broad sense (i.e., expressions referring to – typically, abstract – entities introduced in the discourse indirectly, such as propositions) is very hard, especially when the annotation produces something less than a full logical form in, say, the DRT sense (Kamp & Reyle, 1993). However, we had also found for the case of visual deixis that in some cases while *identifying* the antecedent of an expression is quite hard, *classifying* an NP as deictic is easier. This proved to be the case for discourse deixis, as well. As a result, we developed a scheme for classifying THIS-NPs that does not require the annotators to mark up the 'antecedent' of the expression. The annotators are instructed to follow the decision tree below:

1. If (i) an <ante> element has *not* been marked up specifying an anaphoric relation of type ident between a <ne> and a previous entity, and (ii) the <ne> is visually deictic (its deix attribute has value yes), classify it as **visual deixis**. (And therefore, active.)
2. Else, if the THIS-NP is connected by an <ante> elements to a previous <ne> by an identity relation, mark it as **anaphoric**. (This applies whether the entity is singular or plural.)

3. Else, if the THIS-NP is a plural entity which contains as elements entities previously introduced, mark it as **plural**;
4. Else, if the THIS-NP involves an elliptical reference to a previous entity (as in (8)), mark it as **ellipsis**;
5. Else, if it is a (non explicitly mentioned) temporal entity, mark it as **time**;
6. Else, if the <ne> is marked as generic, and its instances are concrete objects, mark it as **type**;
7. Else, if the NP refers to an abstract object 'introduced' only implicitly by the previous discourse, such a proposition or an abstract concept, mark it as **discourse deixis**;
8. Else, mark it as **problem**.

We tested the reliability of this scheme by measuring the agreement among ourselves on about 87 THIS-NPs in the corpus. We disagreed on 3 THIS-NPs and 5 were classified as problematic; with 6 possible values, we get κ = .82, significant at the .01 level.

4 Results

4.1 The distribution of THIS-NPs

All THIS-NPs in our corpus not classified as 'problem' were active in the sense above. The observed distribution of THIS-NPs in our corpus is as follows:

Class	Number (Percentage)
Anaphora	45 (40%)
Visual Deixis	28 (25%)
Discourse Deixis	19 (17%)
Type	9 (8%)
Plurals	1
Ellipsis	1
Time	1
Problem	5
Disagreement	3
Total	112

4.2 The correlation between focus and THIS-NPs

For each instantiation of the notion of 'in focus', described in Section 3.1, we observed some variation depending on the values of parameters, but the results were nevertheless clear-cut:

- We found between 8 and 11 violations to the hypothesis that a THIS-NP is used to refer to entities other than CB(U_{i-1}), which is therefore verified by 90-93% of THIS-NPs;
- the hypothesis that THIS-NPs are used to refer to entities other than CP(U_{i-1}) is verified by 75-80% of THIS-NPs;
- the hypothesis that a THIS-NP is used to refer to entities other than CB(U) is verified by 61-65% of THIS-NPs;

Interpreting 'not in focus' as 'not CB(U_{i-1})' leads to better empirical results. (All the differences are significant.)

4.3 Violations of the THIS-NP hypothesis

We analysed the 13 uses of THIS-NPs that were exceptions to the THIS-NP Hypothesis even under its best-performing version.[9] All entities referred to by THIS-NPs in the violation examples are IN FOCUS; this is consistent with the proposal by Gundel *et al.*, provided that reasons for violating the Maxim of Quantity are found. We tested whether pronouns could be used in place of THIS-NPs in these examples, and found that these cases can be divided in three classes: (i) 5 cases in which pronominalisation is possible, (ii) 3 cases in which a pronoun would be possible but awkward, and (iii) 4 cases in which a pronoun would seem rather infelicitous. We discuss some of these cases below.

In (9), a pronoun could be used instead of *this work*, although a slight change in word order would make the example sound better: replacing *appears twice on this work* with *appears on it twice* rather than *appears twice on it*:

(9) The fleurs-de-lis on the top two drawers indicate that the cabinet was made for Louis XIV. As it does not appear in inventories of his possessions, it may have served as a royal gift. The Sun King's portrait appears twice on <u>this work</u>.

In (9), the last mention of the entity was made via a pronoun, but note that another entity is pronominalised in the same sentence, *his*, i.e., *Louis XIV*, and the focus subsequently shifts to that entity, *the Sun King* in sentence three. By using a demonstrative, rather than a pronoun, the speaker seems to prepare the listener for this shift.

In two examples, pronominalisation is possible, even if the referent is mentioned after a paragraph break:

(10) Modeled in the form of three of laurel branches tied with a ribbon, these massive wall lights with their detailed chasing and burnishing reveals the extraordinary skill of their maker, a silversmith to Louis XV, King of France. Each wall light is slightly different, and no one model repeats another.
<u>These four wall lights</u> are among eight made in 1756 ...

In (11), the entity is also mentioned after a paragraph break. A pronoun would be possible but awkward:

(11) Do not keep your patches if your doctor decides to stop treatment. Return them to your pharmacist who will arrange for their destruction.
REMEMBER <u>these patches</u> are only for you.

---

[9] The two instantiations of the ranking function – grammatical role and information status – both resulted in 11 violations, but they differed slightly as to which examples they produced.

This example seems to differ from (10) in that there is an implicit argument of the imperative (*you*), which perhaps is more salient than the referent of *these patches*.

Example (12) – note the two THIS-NPs – is a quite interesting example for which we do not have a ready analysis:

(12) This brooch is made of titanium ... It was made by Anne-Marie Shillitoe, an Edinburgh jeweller, in 1991. It's a good example of a modern material being used in jewelry. In fact, this piece is not one of the very earliest examples of titanium jewelry; The technique for colouring in this piece has already become quite sophisticated.

We believe that a 'principle of variety' is at play here and interacts with the principle of always using the stronger form possible – for the first instance of *this piece* (see also (Poesio *et al.*, 2004)). But it is not clear why a pronoun would appear awkward in the later occurrence.

In some cases, a pronoun would seem rather infelicitous, in particular if the antecedent of the THIS-NP occurs in a title:

(13) **Has the cream any side effects?**
Most people find using this cream causes no problems when used in the right amount ...

Example (14) shows another example in which a pronoun would be infelicitous:

(14) This piece is also a brooch, but from an earlier period; around 1920. It's particularly interesting because of its colour scheme. Purple, white and green were the colours of the suffragette movement; women would wear a brooch like this to show solidarity or affiliation with the movement.

We believe that the infelicity here arises from a reference to a type. As we showed in section 3.2, reference to types is one of the conditions that licenses THIS-NPs.

5 Discussion and conclusions

We reported on an empirical investigation into the use of THIS-NPs in two genres: museum descriptions and pharmaceutical texts. The THIS-NP Hypothesis that we tested extends and clarifies previous proposals on the conditions that license the use of this-NPs. Specifically, we proposed a specific definition of what it means for an entity to be 'in focus' and provided a detailed analysis of a subset of the cases in which an entity can be considered 'activated'; we introduced the term 'active' to refer to these cases. We devised and tested a reliable annotation scheme for classifying THIS-NPs as active. Three instantiations of the THIS-NP Hypothesis were tested on our data. The version that leads to the fewest violations of the Maxim of Quantity is the following:

The THIS-NP Hypothesis: THIS-NPs are used to refer to entities which are ACTIVE in the sense specified above. THIS-NPs should be preferred to pronouns for entities other than CB(U_{i-1}).

From a semantic perspective, we believe that this work – both the results of our corpus analysis and our 'salience-based' analysis of the use of THIS-NPs – are in agreement with the spirit, if not all details, of recent work challenging Kaplan's 'referential' analysis of demonstratives (Roberts, 1999; Zeevat, 1999). We leave for future work a detailed comparison between the THIS-NP hypothesis and these recent presuppositional accounts.

Acknowledgments

Part of this research was done while Massimo Poesio was at the University of Edinburgh, supported by an EPSRC Advanced Research fellowship, and while Natalia N. Modjeska was supported by the EPSRC grant No. GR/M75129.

References

André, E., M. Poesio and H. Rieser (eds.) 1999. *Proc. of the ESSLLI Workshop on Deixis, Demonstration and Deictic Belief in Multimedia Contexts*. Utrecht: Univ. of Utrecht.

Asher, N. 1993. *Reference to Abstract Objects in English*. Dordrecht: D. Reidel.

Byron, D. 2002. "Resolving Pronominal References to Abstract Entities." *Proc. of the ACL,* 80-87.

Carletta, J. 1996. "Assessing Agreement on Classification Tasks: The Kappa Statistic." *Computational Linguistics* 22.249-254.

Eckert, M. and M. Strube 2001. "Dialogue Acts, Synchronising Units and Anaphora Resolution." *Journal of Semantics*.

Grosz, B. 1977. *The Representation and Use of Focus in Dialogue Understanding*. Ph.D. thesis, Stanford University.

----------, A. Joshi and S. Weinstein 1995. "Centering: A Framework for Modeling the Local Coherence of Discourse." *Computational Linguistics* 21.202-225.

---------- and C. Sidner 1986. "Attention, Intention, and the Structure of Discourse." *Computational Linguistics* 12(3).175-204.

Gundel, J., N. Hedberg and R. Zacharski 1993. "Cognitive Status and the Form of Referring Expressions in Discourse." *Language* 69.274-307.

Haviland, S. and H. Clark 1974. "What's New? Acquiring New Information as a Process in Comprehension." *Journal of Verbal Learning and Verbal Behavior* 13.512-521.

Hitzeman, J., A. Black, P. Taylor, C. Mellish and J. Oberlander 1998. "On the Use of Automatically Generated Discourse-level Information in a Concept-to-speech Synthesis System." *Proc. of the Int. Conference on Spoken Language Processing* (ICSLP98), Australia.

Jarvella, R. and W. Klein (eds.) 1982. *Speech, Place and Action – Studies in Deixis and Related Topics*. New York: John Wiley.

Kamp, H. and U. Reyle 1993. *From Discourse to Logic*. Dordrecht: D. Reidel.

Kaplan, D. 1979. "Dthat". *Syntax and Semantics v. 9, Pragmatics*. P. Cole (ed.). Academic Press, New York. 221-243.

Lascarides, A. and N. Asher 1993. "Temporal Interpretation, Discourse Relations and Commonsense Entailment." *Linguistics and Philosophy* 16.437-493.

Linde, C. 1979. “Focus of Attention and the Choice of Pronouns in Discourse.” *Syntax and Semantics* 12, T. Givon (ed.). Academic Press.

Navarretta, C. 2000. “Abstract Anaphora Resolution in Danish.” *Proc. of the 1st SIGdial Workshop on Discourse and Dialogue*, L. Dybkjaer, K. Hasida and D. Traum (eds.). 56–65. ACL.

Oberlander, J., M. O’Donnell, A. Knott and C. Mellish 1998. “Conversation in the Museum: Experiments in Dynamic Hypermedia with the Intelligent Labeling Explorer.” *New Review of Hypermedia and Multimedia* 4.11-32.

Passonneau, R. and D. Litman 1993. “Feasibility of Automated Discourse Segmentation.” *Proc. of 31st Annual Meeting of the ACL*.

---------- 1993. “Getting and Keeping the Center of Attention.” *Challenges in Natural Language Processing*, M. Bates and R. Weischedel (eds.). Cambridge: Cambridge University Press, Ch. 7, 179-227.

Poesio, M. 1993. “A Situation-theoretic Formalization of Definite Description Interpretation in Plan Elaboration Dialogues.” *Situation Theory and its Applications*, vol. 3, P. Aczel, D. Israel, Y. Katagiri and S. Peters (eds.). Stanford, CA: CSLI. Ch. 12, 339-374.

---------- 1994. Focusing in the Interpretation of Definite Descriptions. *Proc. of the Workshop on Focus in NLP*. P. Bosch and R. van der Sandt (eds.). Schloß Wolfsbrunnen, June. Available as Working Paper 6, IBM Deutschland.

---------- 2000. “Annotating a Corpus to Develop and Evaluate Discourse Entity Realization Algorithms: Issues and Preliminary Results.” *Proc. of the 2nd LREC*, 211-218, Athens.

---------- and U. Reyle 2001. “Underspecification in Anaphoric Reference.” *Proc. of the 4th Int. Workshop on Computational Semantics*, E. Thijsse, H. Bunt and I. van der Sluis (eds.), 286-300, Tilburg.

----------, R. Stevenson, B. Di Eugenio and J. Hitzeman 2004. “Centering: A Parametric Theory and its Instantiations.” *Computational Linguistics* 30(3).

---------- and D. Traum 1997. “Conversational Actions and Discourse Situations.” *Computational Intelligence* 13.309-347.

Prince, E. 1992. “The ZPG Letter: Subjects, Definiteness, and Information Status.” *Discourse Description: Diverse Analyses of a Fund-raising Text.* S. Thompson and W. Mann (eds.). Amsterdam: John Benjamins. 295-325.

Roberts, C. 1999. “Demonstratives as Definites.” *Proc. of the ESSLLI Workshop on Deixis, Demonstratives and Deictic Belief in Multimedia Contexts*. E. Andre, M. Poesio and H. Rieser (eds.), 85-114.

Sanford, A. and S. Garrod 1981. *Understanding Written Language*. Chichester: Wiley.

Scott, D., R. Power and R. Evans 1998. “Generation as a Solution to its own Problem.” *Proc. of the 9th Int. Workshop on Natural Language Generation*, Niagara-on-the-Lake, CA.

Strube, M. and U. Hahn 1999. “Functional Centering: Grounding Referential Coherence in Information Structure.” *Computational Linguistics* 25.309-344.

Walker, M. A. 1998. “Centering, Anaphora Resolution, and Discourse Structure.” *Centering Theory in Discourse.* M. Walker, A. Joshi and E. Prince (eds.). 19.401-435.

----------, A. Joshi, and E. Prince (eds.) 1998. *Centering Theory in Discourse*. Oxford: Clarendon Press.

Webber, B. 1991. “Structure and Ostension in the Interpretation of Discourse Deixis.” *Language and Cognitive Processes* 6.107-135.

Zeevat, H. 1999. “Demonstratives in Discourse.” *Journal of Semantics* 16.279-313.

Index of Subjects and Terms

CURRENT ISSUES IN LINGUISTIC THEORY

E. F. K. Koerner, Editor
Zentrum für Allgemeine Sprachwissenschaft, Typologie
und Universalienforschung, Berlin
efk.koerner@rz.hu-berlin.de

Current Issues in Linguistic Theory (CILT) is a theory-oriented series which welcomes contributions from scholars who have significant proposals to make towards the advancement of our understanding of language, its structure, functioning and development. CILT has been established in order to provide a forum for the presentation and discussion of linguistic opinions of scholars who do not necessarily accept the prevailing mode of thought in linguistic science. It offers an outlet for meaningful contributions to the current linguistic debate, and furnishes the diversity of opinion which a healthy discipline must have. A complete list of titles in this series can be found on the publishers website, **www.benjamins.com**

222 **HERSCHENSOHN, Julia, Enrique MALLÉN and Karen ZAGONA (eds.):** Features and Interfaces in Romance. Essays in honor of Heles Contreras. 2001. xiv, 302 pp.
223 **FANEGO, Teresa, Javier PÉREZ-GUERRA and María José LÓPEZ-COUSO (eds.):** English Historical Syntax and Morphology. Selected papers from 11 ICEHL, Santiago de Compostela, 7–11 September 2000. Volume 1. 2002. x, 306 pp.
224 **FANEGO, Teresa, Belén MÉNDEZ-NAYA and Elena SEOANE (eds.):** Sounds, Words, Texts and Change. Selected papers from 11 ICEHL, Santiago de Compostela, 7–11 September 2000. Volume 2. 2002. x, 310 pp.
225 **SHAHIN, Kimary N.:** Postvelar Harmony. 2003. viii, 344 pp.
226 **LEVIN, Saul:** Semitic and Indo-European. Volume II: Comparative morphology, syntax and phonetics. 2002. xviii, 592 pp.
227 **FAVA, Elisabetta (ed.):** Clinical Linguistics. Theory and applications in speech pathology and therapy. 2002. xxiv, 353 pp.
228 **NEVIN, Bruce E. (ed.):** The Legacy of Zellig Harris. Language and information into the 21st century. Volume 1: Philosophy of science, syntax and semantics. 2002. xxxvi, 323 pp.
229 **NEVIN, Bruce E. and Stephen B. JOHNSON (eds.):** The Legacy of Zellig Harris. Language and information into the 21st century. Volume 2: Mathematics and computability of language. 2002. xx, 312 pp.
230 **PARKINSON, Dilworth B. and Elabbas BENMAMOUN (eds.):** Perspectives on Arabic Linguistics. Papers from the Annual Symposium on Arabic Linguistics. Volume XIII-XIV: Stanford, 1999 and Berkeley, California 2000. 2002. xiv, 250 pp.
231 **CRAVENS, Thomas D.:** Comparative Historical Dialectology. Italo-Romance clues to Ibero-Romance sound change. 2002. xii, 163 pp.
232 **BEYSSADE, Claire, Reineke BOK-BENNEMA, Frank DRIJKONINGEN and Paola MONACHESI (eds.):** Romance Languages and Linguistic Theory 2000. Selected papers from 'Going Romance' 2000, Utrecht, 30 November–2 December. 2002. viii, 354 pp.
233 **WEIJER, Jeroen van de, Vincent J. van HEUVEN and Harry van der HULST (eds.):** The Phonological Spectrum. Volume I: Segmental structure. 2003. x, 308 pp.
234 **WEIJER, Jeroen van de, Vincent J. van HEUVEN and Harry van der HULST (eds.):** The Phonological Spectrum. Volume II: Suprasegmental structure. 2003. x, 264 pp.
235 **LINN, Andrew R. and Nicola McLELLAND (eds.):** Standardization. Studies from the Germanic languages. 2002. xii, 258 pp.
236 **SIMON-VANDENBERGEN, Anne-Marie, Miriam TAVERNIERS and Louise J. RAVELLI (eds.):** Grammatical Metaphor. Views from systemic functional linguistics. 2003. vi, 453 pp.
237 **BLAKE, Barry J. and Kate BURRIDGE (eds.):** Historical Linguistics 2001. Selected papers from the 15th International Conference on Historical Linguistics, Melbourne, 13–17 August 2001. Editorial Assistant: Jo Taylor. 2003. x, 444 pp.
238 **NÚÑEZ-CEDEÑO, Rafael, Luis LÓPEZ and Richard CAMERON (eds.):** A Romance Perspective on Language Knowledge and Use. Selected papers from the 31st Linguistic Symposium on Romance Languages (LSRL), Chicago, 19–22 April 2001. 2003. xvi, 386 pp.
239 **ANDERSEN, Henning (ed.):** Language Contacts in Prehistory. Studies in Stratigraphy. Papers from the Workshop on Linguistic Stratigraphy and Prehistory at the Fifteenth International Conference on Historical Linguistics, Melbourne, 17 August 2001. 2003. viii, 292 pp.