Vocabulary Studies in First and Second Language Acquisition

The Interface Between Theory and Application

Edited by
Brian Richards, H. Michael Daller,
David D. Malvern, Paul Meara,
James Milton, Jeanine Treffers-Daller

Vocabulary Studies in First and Second Language Acquisition

This page intentionally left blank

Vocabulary Studies in First and Second Language Acquisition

The Interface Between Theory and Application

Edited by

Brian Richards
University of Reading

Michael H. Daller
University of the West of England, Bristol

David D. Malvern
University of Reading

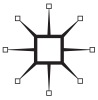
Paul Meara
Swansea University

James Milton
Swansea University

and

Jeanine Treffers-Daller
University of the West of England, Bristol

palgrave
macmillan



First published 2009 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries.

ISBN-13: 978–0–230–20668–7 hardback
ISBN-10: 0–230–20668–9 hardback

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.
A catalog record for this book is available from the Library of Congress.

10 9 8 7 6 5 4 3 2 1
18 17 16 15 14 13 12 11 10 09

Printed and bound in Great Britain by
CPI Antony Rowe, Chippenham and Eastbourne

Contents

List of Figures

List of Tables

Preface

Paul Meara
Swansea University, UK

Not so very long ago, it would have been completely unthinkable for any serious applied linguist to say that they were doing research on vocabulary acquisition. When I did my training, back in the 1960s, applied linguistics was acutely self-conscious, and very sensitive about being the application of *linguistic* theory, and since linguistic theory was pretty dismissive of vocabulary, hardly anyone thought that this area was one worth looking at. Structural linguists, Hockett for example, and their generative descendants, tended to regard vocabulary as little more than an inconvenience, and proper applied linguists saw vocabulary as a distraction from the real business of language learning, which was the mastery of grammar. Even Canale and Swain's (1980) seminal paper on communicative competence, which informed and guided much of the research in applied linguistics at the end of the twentieth century, relegated vocabulary to a minor role within grammatical competence. This book is an indication of how much things have changed since those early days.

With hindsight, we can probably date this change to the publication of Paul Nation's *Teaching and Learning Vocabulary* in 1990 – though a manuscript version had been circulating for some years before that (Nation, 1983). This book reintroduced applied linguists to a long tradition of research on the role of vocabulary in language teaching. Some of this work was tacitly acknowledged in the UK and Canada. For example, Palmer's work on vocabulary lists in the 1920s and 1930s had recently been highlighted by Howatt's (1984) account of Palmer's work at the Japanese Ministry of Education (Mombusho), and West's (1953) articulation of this work played a large role in the development of dictionaries aimed at EFL learners. But Nation's comprehensive analysis of the role of vocabulary in L2 teaching and learning went much further than this: it brought back into play a whole series of studies carried out by education specialists which applied linguists were largely ignorant of. It made us realize that learning a vocabulary is much more than the acquisition of a list of unorganized words, and that there were many questions to be asked about words and their acquisition. How big is a typical vocabulary? How quickly does a vocabulary grow? How much variation is there in individual learners? How much vocabulary do you need to perform tasks like a native speaker? What sort of vocabulary do specific tasks require? What is the relationship between active and passive vocabulary? How does L2 vocabulary

use differ from L1 vocabulary use? Can we identify thresholds in vocabulary learning? In short, Nation's book raised questions about the psychology of words which applied linguists were ill-equipped to answer, and created a need for new testing tools and new research methodologies.

Implicit in all of this was a massive research agenda, an agenda which was taken up vigorously by young scholars, particularly in Japan. And over the years, a number of serious volumes appeared all of which have contributed in their own way to the massive shift in priorities which Nation's work kick-started. The VARGA bibliography records only 20 vocabulary-related articles for 1975 – not much more than the number of papers contained in this volume. Thirty years on, the VARGA list for 2005 contains almost 20 times that number.

My own professional career has more or less coincided with this shift. What has particularly impressed me over this time is the huge increase in technical sophistication that has accompanied the growth in research volume. The development of standardized testing techniques and standardized research tools with well-understood measurement characteristics seems to be a huge improvement on some of the earlier work. The development of mathematical models which provide an underpinning for some of our theories about lexical development and vocabulary use also seems to be a good thing. We still have a long way to go in this area, but it is easy to see how some of the claims we might want to make about vocabulary acquisition are now testable in ways which would have been unthinkable only a few years ago. It is also noticeable that the increasing sophistication of research on vocabulary acquisition is making this area one that is more attractive for psycholinguists, neurolinguists, language testers, computational linguists and other specialists to work in. This too can only be a good thing for everybody, since it makes us question our assumptions more closely, and introduces us to new sets of research tools that allow us to approach old questions in new and exciting ways.

This volume presents recent original research on vocabulary that explores common themes and current issues in both first and second language over a wide range of ages and stages. A key feature is that, in every case, the themes and issues relating to vocabulary have implications for educational practice and policy. These include preparation for the academic language of school; developing foundations for literacy; teaching and learning first and second language vocabulary in the classroom; language dominance and vocabulary knowledge; the relationship between L2 lexical learning and the acquisition of morphosyntax; the impact of studying abroad; predicting academic success and failure; and language assessment in educational research and summative examinations. A variety of contexts are included. Beginning in the home with an analysis of first language parental language input, the chapters move through the early pre-school years into mainstream schools, bilingual classrooms and bilingual adult communities and

into foreign language teaching and learning and academic success at school and higher education.

Thirty years ago, it would have been impossible to assemble a book of this sort. I hope that the early pioneers of vocabulary research will be proud of what their dogged spadework has produced on what must have seemed like very stony ground at the time. And I look forward to seeing the sort of books we can produce in another 30 years' time.

Paul Meara
October 2008

Acknowledgements

We are indebted to the Economic and Social Research Council for funding the seminar series 'Models and measures of vocabulary acquisition, knowledge and use: the interface between theory and applications' through a grant awarded to the editors of this volume involving the University of the West of England, Bristol, the University of Reading and Swansea University (RES-451-254185). The resulting collaboration between the recipients and other contributors to the seminar series made the production of this collection possible.

Notes on the Contributors

Thomaï Alexiou is a Lecturer in the Department of Theoretical and Applied Linguistics, School of English at Aristotle University, Greece. She completed a PhD in Applied Linguistics at Swansea University in 2005 and has been teaching English for more than ten years in Greece and Britain. She has published articles in several European countries. Her research interests concern the methodology of teaching languages and she has special expertise in cognitive development, aptitude and language learning for very young learners. She has also written books in English for children.

Huili Chen is currently working as a Content Development Manager at Elsevier, the world's leading publisher of science and health information. In 2007 she completed a Master's dissertation at the University of Oxford on the different processes that advanced Chinese learners of English apply to learning low-frequency English words which are rendered by different translation types in Chinese.

Michael H. Daller is Senior Lecturer for Linguistics and English Language at the University of the West of England, Bristol. He received his PhD in Linguistics from Utrecht University in The Netherlands in 1996 with a dissertation entitled 'Migration and Multilingualism'. He has edited and authored books, and published articles on migration and multilingualism, assessing and modelling vocabulary knowledge and on the measurement of bilingual language proficiency.

Annabelle David is Postdoctoral Research Fellow at Newcastle University, UK. Her research focuses on second and bilingual language acquisition, especially of vocabulary. She also has an interest in the use of digital technologies to facilitate L2 research. Her work has appeared in the *Journal of French Language Studies*, *Language Learning Journal* and the *International Journal of Bilingual Education and Bilingualism*.

David K. Dickinson is a Professor in the Department of Teaching and Learning at Vanderbilt University. He has long been interested in the relationship between oral language and literacy development. Currently he is engaged in intervention research designed to help teachers who work with low-income children in centre-based programmes better support children's language and early literacy development.

Viv Edwards is Professor of Language in Education at the University of Reading where she is also Director of the National Centre for Language and

Literacy. She is editor of the international journal, *Language and Education*, and has researched and published widely in the area of learning and teaching in multilingual classrooms. Her publications include *Learning to be Literate: Multilingual Perspectives* (2009) and *Multilingualism in the English-Speaking World* (2004).

Tanya R. Flushman is a doctoral student at Vanderbilt University's Peabody College in the Department of Teaching and Learning. Prior to her doctoral studies, she taught in California and Mexico and obtained a Master's Degree in Curriculum and Instruction. Her areas of interest include early childhood literacy and language development, particularly of second language learners.

Pauline Foster is Reader in Applied Linguistics at St Mary's University College. She holds an MA from the University of Surrey and a PhD from King's College London. Her research publications are on second language acquisition, task-based language learning, classroom interaction and formulaic language.

Jill B. Freiberg is currently working on her doctorate at Vanderbilt University. Her focus is on development, learning and diversity, with specific research interests in language and early literacy development. She previously taught first and second grade in urban elementary schools, and trained pre-service Teach for America corps members in elementary literacy methods.

Tao Guo obtained a doctorate from the Department of Education at the University of Oxford on patterns of teacher code-switching in China and particularly on students' strategic responses to teacher code-switching. He is an experienced teacher of English as a Foreign Language and is currently teaching at Shanghai International Studies University.

Lotte Henrichs is currently working on her doctoral dissertation at the University of Amsterdam. Her research is on the acquisition of the academic register by young children, with a special focus on co-construction by parents and teachers. She has published on preschoolers' academic language proficiency in Dutch journals.

Ernesto Macaro is Professor of Applied Linguistics at the University of Oxford where he teaches on the Master's in Applied Linguistics and Second Language Acquisition and on the postgraduate language teacher education programme (PGCE). He has carried out research and published widely on the topics of code-switching in second language classrooms and on language learner strategies.

David D. Malvern is Professor of Education, the University of Reading, UK. He has been Research Officer at the Royal Society, Visiting Professor at the

Department of Educational Psychology, McGill University, and a European Union and British Council Consultant.

Paul Meara began his research career studying the psycholinguistics of syntax, but soon realized that vocabulary acquisition was a lot more interesting. He now works in Swansea University, where he runs a group of part-time PhD students, all working on a range of vocabulary-related research topics.

James Milton is subject leader in Applied Linguistics at Swansea University. His research focuses on vocabulary acquisition and measurement in foreign languages. Recently published works include *Modelling and Assessing Vocabulary Acquisition* (2007) and *Measuring Second Language Vocabulary* (2009). A current interest is finding a good basis to fitting vocabulary size measurements within the Common European Framework of Reference for Languages.

Florence Myles is Director of the Centre for Research in Linguistics and Language Sciences at Newcastle University, and Professor of French Linguistics in the School of Modern Languages. She is co-author of the widely acclaimed book *Second Language Learning Theories*, and has published on various aspects of second language acquisition of French, especially in the area of morpho-syntax. She is co-founder with Rosamond Mitchell of the French Learner Language Oral Corpora (FLLOC), a large database of oral language from classroom learners of French: www.flloc.soton.ac.uk.

Brian Richards is Emeritus Professor of Education at the University of Reading. A former teacher of German, French and English as a foreign language, he completed a doctorate on individual differences in first language acquisition at the University of Bristol in 1987. He has published widely in the fields of first and second language and language assessment and together with David Malvern is co-author of *Lexical Diversity and Language Development* published by Palgrave Macmillan in 2004.

Vivienne Rogers is currently working on her doctorate at Newcastle University. She previously completed her MA in language acquisition at Durham and her undergraduate degree at the University of Aberdeen. Her research is in second language acquisition of French syntax, particularly in respect of negation, adverbs and clitic pronouns.

Sarah Rule is a Research Fellow associated with the University of Southampton. She holds a PhD in the Second Language Acquisition of French from the University of Southampton, and has published articles on French second language acquisition focusing on morphology and syntax.

Rob Schoonen is Associate Professor of Second Language Acquisition at the University of Amsterdam where he teaches general linguistics,

psycholinguistics and research methodology. His main research interests are second and foreign language acquisition and proficiency, and language testing. He has (co-) authored articles in journals such as *Applied Linguistics, Journal of Educational Psychology, Language and Education, Language Learning* and *Language Testing*.

Alison Sealey is Senior Lecturer in Modern English Language at the University of Birmingham. She has published widely on policy and practice in the English curriculum, and also writes on the interface between social theory and applied linguistics. She is the author of several books, including *Childly Language: Children, Language and the Social World*, *Learning about Language: Issues for Primary Teachers*, and, with Bob Carter, *Applied Linguistics as Social Science*.

Peter Skehan works at the Chinese University of Hong Kong. His research interests include second language acquisition, task-based instruction, and language aptitude. He is the author of *A Cognitive Approach to Language Learning*, and co-edited *Researching Pedagogic Tasks* with Martin Bygate and Merrill Swain.

Lili Tian is a doctoral student in the Department of Education at the University of Oxford. She is conducting research into the effect of teacher code-switching on vocabulary acquisition during classroom interaction. She is an experienced teacher of English as a Foreign Language and has taught at Yantai University in China.

Jeanine Treffers-Daller is Professor of Linguistics at the University of the West of England, Bristol. She holds a PhD in Linguistics from the University of Amsterdam (1991), and has published widely in the field of bilingualism and language contact, in particular on code-switching and borrowing in French/Dutch contact in Brussels. Currently her research focuses on the role of transfer in language change and second language acquisition, as well as on vocabulary richness in bilinguals and second language learners. She is a member of the Editorial Board of *Bilingualism: Language and Cognition* and of the Impact series published by Benjamins.

Huijuan Xue holds a PhD in Business and Management and works as International Development Coordinator at the University of the West of England, Bristol. Her areas of expertise are English teaching, learning and testing, and international education. She has a wide range of publications including over 20 journal articles and co-edited books.

This page intentionally left blank

1

Lexical Features of Parental Academic Language Input: the Effect on Vocabulary Growth in Monolingual Dutch Children

Lotte Henrichs and Rob Schoonen
University of Amsterdam, Netherlands

Introduction

The importance of a rich vocabulary for successful literacy development of children has been well established over the past decades. It has been shown in several studies that children's vocabulary knowledge is of great value in predicting eventual school success (Biemiller and Boote, 2003; Bornstein, 1998; Bornstein, Haynes and Painter, 1998; Cunningham and Stanovich, 1997; Tymms, Merrell and Henderson, 1997). Most likely, the relationship between vocabulary and school success is mediated by children's reading proficiency. Several studies have shown that inadequate vocabulary knowledge hinders reading comprehension from the very beginning of literacy development in elementary school (cf. Biemiller and Boote, 2003). Since reading is a major skill in every school subject, insufficient reading comprehension skills in turn will cause problems in numerous fields of the elementary school curriculum.

During the language acquisition process large numbers of words have to be learned (cf. Clark, 2003). There are considerable individual differences among children regarding their vocabulary size and the speed with which they acquire new words. Biemiller (2006a; Biemiller and Boote, 2003) reports a difference of 4000 root words between the lowest and the highest quartile of second grade children in his studies. Because of this large variability, estimates of the average vocabulary size are hard to give, and tend to differ over studies in various countries. For example, Biemiller (2006a) found that by the end of Grade 2 (six years old), normally developing English-speaking children have acquired around 6000 root word meanings. According to his findings children subsequently acquire approximately 1000 words a year. Vermeer (2001) reports that Dutch children have acquired this number of 6000 root word meanings no earlier than the age of eight, from which point on they acquire 3000 additional words yearly. The fact that researchers differ

in their findings regarding vocabulary size is probably due to differences in approaches (data collection and analyses, cf. Richards and Malvern, 2007), besides the aforementioned large variability.

Various explanations for the large variability in vocabulary size among children have been studied over the past decades. Environmental factors such as socio-economic and cultural factors and factors relating to the home language environment have been found to influence vocabulary development (Hart and Risley, 1995; Heath, 1983). When considering the influence of socio-economic status (SES) and cultural background of the parents on children's vocabulary development, the question remains *how* these factors affect vocabulary growth. It is likely that SES is associated with parents' talk to their children, which in turn affects vocabulary development. Indeed, it was found by Hoff (2003) that the effect of SES on children's vocabulary is fully mediated by language input. In a study that investigated *why* parents of different SES backgrounds tend to differ in the way they communicate with their children, Rowe (2008) recently found that knowledge of child development and parenting beliefs in turn mediate the effect of SES on language input.

Hoff and Naigles (2002) found that higher levels of quantity, lexical richness, and syntactic complexity of the input that parents provide to their children positively influence two-year-olds' productive vocabulary. It has been consistently shown that frequency of the input provided to young children has a substantial impact on vocabulary growth (Huttenlocher, Haight, Bryk, Seltzer and Lyons, 1991; Pan, Rowe, Singer and Snow, 2005; Ravid and Tolchinksky, 2002; Vermeer, 2001). Moreover, Weizman and Snow (2001) found that the degree to which sophisticated lexical items are used predicts 50 per cent of the variance in vocabulary of second grade children, and is thus even more strongly related to vocabulary growth than sheer quantity. In their study, sophistication was defined as words not belonging to the Dale–Chall word list (Chall and Dale, 1995, reported in Weizman and Snow, 2001). In sum, among the various environmental factors that have been indicated to influence vocabulary development, language input might play a key role as a mediating factor.

Thus, regarding the effects of language input, it has been shown that not only quantity, but also quality of language input is an important predictor of children's vocabulary growth. This warrants a closer look at the quality of parental language input, and this is what we aim to do in the current study.[1] More specifically, we will look at features of a particular language register that is claimed to play an important role in achieving school success, that is, the *academic register*. These characteristic features can be found on all linguistic levels, but for the present study we limit ourselves to lexical features of the academic register. The aim of this chapter is to chart to what extent parents use the lexical characteristics of the academic register in interaction with their child in both semi-structured (school-like) and spontaneous interaction and

how this language use possibly changes with increasing age of the children. Additionally, we want to investigate whether a relationship can be found between parental use of the lexical features of the academic register and children's receptive vocabulary knowledge.

The academic register

The assumptions that underlie our focus on the academic register stem from the theoretical framework of Schleppegrell (2001, 2004). Schleppegrell states that academic language (or 'the language of schooling') is the expected language use in schools from a very young age. She argues that in formal settings such as school, children are expected to participate in conversations on cognitively complex topics. Moreover, they are expected to display knowledge while engaging in such conversations. Schleppegrell also argues that while displaying knowledge, children are expected to present themselves as 'authorities'. Where they cannot rely on shared 'physical' context they need to use linguistic tools to establish explicitness instead of for example pointing and using pronominal markers. Studies investigating language use in school show how oral expositions of knowledge that bear features of written language (e.g. explicitness, context-independency) tend to be highly valued by teachers (Collins and Michaels, 1986; Michaels, 1986; Schleppegrell, 2004).

The academic register is constituted by particular language features that are more prevalent in this register than they are in the informal register. The features cover the whole range of language use, that is lexical features (e.g. high lexical diversity and high lexical density), morphosyntactic features (high proportion of multi-clause sentences, varied mood choice), textual features (high demands on coherence, use of connectives, high level of abstraction) and socio-pragmatic features (many assertive speech acts, many open-ended questions). In this chapter, we will focus on *lexical features*.

Some children may have encountered academic register features fairly often in their home language environment, whereas others may not, and the degree to which children have become familiar with them might explain differences in vocabulary growth. It is expected that a higher degree of academic register features within parental language input during the preschool period and after the transition to school will have an impact on children's vocabulary knowledge. The reason for this hypothesized relationship is the well-established importance of both vocabulary knowledge and academic language skills for school success. In addition, the degree to which parents' language input resembles the academic register might change over time. Possibly, parents adjust their conversational style to the current developmental stage of their child.

First, we will elaborate on two focal features at the lexical level of the academic register: *lexical density* and *lexical diversity* of the input. Second, we

will briefly discuss the design and method of the study, and lastly, we will present the results and discuss possible implications of the findings.

Lexical density

The lexical density of a text refers to the relationship between the number of words with lexical properties as opposed to the number of words with grammatical properties. The measure is generally expressed as a percentage of the number of lexical words over all words in a text (Laufer and Nation, 1995; O'Loughlin, 1995). According to O'Loughlin (1995), the term was introduced by Ure (1971) who suggested that lexical density is a valid measure of the degree of 'literacy' versus 'orality' in a text, irrespective of the text being written or spoken. It is fair to expect that from the very first stages of schooling, a 'literate' conversational style is the expected mode. This expectation is warranted by the given that in informal interactional settings it is quite common to rely on a shared context by means of using deictic or other pronominal cues, whereas in academic settings this is usually not the expected style. A speaker in an academic setting needs to provide the listener with enough information about the specific topic of the discourse, and therefore has to be explicit. In order to convey information in an explicit manner, lexical words are needed. Therefore, a high lexical density is put forward as a feature of the academic register.

Thus, it is expected that in spoken interaction in educational settings, lexical words are used to refer to entities or to situations, where in informal interactions these might be referred to by (deictic) pronouns or other function words. Consider the following examples. Example (a) shows an utterance from our corpus, literally translated into English. The child comments on the researcher leaving the house while he and his mother are having lunch:

(a) *Ze gaat gewoon zonder jas aan!*
She is just leaving without her coat!

When this boy is to retell this event the next day in school, he needs to be more precise about what happened, where it happened, and to whom it happened. A (hypothetical) utterance like example (b) would be more appropriate in such a setting:

(b) *De vrouw die gisteren bij ons thuis was ging weg zonder jas aan.*
The woman who visited our house yesterday left without her coat.

According to the coding scheme used for this study, the lexical density of the Dutch sentence (a) is 0.33 (2 lexical words divided by 6) whereas the lexical density of the Dutch sentence (b) is 0.46 (6 lexical words divided by 13). Example (b) entails more (necessary) information than was needed in

(a) because mother and child shared the same context and therefore the same frame of reference. Because the information exchange in (b) is not about the here-and-now, it is not possible to rely on shared context and thus the context needs to be created linguistically. Consequently, sentence (b) consists of more content words, in order to add to the explicitness of the text. Thus, the number of content words as opposed to the total number of words is an indication of the amount of information that is packed into the sentence.

Lexical diversity

In formal (educational) settings speakers are expected to use a subject-specific and technical vocabulary, and to be explicit and clear about the message they want to convey. Also, in such settings, speakers tend to add information to the initial topic by using different words for the same object or event (Spycher, 2007). The degree to which new words are introduced and used in a text shows in the lexical diversity of the particular text. Lexical diversity can therefore be regarded as a measure for rich or varied language use. It is expected that language used in educational environments shows a higher degree of diversity than language used in informal interactional settings because educational settings require as much information as possible to be conveyed.

The measurement of lexical diversity is not as straightforward as one would expect, but recently a new measure has been introduced that seems to overcome some of the drawbacks of the traditional type-token ratio (TTR). This is the index *D* (Durán, Malvern, Richards and Chipere, 2004; Malvern and Richards, 2002; Malvern, Richards, Chipere and Durán, 2004; see also below in the method section).

In sum, both lexical density and lexical diversity are expected to be indicative of the academic register. We will use both measures to establish the relationship between parental language input and vocabulary development.

Research questions

In this chapter we will address the following research questions:

1. Do lexical density and lexical diversity in the language input of parents to their young children change over time?
2. Do lexical density and lexical diversity in the language output of children change over time?
3. Is there a relationship between lexical density and lexical diversity of parental language input on the one hand and the level of children's receptive vocabulary knowledge on the other?
4. Do lexical density and lexical diversity of parental input at one measurement time *predict* children's receptive vocabulary scores at a successive measurement time?

In order to be able to answer these questions, we analysed conversations of caregiver–child dyads. The data stem from an interdisciplinary study called 'The Development of Academic language in School and at Home' (DASH project). In the DASH project, the development of academic language is investigated in three- to six-year-old children. The current chapter reports an in-depth study of a subsample of the DASH project; in this in-depth study, the language input of parents is described in more linguistic detail than in the larger project, which allows us to address the above-mentioned questions.

Method

Design of the study

The in-depth study had a longitudinal design and concerns 25 monolingual Dutch children (11 girls and 14 boys) and their primary caregiver (24 mothers and 1 father). The families were of varying SES backgrounds with educational levels ranging from lower vocational education to academic education. The caregiver–child dyads were recorded at four points in time in a three-year period. The children were aged 3;2 at time 1 (T1), 3;10 at time 2 (T2), 4;2 at time 3 (T3), and 5;10 at the final recordings (T4). The main aim of the in-depth study is to obtain a detailed picture of the children's home language environment, focusing on lexical features of the academic language register. The dyads were video-recorded at each measurement point during four interactional settings, which we call the *interaction tasks*. The interaction tasks were designed to reflect different conversational settings. Three of the tasks were semi-structured; these were a *picture description task*, a *block-construction task* and a *book reading task*. The fourth task was the recording of a *mealtime conversation (lunch)* in order to obtain more spontaneous speech data. The tasks all involved verbal interaction. Accordingly, the parents were encouraged to involve their child in describing the picture and in talking about the book that was just read. The block construction concerned the joint building of a marble slide from a picture model. Despite the structured setting, the researcher stressed that the tasks should be performed just as the parent would do in a natural situation in which a joint activity would be performed. To enhance the naturalness of the situation, the researcher left the room after the instructions were completed and the camera was set up. The recording of these four interaction tasks was repeated at each time point and thus provided comparable longitudinal data.

Transcription and lemmatization

All interaction tasks were transcribed verbatim according to the CHAT format as described in the CHILDES manual (MacWhinney, 2000). For each interaction task separately, a maximum length (in minutes) to be transcribed was set. This length was based on the mean number of minutes the

parents needed to complete the interaction tasks. This mean length plus one additional minute was chosen as a maximum for transcription.

After transcription, each transcript was lemmatized, using the MOR program which is also provided for within the CHILDES system (MacWhinney, 2000). Following this procedure, the forms '*teach*', '*taught*' and '*teaches*', for instance, are treated as one type: the lemma '*teach*'. Homographs (such as *walk*) were marked as verbs or nouns in order to be counted as two different types. Also the transcripts were checked for inconsistencies in spelling in order to prevent a false increase in the number of types (cf. Richards and Malvern, 2007). Interjections such as *'hm'* or '*ooh*' were excluded from analyses. The final corpus that was analysed for the current study contained 54,110 utterances and 249,950 tokens, summed for mothers and children.

Measures of language input

The (lemmatized) transcripts of the interaction tasks form the basis of our input measures. As stated above, our input measures concern lexical density and lexical diversity of the parental language input.

Lexical density is defined as the proportion of 'content words' (i.e. nouns, verbs, adjectives and adverbs that express lexical meaning) over all words used in a task, and was computed manually (cf. Laufer and Nation, 1995; O'Loughlin, 1995). The MOR analysis as described above provides lexical categories. Verbs, nouns, adverbs and adjectives (i.e. content words or open-class words) could thus be distinguished and constituted the numerator of the measure for lexical density. In the case of adverbs, only adverbs that express lexical meaning were coded as content words. Thus, the adverb *'quickly'* would be considered a content word, whereas the adverb *'then'* would not. Proper nouns were considered content words, because as compared to pronouns, they explicitly refer to persons. The total number of words (tokens) in the transcript constituted the denominator.

Lexical diversity was indicated by the index *D*. *D* is based on mathematically modelling word probabilities and is less sensitive to text length than the type-token ratio and measures derived from it (Malvern and Richards, 2002; Malvern et al., 2004; McCarthy and Jarvis, 2007; see also Van Hout and Vermeer, 2007 for an alternative view). *D* is produced by the computer program VOCD, available within the CLAN programs (MacWhinney, 2000). VOCD uses a random token sampling method, which avoids the problem of obtaining a measure based on the clustering of the same vocabulary items at particular points in the transcript. It calculates the mean type-token ratio (TTR) from 100 random samples for 16 different sample sizes; the first set of 100 samples consist of 35 tokens, the second set of 100 samples consists of 36 tokens, until the last set of 100 samples which consists of 50 tokens. Accordingly, 16 mean TTR scores are computed. Subsequently, the program applies a curve-fitting procedure to find the best fit between the observed data of the 16 TTRs and a theoretical curve relating TTR and sample size

(Malvern et al., 2004). *D* is related to the decrease of TTR with increasing sample size.

Measures of children's language

Three measures of children's output are considered in this chapter. Similarly to caregiver data, children's lexical diversity and children's lexical density are investigated. Additionally, children's receptive vocabularies at the four points in time were studied. Children's receptive vocabulary knowledge was measured using two subtests of the computerized version of the Diagnostic Test for Bilingual Development (DTT) (Verhoeven, Narain, Extra, Konak and Zerrouk, 1995): *receptive vocabulary* and *concepts*. Despite the monolingual Dutch background of the current group of participants, a test suitable for bilingual assessment was required because of the cross-linguistic character of the DASH project as a whole. Obviously, for the current group, the test was administered in Dutch. Both subtests measure vocabulary, albeit on two different aspects. The receptive vocabulary subtest assesses knowledge of single words. The concepts subtest assesses not only lexical word knowledge but also knowledge of concepts such as 'all', 'in' or 'equal'. A trained research assistant administered the test to the children. The receptive vocabulary subtest consists of 60 test items (Cronbach's alpha = .85 at T2). On a computer screen, the child is provided with four pictures and subsequently the child is asked to point to the picture that depicts the target word. For example, the research assistant would say the Dutch word *'opstapelen'* (English translation *'to pile up'*) and the child should point to the appropriate picture. The vocabulary-concepts subtest consists of 65 test items (Cronbach's alpha = .79 at T2). Each item consists of three, four or five pictures and the research assistant produces the word or the sentence that refers to the target picture. For example, the research assistant would say: *'op één plaatje dragen alle kinderen een hoed'* (*'On one picture all children are wearing a hat'*). Again, the child is asked to point to the appropriate picture or pictures as some items require the selection of two pictures. For both subtests one point is allocated to each correct answer.

At T1, only half of the two vocabulary tests were administered because the test would take too long for three-year-old children. Therefore, only the odd items were administered. As a result, the composite variable of T1 initially consisted of half of the items compared to T2, T3 and T4. In order to maintain comparability, we estimated the scores that would have been obtained by the children when the even items were also administered. The missing data were estimated by means of a linear model, based on the scores of both halves of the vocabulary subtests that were administered at T2. The final measure of receptive vocabulary consisted of the sum of the scores on both subtests at each time point. The maximum score that could theoretically be obtained was the sum of the items of the receptive vocabulary subtest and the concepts subtest, which is 125.

Structured and spontaneous language data

It is expected that the features of academic language manifest themselves differently in structured tasks (picture task, block construction and book reading) as compared to spontaneous tasks (mealtime conversation). It is likely that the use of academic language features is comparable over the three structured tasks, whereas the manifestation of these features will be different in the spontaneous task. Therefore, the focal measures of language input (i.e. lexical density, lexical diversity) were studied separately for the structured tasks and the spontaneous speech. In order to do so, the scores for lexical density and lexical diversity in the three structured tasks were summed. Scale reliabilities (Cronbach's alpha) for the parental input measures were between .54 and .83. These alpha coefficients are not very high, but not uncommon for this kind of data and can be regarded as acceptable. Thus, scales were constructed for the language input measures at each measurement point. For the output measures, that is, the children's contribution to the interactions that constituted our data, Cronbach's alpha was acceptable at T1, T2 and T3 ($.54 < \alpha < .73$), but was unacceptably low at T4 ($.23 < \alpha < .25$). Children's values of lexical density and lexical diversity during the structured tasks varied too much to be able to use a summed score. As a consequence, for the children at T4, the descriptive statistics for the three structured tasks were kept separate (see Table 1.2) and were not used in the repeated measures analyses. The language data for the parents are presented in Table 1.1. Means and standard deviations of the summed scores are presented. The language data of the children are presented in Table 1.2.

The research questions that concern the change over time of lexical density and lexical diversity, that is to say, research questions 1 and 2, will be

Table 1.1 Means (standard deviations) of caregivers' language features for interaction tasks

	T1	*T2*[a]	*T3*	*T4*
Mean no. of words				
Structured tasks	510 (163)	493 (163)	512 (122)	397 (133)
Mealtime conversation	575 (219)	613 (253)	539 (191)	601 (234)
Mean no. of utterances				
Structured tasks	102 (27)	100 (23)	94 (20)	64(22)
Mealtime conversation	114 (39)	114 (37)	95 (29)	97 (32)
Lexical diversity				
Structured tasks	51.75 (6.95)	52.01 (6.88)	53.17 (6.60)	50.79 (6.95)
Mealtime conversation	68.7 (11.68)	69.68 (8.76)	70.82 (12.96)	69.07 (7.77)
Lexical density				
Structured tasks	.43 (.02)	.44 (.02)	.42 (.02)	.42 (.03)
Mealtime conversation	.46 (.04)	.44 (.05)	.44 (.05)	.43 (.03)

[a] At T2 two recordings failed due to technical problems, thus $N = 23$ for this measurement time.

Table 1.2 Means (standard deviations) of children's language features for interaction tasks

	T1	*T2*[a]	*T3*	*T4*
Mean no. of words				
Structured tasks	113 (48)	131 (49)	120 (50)	PT[b] 204 (88) BT 102 (57) BO 156 (63)
Mealtime conversation	184 (70)	302 (141)	212 (99)	258 (84)
Mean no. of utterances				
Structured tasks	47 (17)	48 (16)	42 (14)	PT[b] 51 (21) BT 32 (16) BO 35 (13)
Mealtime conversation	65 (25)	80 (35)	55 (25)	62 (19)
Lexical diversity				
Structured tasks	–[c]	34.76 (11.56)	39.62 (11.73)	PT[b] 40.29 (14.12) BT 29.06 (10.82) BO 43.86 (12.89)
Mealtime conversation	38.6 (13.25)	51.29 (9.94)	51.45 (12.14)	60.31 (14.49)
Lexical density				
Structured tasks	.42 (.07)	.42 (.06)	.39 (.06)	PT .43 (.05) BT .42 (.08) BO .46 (.04)
Mealtime conversation	.41 (.05)	.43 (.05)	.42 (.06)	.42 (.05)

[a] At T2 two recordings failed due to technical problems, thus $N = 23$ for this measurement time.
[b] PT = picture task, BT = block task, BO = book reading. At T4, too much variance occurred for children's output, thus no reliable scale score for the structured tasks could be computed (unlike T1, T2 and T3, where alpha coefficients were all $> .60$). Therefore individual values are reported.
[c] Due to small speech sample sizes (tokens < 50) too many missing values occurred for *D* on T1 to compute a valid mean *D*.

answered by means of an ANOVA with repeated measures. The research questions that concern relationships between these variables of language input and child vocabulary knowledge and growth, research questions 3 and 4, will be answered by means of correlational analyses.

Missing data

Notwithstanding that all 25 families continued to participate during the longitudinal study, some missing data were unavoidable due to causes explained below. In Table 1.3 the missing data for each measurement time are displayed. In the analyses that follow, the numbers in Table 1.3 explain the lower degrees of freedom in the ANOVAs answering the first two research questions and the lower *N* values in the correlation analyses answering the third and fourth research questions.

Table 1.3 Missing values at each measurement point

	T1	*T2*	*T3*	*T4*
Receptive vocabulary word knowledge/concepts	3/2	1/1	0/0	0/0
Lexical diversity input structured/mealtime	0/0	2/2	1/2	0/0
Lexical diversity output structured/mealtime	25/2	3/2	1/1	25/0
Lexical density input structured/mealtime	0/0	2/2	0/0	0/0
Lexical density output structured/mealtime	0/0	2/2	0/0	0/1

There are a number of different causes for the missing data. Regarding lexical diversity, an important cause is the fact that a minimum of 50 tokens is required to compute the index *D*. At T1, a lack of tokens occurred very frequently. In case of the block construction task, half of the children produced too few tokens to compute lexical diversity. As a result, it was decided to leave out children's lexical diversity data for T1 considering that it was not possible to compute a reliable composite measure. At T2, two recordings were lost because of technical problems with the camera. Furthermore, with the mealtime recordings families occasionally forgot to delay lunch until the appointment with the researcher. In these cases the mealtime conversation could not be recorded, and thus the focal variables for this interaction task were missing and could not be reported. At T1, in three cases it was not possible to administer the vocabulary test. The children were either too shy or too distracted to assess their vocabulary knowledge reliably. At T2 this happened only once, and at T3 and T4 all children completed the vocabulary test. The missing data are displayed in Table 1.3.

Results

Lexical density and lexical diversity of the input over time: the structured tasks

In Figures 1.1 and 1.2, the caregiver data for lexical density and lexical diversity are presented graphically. Looking at Figure 1.1, we see that lexical density of the caregivers during the structured tasks (the dashed line with triangles) changes over time. ANOVA with repeated measures confirms that this change is significant, revealing a main effect of time for lexical density of the input ($F(2.06, 22) = 4.54$, $p = .02$, $\eta^2 = .171$) (Because the assumption of equal variance at each measurement time was violated, the more conservative statistic of Greenhouse-Geisser, with adjusted degrees of freedom, was used in the repeated measures analysis.) Post hoc analyses revealed that this significant effect of time was due to a significant increase in lexical density from T1 to T2, and a significant decrease in lexical density from T2 to T3. As the nearly horizontal dashed line in Figure 1.2 leads us to expect, the lexical diversity of the input during the structured tasks, on the

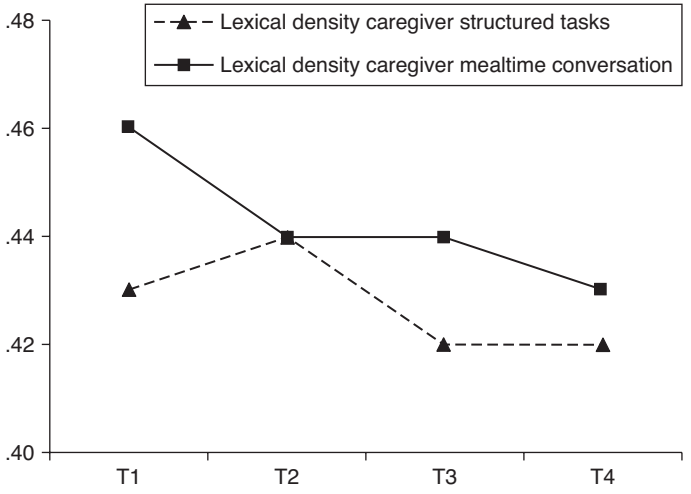


Figure 1.1 Lexical density of caregiver input during structured tasks and mealtime conversation

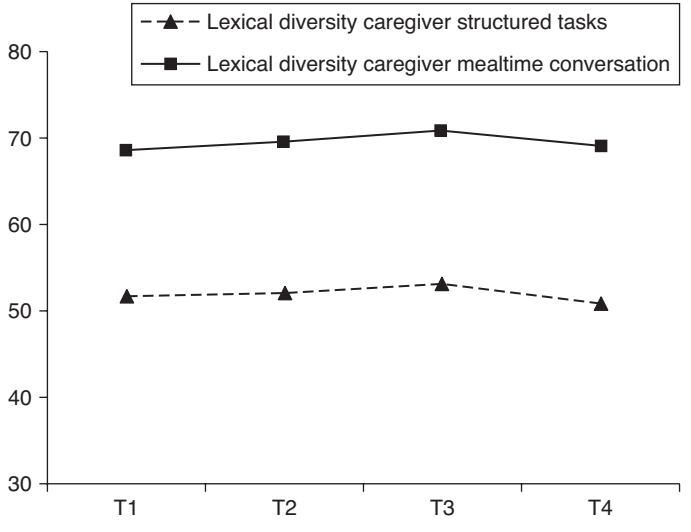


Figure 1.2 Lexical diversity of caregiver input (D_{input}) during structured tasks and mealtime conversation

other hand, does not increase over time. An ANOVA with repeated measures confirms this observation. The minor increase from T1 to T3 and the drop at T4 are not sufficient to produce a significant main effect ($F(3, 21) = 1.44$, $p = .24$).

It can thus be concluded that lexical diversity of the input during structured tasks is rather stable over time. Lexical density of the input during structured tasks, however, changes over time, and this change does not only involve growth.

Lexical density and lexical diversity of the input over time: the mealtime conversation

Even though the continuous line with squares in Figure 1.1 suggests change in lexical density of parental input during mealtime conversation, this change appears to be non-significant ($F(3, 20) = .83$, $p = .49$). Similarly, the nearly horizontal continuous line in Figure 1.2 suggests that lexical diversity of the input during the mealtime conversations, does not increase or decrease significantly over time. The lack of a main effect of time for lexical diversity of the input during the mealtime conversation shows in the ANOVA: $F(3, 20) = .43$, $p = .72$. Thus, lexical density and lexical diversity of the input during spontaneous speech interaction do not change over time in our sample.

Lexical diversity and lexical density in children's output over time: structured tasks

As was stated above in the method section, lexical diversity and lexical density for T4 were not comparable enough over tasks to be assembled into one score. Secondly, at T1, children too often produced less than 50 tokens, which is the minimum number that allows for *D* to be computed (see method section). As a result, for lexical diversity of the output of children during the structured tasks we can only report on the change between T2 and T3. For lexical density, we can report on the data between T1 and T3. As the descriptive statistics in Table 1.2 indicate, lexical density of children's output in the structured tasks hardly changes over time. Even though Figure 1.3 suggests change over time (the dashed line with triangles), no significant effect of time could be established regarding lexical density of children's output during the structured tasks ($F(2, 22) = 1, 40$, $p = .26$. The observed decrease from T2 to T3 is not statistically significant. The separate descriptive statistics for the three structured tasks in Table 1.2 suggest a slight increase in children's lexical density from T3 to T4, but because the scale scores at T4 are missing, this trend cannot be tested statistically.

Regarding children's lexical diversity during the structured tasks, an ANOVA revealed a main effect of time between T2 and T3 ($F(1, 18) = 8, 2$, $p < .05$, $\eta^2 = .148$; see also Figure 1.4, dashed line with triangles). The separate descriptive statistics for T4 do not suggest further growth from T3 to T4. Again, this impression could not be tested because of the large individual variance among tasks at the fourth measurement point.

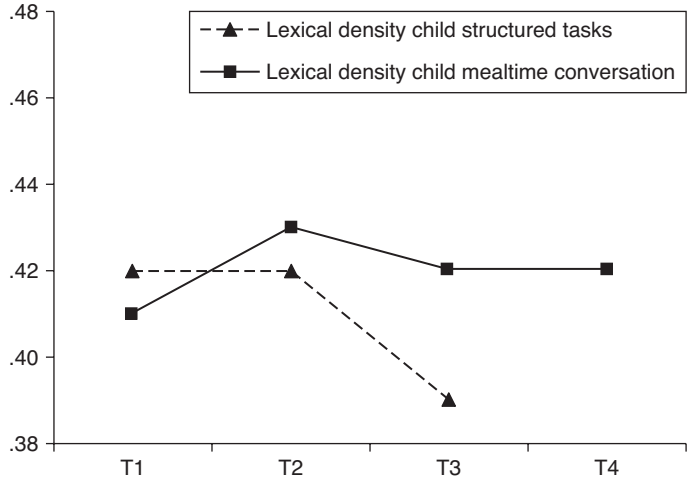


Figure 1.3 Lexical density of child output during structured tasks and mealtime conversation

Lexical diversity and lexical density in children's output over time: the mealtime conversation

As with the structured tasks, children's lexical density during the mealtime conversation did not change significantly over time ($F(3, 20) = .69$, $p = .57$, see also Figure 1.3, the straight line with squares). Again similar to the findings regarding the structured tasks, children's lexical diversity in the mealtime conversation (Figure 1.4, the straight line with squares) did increase significantly ($F(3,18) = 18.31$, $p < .01$, $\eta^2 = .504$). Post hoc tests showed that this significant increase was due to a significant increase in lexical diversity from T1 to T2 and the increase from T3 to T4. It can be seen from Table 1.2 that children's lexical diversity during mealtimes increased only slightly from T2 to T3.

Comparing lexical density and lexical diversity across the two settings

Note that both for caregivers and for children, lexical diversity of the mealtime conversation is considerably higher than the lexical diversity of the structured tasks (Figures 1.2 and 1.4). Paired *t*-tests (with Bonferroni correction for multiple comparisons) confirmed that this difference is statistically significant both for parents and for children at each measurement time. For parents: at T1, $t(24) = 7.78$, $p < .001$; at T2 $t(22) = 9.61$, $p < .001$; at T3 $t(22) = 7.98$, $p < .001$; and at T4 $t(23) = 10, 29$, $p < .001$, for children at T2 $t(21) = 5.02$, $p < .001$ and at T3 $t(20) = 5.69$, $p < .001$. The difference between lexical diversity during the structured tasks on the one hand and during the spontaneous task on the other, is relevant when interpreting the

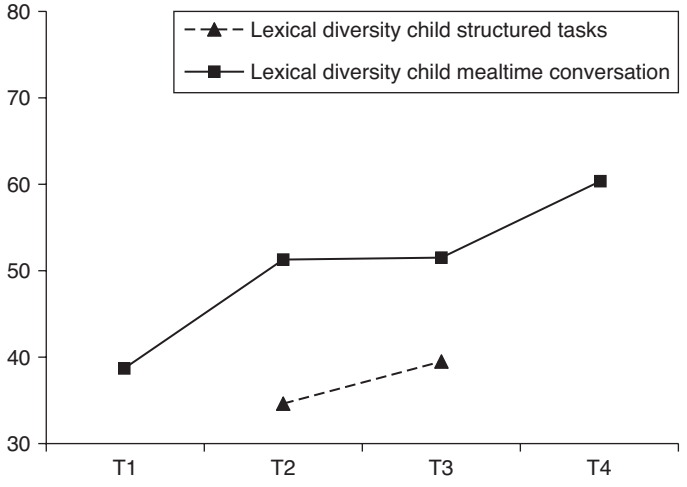


Figure 1.4 Lexical diversity of child output (D_{output}) during structured tasks and mealtime conversation

answers to the research questions regarding the relationship between lexical diversity of the input and children's vocabulary scores.

At T1 and T3, lexical density of parental input was also higher during the spontaneous setting of mealtime conversation as compared to lexical density during the structured tasks. For parents test statistics at T1 were $t(24) = 5.11$, $p < .001$ and at T3 $t(23) = 2.26$, $p < .05$. At T2 and T4 there were no differences regarding lexical density between the structured setting and the spontaneous setting. Children's lexical density did not differ between the structured tasks and the spontaneous setting.

Child receptive vocabulary over time

Table 1.4 shows that the vocabulary test scores of the children grow considerably over the three years that they participated in the study. The growth is statistically significant for each subsequent measurement time. In Figure 1.5 children's receptive vocabulary growth is displayed graphically. A strong main effect of time was found ($F(3, 20) = 143.36$, $p < .001$, $\eta^2 = 0.878$). Additional post hoc tests showed that this effect was significant for each time interval, and strongest for the change between T3 and T4, as can also be seen in Figure 1.5. However, it should be noted that the time interval between T3 and T4 is considerably longer than between T1 and T2 and between T2 and T3.

In Table 1.5, correlations for the vocabulary scores across time are displayed. They show that the scores on the vocabulary test at the subsequent measurement times are strongly and significantly correlated across time.

Table 1.4 Children's scores on receptive vocabulary test, $20 \leqslant N \leqslant 25$ (see Table 1.3)

Vocabulary score	*M (SD)*	*Standard error of mean*
T1	62.08 (15.96)	3.40
T2	77.96 (13.96)	2.91
T3	85.88 (13.24)	2.68
T4	109.40 (6.86)	1.38

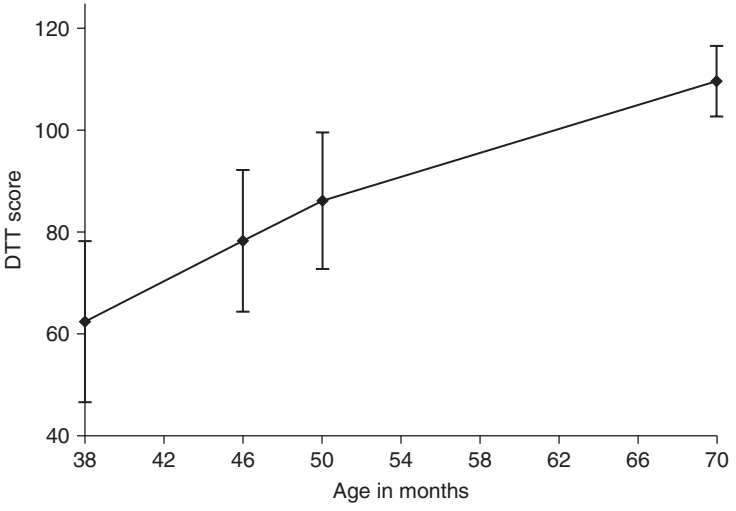


Figure 1.5 Receptive vocabulary growth between T1 (3;3) and T4 (5;10)

Table 1.5 Correlations across time for children's receptive vocabulary test scores

	T1	*T2*	*T3*	*T4*
T1	–	.61*	.72*	.72*
T2		–	.83*	.65*
T3			–	.87*
T4				–

*$p < .05$ level, two-tailed.

Lexical diversity and lexical density of the input and child vocabulary knowledge

After having looked at the development of the use of lexical features of the academic register, the last research question concerns the relationship between input on the one hand, and children's vocabulary scores on the other.

Table 1.6 Concurrent correlations (diagonal) between child receptive vocabulary score and D_{input}, and partial correlations (off-diagonal), structured tasks, $20 \leqslant N \leqslant 25$ (see Table 1.3)

Vocabulary	*D_{input} during structured tasks*			
	T1	*T2*	*T3*	*T4*
T1	**.41**			
T2	.32	**.68***		
T3	.04	−.03	**.47***	
T4	.25	.25	−.05	**.46***

Note: bold print indicates a concurrent correlation between lexical diversity of the input and child vocabulary scores. Partial correlations (controlling for preceding vocabulary score) are reported in normal print.
*$p < .05$, two-tailed.

Table 1.7 Concurrent correlations (diagonal) between child receptive vocabulary score and D_{input}, and partial correlations (off-diagonal), mealtime conversation, $20 \leqslant N \leqslant 25$ (see Table 1.3)

Vocabulary	*D_{input} mealtime conversation*			
	T1	*T2*	*T3*	*T4*
T1	**.06**			
T2	.44*	**.48***		
T3	−.08	.12	**.35**	
T4	−.08	−.04	−.06	**.37**

Note: bold print indicates a concurrent correlation between lexical diversity of the input and child vocabulary scores. Partial correlations (controlling for preceding vocabulary score) are reported in normal print.
*$p < 0.05$, two-tailed.

No relationship was found between children's vocabulary scores over time and the degree of lexical density in parental input. Correlations between lexical density during the structured tasks and vocabulary scores at each measurement point ranged between −.20 and .34. These weak correlations were non-significant without exception. Similar results were found for parental lexical density during the mealtime conversation. Correlations with children's receptive vocabulary scores at the subsequent measurement times were weak and non-significant, *r* ranging from −.42 to .23.

Regarding parental lexical diversity, by contrast, significant relationships with children's vocabulary scores could be established. In Table 1.6 concurrent correlations are displayed between lexical diversity of parental input during the structured tasks and children's vocabulary scores at the same measurement time (in bold print). In Table 1.7, these same concurrent

correlations are shown for the mealtime conversation. To study whether the lexical diversity at one point in time was also related to vocabulary scores at a later point in time, partial correlations are displayed in Tables 1.6 and 1.7. For example, in order to obtain a measure of the unique relationship between lexical diversity at T1 and child vocabulary score at T3, we controlled for the relationship between lexical diversity at T1 and vocabulary at T2. Since the subsequent vocabulary scores were strongly and significantly correlated over time (see Table 1.5), these correlations also need to be taken into consideration when seeking to predict vocabulary scores from the input measures.

Looking at the concurrent correlations for the structured tasks first, we see that the relationship between lexical diversity of the input and child vocabulary scores is significant at the $p < .05$ level at T2, T3 and T4 (Table 1.6). The partial correlations show moderate to low relationships, which are non-significant without exception. It does stand out, however, that the correlations are particularly low for the relationship with vocabulary at T3 which might be explained by the fact that the vocabulary assessment at T3 was shortly after T2. Because the time interval between T2 and T3 is so short, by controlling for the concurrent relationship between lexical diversity of the input and child vocabulary at T2, only a very small amount of variance remains to be explained. It is therefore likely that no unique relationship exists between input at T2 and vocabulary at T3.

Table 1.7 shows the results of an identical correlational analysis (concurrent and partial) for lexical diversity of parental input during the spontaneous setting of mealtime, and child vocabulary knowledge. Comparing these figures to the results for the structured tasks, it stands out that it is only at T2 that we see a significant concurrent correlation between lexical diversity of the input and child vocabulary knowledge. The concurrent correlations at T1, T3 and T4 are not significant. In contrast to the results for the structured tasks, where we saw no significant partial correlations, the results for the mealtime conversation in Table 1.7 do show a significant partial correlation: the relationship between lexical diversity of the input at T1 and child vocabulary knowledge at T2, controlling for vocabulary knowledge at T1. This means that there is a unique relationship between lexical diversity of the input during mealtime at T1 and vocabulary score at T2. Other partial correlations are very low and non-significant.

Discussion and conclusion

The main aim of this part of our study was to investigate whether a relationship can be established between the degree to which parents use academic register features and their children's vocabulary growth, as assessed by a standardized vocabulary test that was repeated four times. The extent to which parents use lexical features of the academic register in interaction

with their young children in both semi-structured and spontaneous situations was charted, and this linguistic behaviour was related to child vocabulary knowledge at four points in time. Lexical density and lexical diversity of parental language were chosen as the focal language input features because of their association with the academic register. It was expected that parents who provide their children with lexically diverse and lexically dense language input foster their child's vocabulary knowledge.

The first and the second research questions involved the change over time of lexical density and lexical diversity of both parents' and children's language use. Results indicated that lexical density in parental input was subject to change, but lexical diversity remained stable over the four measurement times in this study. The change in the degree of lexical density over time did not meet our expectations. Instead of an expected steady increase, lexical density in parental language increased and then decreased. We suggest two possible explanations for this unexpected decrease. First, parents possibly refrain from mere labelling as their child grows older, something that might explain higher lexical density at the first two measurements. As a consequence of less labelling, an increase in the amount of function words will arise. Second, it could be seen from Table 1.2 that parents produce fewer utterances from T3 onwards, as a result of which the relative contribution of the children increased. Possibly because parents let their child do more of the talking, lower lexical density of the input was the result. Feedback utterances such as 'well done', 'indeed', 'yes' and so forth were highly frequent in the transcripts at T3 and T4. Alternatively, habituation to the task when parents conducted it for the third time may have resulted in different behaviour as compared to the previous two measurement times.

Lexical density in the output of children did not change over time, neither in the structured tasks nor in the spontaneous task. It needs to be taken into consideration, though, that children of this age are still learning both open class and closed class words. As children become more talkative over the years, both the numerator and the denominator will change over time in a way that results in an unchanging lexical density score. Our choice for lexical density as a characteristic of academic language was based on previous research that concerned mainly written texts or L2 acquisition research. Possibly the transfer to L1 oral language is less straightforward or even more problematic than was initially expected.

Children's lexical diversity, however, did increase over time, both in the structured tasks and in the spontaneous task. Thus, children became more diverse in their word choice, an indication of an increasing productive vocabulary. This result relates to the increase we predicted and obtained in children's scores on the receptive vocabulary test. As expected, the children's vocabulary increased significantly in the course of almost three years.

In sum, our data show us that differences exist among parents regarding the extent to which they provide their children with lexically diverse language

but that the extent to which parents are lexically diverse is rather stable over time. Lexical density, however, is less stable over time and does not vary a lot among parents. In a previous study in which kindergarten teachers were involved, we found that teachers used more explicit language during book reading, that is to say, showed lexically denser language use than was the case for parents (Henrichs, 2007). Therefore, it had been expected that some parents more than others might show parallels with teachers' language input, especially on structured tasks. As stated above, however, there was no such variation among parents in our data.

The third and fourth research questions concerned the relationship between lexical diversity and lexical density of the input on the one hand and children's vocabulary knowledge and growth on the other. We found positive associations between lexical diversity of parental input and children's vocabulary. Moderately strong significant correlations were found between lexical diversity of the input during joint picture description, block building and book reading and children's scores on a vocabulary test, assessed four times over a period of three years. Few significant correlations, however, were found between parents' lexical diversity during the mealtime conversation and children's vocabulary knowledge. The relationship between lexical diversity during the structured tasks and children's vocabulary knowledge is interesting, for it *suggests* that lexical diversity in one context (structured settings) might be more beneficial than lexical diversity in another context (spontaneous setting of mealtime). However, we cannot assume such a causal relationship from concurrent correlations as parents are known to be responsive to their children's language level. A follow-up study in which the sophistication (or frequency) of words used in the conversation is taken into account could show whether we are dealing with different domains of vocabulary. That is, possibly the structured settings elicited the use of more low-frequency words, thus constructing a genre that is related to vocabulary acquisition (Weizman and Snow, 2001). With respect to this issue, it is also important to keep in mind that lexical diversity was significantly higher during the mealtime conversations as compared to the structured tasks (see Figure 1.2). This phenomenon is described in other studies as well, and can be explained by the variety of topics that occur in free conversation (cf. Van Hout and Vermeer, 2007). During the structured tasks, the materials at hand control the topic of conversation to a certain extent, simply because the materials represent a given topic. It is therefore an interesting finding that it was lexical diversity during those tasks that were designed to reflect school-like settings that showed the most, and strongest, relationships with child vocabulary knowledge. This finding suggests that it is not the mere diversity of words that is related to children's vocabulary development, but that the diversity of low-frequency words that may be responsible for the relationship with child vocabulary scores.

Lexical density did not show any relationship with children's vocabulary scores. It was hypothesized that parents who display higher lexical density and thus use many content words, would foster their child's vocabulary development. However, contrary to our expectations, an association between the proportion of content words and children's vocabulary knowledge could not be established in our data. Again, as was elaborated on above, this finding might be explained by the fact that parents did not differ sufficiently from each other in lexical density to identify an association with the dependent variable, that is to say, children's vocabulary knowledge.

Finally, we investigated whether growth in vocabulary could be associated with lexical diversity of parental input. In other words, we wanted to know whether we could predict child vocabulary scores at a given time by lexical diversity of the input at a preceding measurement time. This was done by looking at the partial correlations between lexical diversity of the input at a given measurement time and the subsequent vocabulary assessment, each time controlling for the preceding vocabulary score. It was found that only the partial correlation between lexical diversity of the input during the mealtime conversation at T1 and children's vocabulary scores at T2 was significant. When controlling for previous vocabulary knowledge, all other correlations became lower and were no longer significant at the 0.05 level. Thus, we had to conclude that for our data, vocabulary growth could not be predicted from the lexical diversity of the input, with the exception of the mealtime conversation at T1. An explanation for this lack of association might lie within a low level of variance within the vocabulary growth of the children. There clearly is variance in the individual vocabulary knowledge level of the children, but it might be the case that their growth rates are fairly equal. This interpretation is supported by the strong intercorrelations between the subsequent vocabulary scores at T1 to T4 (see Table 1.5).

In this chapter we wanted to focus on the lexical aspects of parents' conversational style when interacting with their children, and look into the relationship between this conversational style and children's vocabulary development. Our intention was to start out from the actual linguistic behaviour of both parents and children instead of socio-economic background, which is often the case. The relationship between levels of child vocabulary and parental SES background is well established (Hart and Risley, 1995; Hoff, 2003; Hoff-Ginsberg, 1991; Pan, Rowe, Spier and Tamis Lemonda, 2004; Snow, 1983). As was mentioned in the introduction, however, recent work has shown that parental SES is reflected in particular language features and parenting beliefs that in turn affect children's vocabulary knowledge. These studies thus speak of a mediating effect of language (Hoff, 2006; Rowe, 2008). The findings in the current study concerning the variety among parents regarding lexical diversity in conversations with their children and its relationship to vocabulary knowledge seem to support this recent development in research in which the mediating role of language is highlighted.

Limitations of the study and suggestions for further research

In this study we compared the conversation during one spontaneous task (mealtime) with the conversations during three structured tasks. The data on the spontaneous task are probably less reliable, which may have attenuated its relationship with vocabulary knowledge. In future research, if one wants to include the spontaneous setting, more than one context should be included.

Second, the finding that lexical diversity in school-like settings shows a relationship with child vocabulary knowledge calls for a qualitative dimension to the index *D* (cf. Daller, Van Hout and Treffers-Daller, 2003). Apparently, we can speak of different *kinds* of lexical diversity, and one particular kind might be more important for children's vocabulary development than another kind. Furthermore, exploration of the vocabulary used in the language input might reveal that diversity of low-frequency words is more beneficial than diversity in high-frequency words, or that diversity within one topic is more 'effective' than diversity due to multiple topic changes.

Third and finally, the measure of lexical density does not seem salient enough to be a discriminating feature for the spoken academic language register. This finding has important implications for future research on academic language use. One of the questions is to what extent lexical density can vary in normal conversation. An increase in content words will usually cause an increase in function words, such as articles and prepositions.

The findings of the current study underline the importance of children's informal learning experiences at home as can be achieved by learning in semi-structured settings. Such informal learning experiences are of importance to familiarize children with school-like settings, and thus make the transition to school as smooth as possible. Becoming familiarized with the academic language register that is common in school (Schleppegrell, 2001, 2004) is one of the aspects that contribute to such a smooth transition. This study showed that at least one of the lexical characteristics of the academic register, lexical diversity of the input, is related to child vocabulary knowledge – a well-established predictor of academic success itself. Conversations during school-like settings at home between parents and their preschoolers or kindergartners seem to provide the opportunity for parents to produce the kind of lexical diversity that is related to vocabulary knowledge of their children.

Note

1. The current study is part of a large-scale collaborative project called the 'The Development of Academic language at School and at Home' (DASH). The DASH project is conducted at the universities of Amsterdam, Tilburg and Utrecht. The project is funded by the Netherlands Organization for Scientific Research (NWO), project number 411-03-060.

2

Vocabulary, Reading and Classroom Supports for Language[1]

David K. Dickinson, Tanya R. Flushman and Jill B. Freiberg
Vanderbilt University's Peabody College

Successful acquisition of the ability to read with comprehension is essential for school success and for full participation in the mainstream technological society. The ability to read with comprehension draws on multiple intellectual skills, with vocabulary and other language abilities being of central importance (Hoover and Gough, 1990; Rapp, van den Broek, McMaster, Panayiota and Espin, 2007). Although vocabulary has long been recognized as important to reading success (Anderson and Freebody, 1981; National Reading Panel, 2000), instruction in the early grades in the United States traditionally has focused heavily on issues related to decoding (i.e. letter knowledge, phonics and associated phonemic awareness ability). Explicit, intentional instruction related to building vocabulary has tended to begin around fourth or fifth grade and in some cases has been delayed until high school. But mounting evidence suggests that by attending narrowly to 'basic skills' at the expense of vocabulary, later reading comprehension abilities suffer.

By failing to support vocabulary effectively, we overlook the most pressing educational needs of many children who are most at risk of later reading failure. Children from economically disadvantaged homes at age four are typically over a year behind their peers in receptive vocabulary (Dickinson, St. Pierre and Pettengill, 2004; Huttenlocher, 2002; Zill and Resnick, 2006), and enter school seriously behind their peers in vocabulary knowledge (Biemiller, 1999, 2006b; Dickinson et al., 2004; Huttenlocher, 2002; Zill and Resnick, 2006). Evidence of the negative effects of these two propositions is the limited progress that has been made in narrowing the achievement gap between children from more and less advantaged homes (National Center for Education Statistics, 2007a). In 2004 the reading comprehension scores of 13-year-old children on the National Assessment of Educational Progress (NAEP) conducted in the United States revealed a 29-point gap between those whose parents graduated from high school and those with parents who were college graduates – exactly the same gap found in 1999, one point more than was seen in 1994, and nine points wider than was detected

in 1980. Similar persistent gaps related to income predictably appear when data are broken down by race and ethnicity (Campbell, Hombo and Mazzeo, 2000).

In this chapter we begin by briefly reviewing evidence that oral language skills are pivotal to reading comprehension, and that early language skills have important and enduring effects on later language and reading skills. We are primarily interested in the support for language that can be provided by classrooms; therefore we discuss the effects of classroom experiences on vocabulary learning. As we review results from one study we also consider methodological issues of importance to researchers interested in describing language environments of classrooms. Next we describe data that make clear that classrooms often fail to provide optimal support for language development and we briefly discuss intervention efforts to address this problem. We then review language learning of children who are learning a new language in school and conclude by suggesting topics of most pressing importance for future research.

Language and reading comprehension

Cognitive process models support the contention that language competencies are key to reading comprehension (Hoover and Gough, 1990; Rapp et al., 2007; Tunmer and Hoover, 1992). For example, the Landscape Model of Reading (van den Broek, Young, Tzeng and Linderholm, 1999) describes comprehension as the result of ongoing efforts to construct and then revise an emerging mental model of text by using transient text cues, drawing on background knowledge and making inferences to create an integrated coherent representation of the text. Vocabulary and other language skills are not explicitly discussed, but are presumed because vocabulary is necessary to access background knowledge, and syntax is required for accurate interpretation of textual details. Consistent with this perspective, Leseman and van Tuijl (2006) argued '… the basic dimension of (school) literacy may not be the technics of recoding letters into sounds and blending them into words and sentences … but mastery of the academic language register and its associated specialized vocabulary and grammar' (p. 214).

While theories of comprehension accord a central role to language, the language-related competence most often correlated with reading skills is vocabulary. Consistent with this perspective is the theoretical account of the relationship between language and reading advanced by Ravid and Tolchinsky (2002), who view acquisition of literacy-related forms of language as being intimately related to skilled reading. In support of this theory Ravid (2006) demonstrated that in adolescence children progressively use more abstract nouns, a semantic type that is more commonly found in expository than narrative texts and in written than spoken narratives.

Additional evidence indicates that reading comprehension requires advanced language skill as well as the semantic knowledge and inferential ability that are associated with vocabulary (Cain, Oakhill and Bryant, 2004; McGill-Franzen and Allington, 1991; Nation, Adams, Bowyer Crane and Snowling, 1999; Nation, Clarke, Marshall and Durand, 2004; Oakhill, Hartt and Samols, 2005; Weekes, Hamilton, Oakhill and Holliday, 2008).

Evidence is mounting that early language abilities play a pivotal role in reading ability, with the evidence being clearest for the role of vocabulary. Scarborough (2001) found moderately strong associations between language and reading development, and longitudinal studies have found that language skills in the preschool and kindergarten years are correlated with literacy abilities in the primary grades (NICHD Early Child Care Research Network, 2005; Storch and Whitehurst, 2002; Whitehurst and Lonigan, 1998, 2001). One study of low-income children examined those whose reading level was below the 30th percentile at the end of first grade (Spira, Bracken and Fischel, 2005). The authors found that grade one performance only weakly predicted grade four reading, but when measures of kindergarten receptive and expressive vocabulary and emergent literacy were added, the predictive power of the model was substantially increased. These findings are consistent with analyses of growth trajectories of children who were followed from age three into middle school (Dickinson and Tabors, 2001). End-of-kindergarten vocabulary and word recognition skills, in combination with children's rate of growth on these measures, accounted for over two-thirds of the variance in fourth grade reading comprehension (Tabors, Porche and Ross, 2003). Noting the strong pattern of intercorrelations among vocabulary and other language skills, phonological awareness and print knowledge, Dickinson and colleagues (Dickinson, McCabe, Anastasopoulos, Peisner-Feinberg and Poe, 2003) hypothesized that language plays a powerful role in the organization of these systems.

Environmental factors affecting early literacy development

Children face longer odds of academic success if they are from low-income backgrounds whether they grow up in the United States (Bishop and Edmundson, 1987; Dickinson, 1987; Hart and Risley, 1995; Strickland, 2001; Tarullo and Zill, 2002; Whitehurst and Lonigan, 1998, 2001) or other industrialized societies (Elder, 2005; Leseman and van Tuijl, 2006; Magnuson and Waldfogel, 2005; McNaughton, 2006). Children from economically disadvantaged homes at age four typically enter school seriously behind their peers in vocabulary knowledge (Biemiller, 1999, 2006b; Dickinson et al., 2004; Huttenlocher, 2002; Zill and Resnick, 2006).

Despite facing longer odds, some children from disadvantaged backgrounds are successful in learning to read and write. To understand factors

during the years prior to formal schooling that enable some children from low-income homes to become successful readers, our Home-School Study of Language and Literacy Development (HSSLD) examined the preschool experiences of a group of 83 low-income children. We hypothesized that children's home and classroom language experiences during the preschool years would be predictive of reading comprehension in the later primary grades (Dickinson and Tabors, 2001). To study the role of classrooms, teachers were audiotaped and videotaped for a full day when children were four years old. Recordings were transcribed by classroom context (e.g. book reading, free play), with set amounts of time in each setting being sampled. Children were assessed using a battery of language and literacy measures in kindergarten, fourth grade and seventh grade. Our analyses included talk that occurred as teachers related to all children, not only the one or two target children from our study; thus our variables measured the overall language environment created as the teacher conversed with children.

We examined three different dimensions of the language environment of classrooms. Two variables examined the structure of conversations. Conversational balance was assessed by comparing the overall number of words teachers used relative to words used by children when they were engaged in conversations. A second variable described teacher efforts to help children extend their comments. We coded speech acts, noting conversational moves designed to help a child extend the current topic of conversation. Extending efforts during free play were divided by all coded free play teacher utterances.

We also examined the level of intellectual challenge in classrooms during group book reading. We coded the content of all the comments and divided it by the total number of utterances. The critical context variable was analytic exchanges – occasions when the meaning of words or the interpretation of a story were discussed.

A third dimension that we examined was exposure to sophisticated vocabulary. The variables we used merit detailed discussion because of the complex methodological issues involved in efforts to describe lexical richness. As Malvern, Richards, Chipere and Durán (2004) have documented, there are many challenges associated with measures designed to describe the diversity of language known and used by speakers and writers. The core problem is that there are complex associations between the number of words used (tokens) and the number of distinct word types. One approach we used to deal with this issue was to equate language samples by analysing language from similar amounts of time. This strategy still produces transcripts that vary in length and, given the association between length and diversity, it advantages teachers who talk more. Another strategy is to divide distinct types by total tokens. We also used this approach because the density of novel words children hear is a better predictor of vocabulary growth than is a simple count of word types (Hoff, 2003; Hoff and Naigles,

2002; Huttenlocher, Haight, Bryk, Seltzer and Lyons, 1991; Pan, Rowe, Singer and Snow, 2005).

Another decision we made related to deciding what word types to analyse. We focused on sophisticated words rather than all word types because, as children move past the initial stages of language acquisition, they learn less commonly used and more complex vocabulary (Anglin, 1993). Our list of sophisticated vocabulary was created using the updated Dale–Chall list of 3000 words known by fourth graders (Chall and Dale, 1995). This list was augmented to create a total set of nearly 8000 words that we used to filter all the words spoken in the classroom thereby identifying the 'sophisticated' vocabulary. Our measure of sophisticated vocabulary did not control for the fact that, even when a proportional variable is used, the variability in the number of word types still affects the resulting measure (Malvern et al., 2004). However, we also conducted analyses examining all tokens. Therefore, these variables described both the density of exposure to sophisticated vocabulary and the degree to which teachers were active in engaging children in conversations.

Our measures captured distinct dimensions of classrooms and only one significant correlation was found among our four language environment variables – teacher efforts to extend topics use was related to the ratio of teacher-to-child talk ($r = .40$, $p < .001$). By taking this multifaceted approach we helped counteract the methodological problems associated with measures of lexical sophistication (Malvern et al., 2004). As shown in Table 2.1, we found modest to moderately strong correlations between all four preschool language environment variables and measures of receptive vocabulary and reading comprehension in kindergarten and grade 4.

Prior regression analyses of end-of-kindergarten vocabulary and early literacy skills revealed that teacher–child interactions when children were in preschool added substantial variance over and above what was accounted for by home literacy support, demographic and child variables (parent

Table 2.1 Correlations between preschool language environment variables and kindergarten and grade four vocabulary and reading comprehension

	Kindergarten PPVT[a] *(N = 74)*	*Grade 4 PPVT (N = 57)*	*Grade 4 Comprehension (N = 57)*
Conversational balance[b]	.29*	.40**	.34**
Extending topics	.38**	.34**	.29*
Sophisticated vocabulary	.29~	.32*	.34*
Intellectual level	.39**	.31*	NS

[a] Peabody Picture Vocabulary Test.
[b] Stronger growth was associated with less teacher talk relative to child talk.
~$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

education, income, gender, mean length of utterance (MLU) at age three) (Dickinson and Tabors, 2001). Recently we re-examined these data to determine whether preschool classroom experiences also were associated with reading skills at the end of fourth grade (Dickinson and Porche, under review). Analyses again controlled for demographic and child variables and added a new and more potent home control, the parents' reported support for literacy during the preschool years (e.g. frequency of reading, use of the library). Our full model predicted 44 per cent of the variance in end-of-kindergarten receptive vocabulary, with 18 per cent of this variance uniquely associated with preschool classroom descriptions. Kindergarten vocabulary was strongly correlated with grade 4 vocabulary ($r = .77$) and reading comprehension ($r = .62$), and grade 4 vocabulary was highly related to grade 4 reading comprehension ($r = .71$). Our full regression model using preschool variables to predict to end-of-grade 4 failed to find an association between preschool classroom variables and fourth grade vocabulary, but note that preschool experiences did account for kindergarten vocabulary which, in turn, was strongly related to grade 4 outcomes. Our full model predicted 28 per cent of the variance in grade 4 reading comprehension, with an impressive 17 per cent of the variance associated with descriptions of children's preschool classroom language experiences. When one considers that there likely were indirect effects of preschool on grade 4 that were mediated by kindergarten classroom experiences, our data strongly suggest that early preschool language experiences are predictive of later reading success.

The findings of this study suggest the potential long-term importance of early childhood classroom support for language and indicate that multiple features of the language environments of classrooms can foster vocabulary learning and acquisition of language and the related abilities that foster reading comprehension. Our correlational findings are bolstered by the fact that much larger correlational studies also find associations between language abilities measured in the preschool years and reading in third and fourth grade (NICHD Early Child Care Research Network, 2005; Storch and Whitehurst, 2002). Additional support for the importance of preschool classroom experiences to children's language growth comes from experimental studies in which teachers were helped to employ strategies similar to those that our observational work identifies as being associated with enhanced language learning. In two studies, Wasik and colleagues (Tabors et al., 2003; Wasik and Bond, 2001; Wasik, Bond and Hindman, 2006) found that increasing teachers' verbal engagement with children and use of strategies designed to teach vocabulary resulted in enhanced learning.

Evidence now indicates that four general strategies tend to be associated with enhanced language learning: (1) extended talk on a single topic; (2) child opportunities to talk as they converse with teachers; (3) exposure to sophisticated vocabulary; and (4) intellectually challenging group

discussions. Unfortunately, careful studies of the quality of support for language available in classrooms indicate that such patterns of conversation are relatively uncommon in classrooms that serve young low-income children and that it is quite difficult to change patterns of language use.

Classroom supports for language learning

Results of the Home-School Study indicate that one-to-one and group conversations in classrooms can support the language learning of children from low-income homes. We now discuss what is known about the frequency with which classroom teachers support children's language learning and briefly consider interventions designed to increase support for language learning.

Early childhood settings

Few research teams have described the details of language support provided to children and most work has been done in preschool classrooms. In a noteworthy early project Tizard and Hughes (1984) examined children in their homes and classrooms and found the richness of language in classrooms to be limited, especially for children from working-class backgrounds. In the early 1990s researchers in the United States spent a week observing 119 classrooms (Layzer, Goodson and Moss, 1993). They found that lead teachers engaged in one-to-one or small-group interactions with children 26 per cent of the time, and that slightly less than 28 per cent of the time teachers were interacting with no children. Half or more of the children had no opportunities for individual attention from an adult during a day in 20 per cent of the classrooms. Another study examined teacher–child interactions in university-affiliated preschool classrooms. Researchers noted the frequency of interactions between teachers and children when they were in close proximity (three feet or less apart) and found that 81 per cent of the time, teachers did not talk to children they were near (Wilcox-Herzog and Kontos, 1998).

In the mid-1990s teachers were observed in 61 Head Start classrooms and interactions were coded using a time-sampling system designed to describe evidence of the use of strategies found by earlier analyses of HSSLD data to be supportive of language learning (Dickinson, 1994; Dickinson and Beals, 1994; Layzer et al., 1993). Observations revealed that teachers engaged four-year-old children in conversations that stayed on and developed a single topic less than 19 per cent of the time during meals and in only 14 per cent of the observed 30-second intervals during free play. They almost never made explicit efforts to teach vocabulary (1 per cent or less of the intervals) and engaged children in the kind of cognitively enriching conversations likely to foster development of language skills associated with literacy development about one-quarter of the time (Dickinson et al., 2004).

Recently similar methods were used to describe patterns of talk among four Head Start teachers. Slightly over two hours of teacher–child talk were recorded during 'centers time', the period of time during which children engage in independent child-initiated activity, over three days in two free-play settings, blocks and dramatic play, with teachers being asked to spend 10 minutes in each of these areas (Dickinson, Darrow and Tinubu, 2008). Transcripts were coded for teachers' efforts to teach information or vocabulary explicitly. Across the four teachers such comments were observed in 1–3 per cent of all teacher conversational turns. During the 30 minutes when the teacher with the lowest rate of such comments was observed she only discussed words or taught new information five times (i.e. one utterance once every six minutes).

One instance when a teacher did seek to teach a new word and associated conceptual knowledge highlights another dimension of the challenges teachers face related to vocabulary instruction. When talking about a white bear a teacher was trying to explain why the bear could be called a 'polar bear'. She said, 'that's why they call him a polar bear 'cause he live in the cold'. Here the teacher used a low-frequency word while providing some semantic information about the name. She sought to build relevant background knowledge about a complex topic, but her explanation fell short of providing a truly solid conceptual base for the notion of polar bears because a truly satisfying explanation would require an extended interaction and likely would require props (a map) and concepts that would be challenging for the child. This exchange highlights the challenges teachers face as they seek to build lexical knowledge and the associated conceptual base among children who lack both. If the teacher herself is not accustomed to explaining such complex ideas, she faces an even more daunting challenge.

We also were interested in how often teachers stayed on and extended a topic because our data and other studies (Barnes, Gutfreund, Satterly and Wells, 1983; Hoff-Ginsberg, 1991; Nelson, 1989) indicate that such talk is associated with improved learning. Extended interactions may foster vocabulary learning because the length of interactions between parents and children is correlated with the number of sophisticated types and tokens used in different contexts (Weizman and Snow, 2001). In addition, children may better comprehend what is being discussed when a topic is discussed in some detail and, as a result, better understand the meanings of new words. This hypothesis is consistent with our finding of the facilitative effect of analytic talk during book reading. Given the importance of extended exchanges, it is unfortunate that, of four teachers we observed three times each, we only found between one and five extended topically related sequences in each 10 minutes of interaction.

Support for language development is of continued importance for school-aged children because language skills continue to vary dramatically and to correlate with income, race and ethnicity (National Center for Education

Statistics, 2007a). Schooling plays a role in fostering continued vocabulary growth in kindergarten and grade 1 because children show greater gains in syntactic understanding and vocabulary during the six months when they are in school than the six months that include summers (Huttenlocher, Levine and Vevea, 1998). Although classrooms can support language learning during the early primary grades, schools do not close the gap in vocabulary knowledge that exists when children enter school (Biemiller and Slonim, 2001). Biemiller (2006b) noted that, at the end of the preschool years, children from different backgrounds differ in the number of root words they know by several thousand. During the school years the rate of growth of vocabulary is roughly parallel across groups. These findings further emphasize the importance of addressing children's vocabulary needs at an early age – preferably during the preschool years.

Primary grade classrooms

There is scant research describing how primary grade classrooms support growth of children's vocabulary and language through conversations that occur throughout the day, but one exception is a study that examined the nature of support for language and other academic skills provided to children in primary grade classrooms. The researchers observed students in their classrooms three times during the academic year, coding the instructional activities they participated in during the day. Children were assessed at the beginning and end of the year to determine how their abilities interacted with their classroom experiences and their language- and literacy-related outcomes. This study found that teacher-facilitated meaning-focused activities such as oral reading, listening comprehension and vocabulary instruction predicted students' reading skill growth (Connor, Morrison and Petrella, 2004). Unfortunately, such activities only accounted for about 24 minutes of overall instructional time each day in first- and second-grade classrooms (Connor, Morrison and Underwood, 2007). Further support for the potential value of such conversations comes from an observational study of a kindergarten classroom where a teacher's skilful questioning and elicitation of explanation led to increased child independence in comprehending text and greater literacy understanding (Hansen, 2004).

Researchers interested in language use in primary grade classrooms have described patterns of teacher–child interaction and found that initiation–response–evaluation (IRE) sequences of questioning and recitation are dominant (Blank and White, 1986; Cazden, 2001; Mehan, 1979). These typically are used to test children's acquired knowledge, rather than draw them into analytical conversations, which help scaffold their language use. While such interactions might include sophisticated vocabulary, they tend to lack other features of interaction we found to be beneficial – a conversational balance that favours child participation, encouragement for children to extend their thinking, and analytical thinking. One primary grade

classroom activity that allows children opportunities for extended turns is sharing time (Cazden, 2001; Gallas, 1992; Michaels, 1981). There is some evidence that the repeated opportunities this activity provides children to construct stories may help them create more coherent narratives and develop proficient semantic and syntactic expressive communication (Gallas, 1992). The potential value of this activity for vocabulary learning comes from an experimentally controlled intervention in which preschool children were encouraged to recount narratives and children were found to have improved vocabulary as a result (Peterson, Jesso and McCabe, 1999).

A number of researchers have implemented interventions designed to help teachers engage children in the type of extended, reflective discourse found to be associated with enhanced vocabulary learning. These efforts seek to create norms for interacting in classrooms that provide explicit guidance regarding conversational norms and encourage extended, analytical discourse. One approach, Accountable Talk, has been employed in classrooms in the United States and has been found to have a beneficial impact on reading comprehension instruction in urban elementary and middle schools (Wolf, Crosson and Resnick, 2004, 2006). Accountable Talk is a structure for teacher-facilitated classroom discourse which fosters a high level of student explanation, engagement and responsibility. Specifically, there are seven observable components: participation rate, teacher's linking ideas, students' linking ideas, asking for knowledge, providing knowledge, asking for rigorous thinking, and providing rigorous thinking (Wolf et al., 2006). Research conducted over a 15-year time span consistently revealed that children in classrooms employing this approach had enhanced academic growth across a range of subjects and grade levels (Michaels, O'Connor and Resnick, 2007). A similar approach, Thinking Together, uses exploratory talk as an instructional method across content areas with primary school students. Thinking Together has been employed in classrooms in England and found to have beneficial effects on features of discourse associated with improved vocabulary learning (Littleton, Mercer, Dawes, Wegerif, Rowe and Sams, 2005). A third approach devised by Goldenberg (1992) builds teachers' abilities to engage children in 'Instructional Conversations', defined as discourse that is connected, reflective and provides children opportunities to participate and listen to each other. An experimental study conducted in upper elementary classrooms with significant numbers of English language learners (ELLs) found this method resulted in enhanced reading comprehension (Saunders and Goldenberg, 1999).

Strategies with a discourse-based approach also have been created to foster better conversations about books. Such methods hold promise for improving vocabulary learning because books are a rich source for low-frequency vocabulary. Reciprocal Teaching (Palinscar and Brown, 1984) is a method that employs discussion of expository books, a genre particularly likely to

yield opportunities for talk about the meaning of words. It has been found to foster gains in comprehension among seventh-grade children that were maintained on other tasks. It also has been found to be effective when used as early as first grade (Coley, 1993). Beck and her colleagues (Beck, McKeown, Hamilton and Kucan, 1997; Beck, McKeown, Sandora, Kucan and Worthy, 1996; McKeown and Beck, 2003) also devised approaches designed to improve conversations about books – Questioning the Author and Text Talk – where teachers lead students in active discussions about the meaning and purpose of a text. Both have been found to have beneficial effects on classroom discourse (Sandora, Beck and McKeown, 1999) although the researchers did not look for changes in vocabulary knowledge. Effects on vocabulary were reported for an intervention designed specifically to teach word meanings (Beck, Perfetti and McKeown, 1982) by providing focused practice that included engaging children in using words in varying contextualized ways including book discussions and novel shared experiences (Beck, McKeown and Kucan, 2002).

In summary, studies conducted in classrooms indicate that teachers can support vocabulary learning and that variability in the support children receive can have important effects. Unfortunately, available evidence indicates that the level of support available to many children is limited in quantity and quality. Interventions have been devised to enhance classroom discourse in primary grade classrooms in ways that are consistent with patterns associated with enhanced vocabulary learning in preschool classrooms, but researchers have not directly examined the impact on vocabulary learning.

Children learning English as a new language in classrooms

More and more children who are second language learners are entering public schools. In the United States there were over 4.7 million schoolchildren designated as limited English proficient (LEP) in 2000–1, almost 10 per cent of the total US school age population (National Clearinghouse for English Language Acquisition, 2003). In addition, the number of these students designated LEP has increased by 95 per cent while the overall enrolment has only increased by 12 per cent. In early childhood, the numbers were even greater. In 2001, 15 per cent of the students enrolled in pre-kindergarten classrooms located within elementary school settings were identified as LEP (National Center for Education Statistics, 2001). In 1998, children spoke over 140 different languages in Head Start classrooms representing 24 per cent of the enrolled population (Head Start Bureau, 2000).

By and large, the educational system in the United States does not work for English language learners. In one large survey in 2000–1 that used reports from 41 state education agencies, it was reported that only 18.7 per cent of the students designated as of limited proficiency in English were

able to reach the state norms for reading achievement in English language reading comprehension exams (Kindler, 2002). As recently as 2007, the NAEP results revealed that 70 per cent of ELLs in fourth grade scored below basic levels of achievement in reading (National Center for Education Statistics, 2007b).

In theory, young ELL children in the United States have three options for preschool programmes. The most common situations are a first language classroom, a bilingual classroom or an English-language classroom (Tabors and Snow, 2001). In the first language classroom, the teacher is typically a native language speaker of the children's first language (and, typically, fluent in English). In this classroom, the first language is developed while also providing conceptual knowledge in that language. In a bilingual classroom, the teacher is fluent in both the first language and English. The students' first language is supported while they are beginning to learn English. In an English-language classroom, the teacher is typically a native speaker of English and all instruction is given in English.

It is widely believed that the most effective way for ELLs to develop oral proficiency and literacy in English is by encouraging and creating a strong background in the native language (Cummins, 1981; Hakuta, 1986; Krashen, 1985; Mace-Matluck, 1982; Ramirez, 1991). Furthermore, it has been found that time spent on native language instruction during younger years does not detract or take away from English reading acquisition in later years (Reese, Garnier, Gallimore and Goldenberg, 2000). Strong vocabularies in native language have been shown to contribute to a greater fluency in the second language (Proctor, August, Carlo and Snow, 2006). A strong theoretical base explains these findings. If children have a strong background in their native language, their conceptual map is already in place and a solid foundation has already been laid upon which to learn English.

However, most ELL preschool children are put into English-only classrooms led by teachers who are monolingual speakers of English (SocioTechnical Research Applications Incorporated, 1996). In fact, there are some states, such as California, Arizona (in 2000) and Massachusetts (in 2002), that have passed legislation that make it almost impossible to provide bilingual education to the average ELL. Relatively little is known about the effectiveness of quality instruction on ELLs in English settings (Shanahan and Beck, 2006).

A small amount of descriptive work and one small intervention study have shed light on the oral language development of young ELLs in English-language classrooms. These studies found certain phases that were exhibited by language minority preschool-aged children beginning with a 'silent period' (Saville-Troike, 1988) and progressing through use of strategies such as repeating, memorizing, use of rote expressions, chorally

responding, self-talk, elaborating, predicting, monitoring, asking for help and role playing (Chesterfield and Chesterfield, 1985).

Research has yet to demonstrate what, if any, kind of instruction will increase the rate at which a child reaches oral proficiency in English or how specific types of instruction in English have a beneficial impact on the oral language development of ELLs (Saunders and O'Brien, 2006). This is particularly the case with early childhood settings. One study suggests that teachers may play a particularly important role in supporting language learning of ELL children. Evidence of the impact of teacher input on vocabulary learning comes from Genishi, Stires and Yung-Chan (2001) who looked at vocabulary acquisition in predominantly Chinese-speaking preschoolers. They found that a common set of vocabulary words developed among the students regarding class rules and routines. The interactions that students had with teachers were those most strongly related to vocabulary development in English, because the children tended to speak Chinese with their peers, and thus those interactions did not seem to support English learning (Genishi et al., 2001). These findings of the importance of the teacher as a source of vocabulary learning echoes findings from the HSSLD study's examination of preschool classrooms discussed earlier that found enduring effects of the quantity and quality of teacher–child conversations on the language learning of children who were native English speakers (Dickinson, 2001b; Dickinson and Porche, under review). Peck (1987) found this same result, but with a peer tutoring programme between second graders and kindergartners. The students who were the better speakers of English spoke more with their 'teacher' and used more developed vocabulary (Peck, 1987).

The longitudinal studies of the contribution of language in early childhood to later literacy described earlier have spurred some to conduct language interventions in early childhood. This need for early intervention also applies to ELLs as is evidenced by the achievement gap seen in fourth grade. There have been small-scale attempts at interventions with ELLs in English-language classrooms and these studies have examined the impact of an instructional approach or strategy. The majority of them have found that effective literacy instruction provided to native speakers of English also has a positive impact on students who are learning English as a second language (Fitzgerald and Noblit, 1999; Shanahan and Beck, 2006). In general, researchers have found that, if given high-quality instruction, ELL children will progress at a rate similar to that of their native language-speaking peers. The one exception to this pattern has been with vocabulary (Fitzgerald, 1995; Fitzgerald and Noblit, 1999; Garcia, 1991). Fitzgerald and Noblit (1999) completed a year-long study of a first-grade classroom. As the classroom teacher, Fitzgerald implemented a high-quality, well-rounded literacy programme and her ELLs were able to make progress similar to that of their native English-speaking peers in all areas except for vocabulary.

All but one of the language minority students in her class scored at the bottom half of the word-meaning test. Fitzgerald and Noblit hypothesized that this result might be attributed either to challenges associated with learning vocabulary or to the fact that the reading programme lacked a specific plan to address vocabulary acquisition. Our hunch is that both factors played a role in accounting for the limited growth in vocabulary. Teachers typically are not aware of the potential of informal conversational settings for teaching vocabulary and therefore fail to take advantage of them. Most planned instruction is guided by published reading programmes and vocabulary is likely not to become the focus of instruction unless a teacher realizes the need for it. Failure to plan for vocabulary instruction could be a by-product of limited instructional time and the difficulty knowing what words to teach. Additionally, teachers must be aware of the depth of their students' word knowledge, building a rich conceptual understanding of words in the second language rather than covering a wide breadth of vocabulary (Verhallen and Schoonen, 1998).

In other interventions, ELL children have successfully learned vocabulary when they were taught explicitly in different linguistic contexts, with repetition and multiple opportunities to use the words (Collins, 2005; Silverman, 2007). This addresses the need for depth of word learning, which is important whether students are instructed in their first or second language (Verhallen and Schoonen, 1998). The finding that children benefit from hearing stories multiple times and from being provided explicit definitions for words is consistent with results of book reading interventions designed for teaching vocabulary in a first language (Biemiller, 2006b; Elley, 1989; Feitelson, Goldstein, Iraqi and Share, 1993; Sénéchal, Ouellette and Rodney, 2006). Two early experimental studies examined the effects of exposing elementary school-aged children to extensive high-quality language in a language other than their native tongue and found clear evidence of benefits from this exposure. Israeli Arab-speaking children learned the academic variety of Arabic required for literacy (Feitelson et al., 1993), and in the South Pacific Elley found beneficial effects on language learning in schools serving rural Fijian children (Elley and Mangubhai, 1983). They found strong effects associated with regular exposure to books. Books have also been used as a basis for intervening with preschool-aged ELL children. Collins (2005) found that repeated readings of books combined with explicit definitions for words led to significant contributions to vocabulary development. In a parallel study involving teacher-led book reading, ELL children learned target words at the same rate as native English children and the ELLs' overall vocabulary rate grew at a faster rate than native English speakers (Silverman, 2007).

Taking a different approach to vocabulary instruction, Vaughn-Shavo (1990) used a comparison group strategy in an intervention with 15 native Spanish-speaking first graders. She found that teaching the 31 target words

in meaningful narratives with their own sentences and picture cards was more effective than teaching words in a sentence context only. The findings from these studies seem to be consistent with the general findings of preschool research described earlier in that a critical feature of the rate that children progress toward mastering a new language in classrooms is the extent to which children interact with the teacher and hear good language models (Dickinson, 2001a). The darker side of research is that ELL children who do not have access to such instruction may make only modest progress in acquiring English skills – especially in the area of vocabulary.

Concluding thoughts

Evidence we have reviewed reflects mounting interest in vocabulary because of its strong associations with reading comprehension and the need to address the vocabulary learning of children who too often later struggle to become proficient readers. While interest is mounting related to these issues, far too little is known about how classrooms foster vocabulary learning of children who are acquiring a first language or children who are newly learning the language used in the classroom. Questions worthy of study include: What types of teacher language use and discourse facilitation predict vocabulary learning and related achievement outcomes? What settings are most conducive to supporting language? Do peer interactions foster language learning and if so under what conditions? How do differences among children affect their ability to benefit from different kinds of classroom support: age, gender, first language and language ability? How do differences among teachers, including beliefs, language, culture and education, affect the way they support language and vocabulary?

Our understanding of how to intervene to foster language growth is also limited. One problem is that the work on interventions lacks integration. Different constructs have been used and, except for book reading studies, strategies found to be promising have not been examined by different research teams in different classroom environments. A closely related challenge is the need to understand how professional development can enable teachers to adopt and sustain use of conversational skills that support children's language learning. Considerable work is needed to see if more efficient ways can be developed to support teachers. Research needs to take into consideration characteristics of the teacher (e.g. language, cultural and educational background), setting (e.g. grade level, backgrounds and language skills of the children), and the nature of the intervention (e.g. tightly prescribed versus open-ended, group size, content area). It is our hunch that the most efficient way to create early childhood classrooms which build language skill and enhance children's lexical knowledge is to provide teachers a curriculum created with the goal of supporting language and vocabulary development and to provide professional development that

builds teachers' knowledge of language and enables them to see how the curriculum can be used to support language learning. We have far to go, but at least we are beginning to recognize the need for the journey.

Note

1. Work on this essay was supported by a grant from the Institute for Educational Sciences R324E060088A, Department of Education. Secondary authors are listed in alphabetical order.

3

Exploring Vocabulary with Young L1 Learners: the Contribution of a Corpus

Alison Sealey
University of Birmingham, UK

Introduction

While there have been many studies of the initial acquisition of L1 vocabulary in infancy, comparatively little research has been conducted on young L1 vocabulary learners in the school context. Of this research, very little demonstrates the influence of developments in corpus linguistic analysis and description. Literature in this field is instead dominated by studies of learners of additional languages, and, consequently, most research findings concern learners in the secondary or tertiary phases of education, leaving the primary learner somewhat neglected. This chapter therefore discusses, with particular reference to young learners of English as a first language, some key themes that emerge from the literature on lexis and on vocabulary learning, and then reports on a study conducted in two primary (elementary) school classrooms with monolingual speakers of English aged between eight and ten years. The study investigated ways in which a corpus can be used in this context: by teachers to supplement their intuitions about vocabulary and to prepare teaching materials and opportunities; and by pupils, as a resource for exploring the wider lexico-grammatical behaviour of the vocabulary items they encounter.

Vocabulary and success

My first observation is that, while vocabulary is recognized as a legitimate challenge for learners of a foreign language, discussion in the L1 literature is often tinged with 'the idea that vocabulary size is a reflection of how educated, intelligent or well read a person is' (Nation and Waring, 1997, p. 7). It is easy to find current examples in everyday discourse of the point of view associated with *Reader's Digest* that 'it pays to increase your word power'. The following online examples appear to play on popular anxiety about vocabulary size: 'Whether you realize it or not, every time you

speak, people use your language to instantly "tip them off" about how educated, competent, and successful you are.' This site ('Ultimate Vocabulary') goes on to link the promise of social advancement with a bigger and 'better' vocabulary: 'It's not surprising … that study after study has shown that a powerful vocabulary is directly linked to success, status, and income.' A similar example from the 'Power Vocabulary Builder' site makes even more extravagant claims:

> You already know it's true … Mastering a superior vocabulary will help you succeed in every area of your life. When you … know just the right words for any occasion … when you have the confidence and respect that only a top-notch vocabulary can deliver … Literally anything is possible. With this kind of confident vocabulary you can: Rapidly earn promotions and higher salaries … Enjoy closer relationships because of better communication …

Written language as the source of ideas about vocabulary

As well as the connection between schooling, literacy and spoken language, another popular idea about language is that writing 'sets the standard'. It is thought that speech 'consistently produces a less varied vocabulary' (Chafe and Danielewicz, 1987, p. 89), and that wide reading is a potential remedy for the lexical deficiency of the spoken language. The 'educated' person displays in spoken interaction the sizeable vocabulary acquired through extensive reading, confirming the perception attributed by Halliday to schools and 'the community at large' that (at least prior to research in the 1970s and 1980s) language 'meant written language … The general picture is that of written language as richly endowed, while speech is a poor man's assemblage of shreds and patches' (1987, pp. 55, 67). Although there have been extensive advances in research, and in its classroom applications, which challenge this perception, 'vocabulary' is closely associated with written language. Indeed, it has been argued that the very concepts of 'words' and 'definitions' are dependent on – possibly artefacts of – written, and especially printed, texts. 'It seems to be only in the context of reading and writing that the whole vocabulary is assigned word boundaries' (Wray, 2002, p. 137; see also Watson and Olson, 1987). In any case, in the context of the English L1 primary education syllabus, 'vocabulary' features most prominently as a dimension of literacy, which is consistent with the belief that success in reading and larger vocabularies are mutually reinforcing. Biemiller (2007, p. 6), for example, claims 'that vocabulary development is a key factor in successful reading development. Children with below-average vocabularies are at-risk of low school achievement.' Likewise, Dickinson and Sprague (2002, p. 276) maintain 'that vocabulary is essential

to later literacy, that growth in vocabulary is related to broader discourse skills, and that engagement in extended discourse that requires decontextualized language skills fosters literacy development'. Advice for teachers from the English government's Department for Children, Schools and Families (DCSF) puts it more succinctly: 'The more words you know, the more words you are likely to learn!' (Primary National Strategy, 2007a; see also Primary National Strategy, 2007b).

Policy and practice in L1 English vocabulary instruction

As I have noted, however, the research evidence on effective classroom practice in respect of the vocabulary of young L1 English learners is relatively sparse. One summary provided by the DCSF (Primary National Strategy, 2007b) confirms the more extensive research effort in this area of North American investigators than those in the UK – despite the fact that, as Nagy put it in 1997, '[m]ost American schoolchildren receive very little vocabulary instruction …' (p. 71), and '… vocabulary instruction of the sort that has been demonstrated to increase reading comprehension is relatively rare in schools' (p. 73). The Primary National Strategy's summary offers some advice to teachers, particularly to support learners with 'low' vocabularies, and recommendations include: '**defining and explaining** word meanings; arranging **frequent encounters** with new words (at least six exposures to a new word); and encouraging pupils' **deep and active processing** of words and meanings in multiple contexts' (2007b, p. 2, emphasis in original). Less evident in this summary is the lack of consensus in the relevant literature about the respective significance of incidental learning and direct instruction, and about the kind of direct instruction likely to be most effective (e.g. Anderson and Nagy, 1992; Cain, 2007; Nagy, 1997; Phythian-Sence and Wagner, 2007). Current advice, reasonably enough, is that 'a range of approaches used together is most effective' (Primary National Strategy, 2007b, p. 2).

Education policy in England is subject to the National Curriculum Programmes of Study for each subject, including English, and the word 'vocabulary' does not feature at all in the 'Speaking and Listening' section at Key Stage 1 which applies to pupils aged 5–7 (DfEE and QCA, 1999; QCA, 2008). At Key Stage 2 (ages 7–11), it occurs twice, in the requirements that pupils be taught: 'to … use vocabulary and syntax that enables [*sic*] them to communicate more complex meanings' and 'about how language varies … according to context and purpose (for example, choice of vocabulary in more formal situations)'. There is one reference to 'challenging' vocabulary in the Programmes of Study for Reading at Key Stage 1, while at Key Stage 2, the word is used once each in relation to the 'understanding and appreciation' of 'literary' and 'non-literary' texts respectively. It is in the Writing dimension of the English curriculum that the idea of extending

learners' vocabulary is made most explicit, although again there are only two occurrences of the word: 'use adventurous and wide-ranging vocabulary' (Key Stage 1); 'broaden their vocabulary and use it in inventive ways' (Key Stage 2).

The other policy document to which English teaching is subject is the *Primary Framework for Literacy and Mathematics*, which incorporates the *National Literacy Strategy* (NLS) 'Framework for teaching' (DfES, 1998). In this document, objectives are specified for teaching in each term of each school year, organized into the 'levels' of word, sentence and text, and 'vocabulary' consistently collocates with 'extension' (19 occurrences). In the teaching specified for pupils from Reception Year (age 4–5) to Year 6 (age 10–11), the string 'new words' occurs 16 times, and what pupils are to do with these is frequently 'make collections'. Words are depicted as having inherent properties, so that some are 'more accurate or interesting than the common choices, e.g. *got, nice, good, then*'. In addition, several other strands run through the specifications for 'word level' work, including developing children's knowledge of:

- spelling conventions and dictionaries;
- rudimentary morphology ('the ways in which nouns and adjectives, e.g. *fix*, *simple*, *solid*, *drama*, *dead* can be made into verbs by use of the suffixes *-ate*, *-ify*, etc.');
- diachronic and regional variation in vocabulary;
- synonyms and antonyms and other aspects of word meaning.

Contemporary research into vocabulary has developed in a number of ways. However, as will become apparent, the findings do not yet appear to have been taken into account by policy makers in this sector of education. The following section outlines some of the insights that I would argue could usefully be incorporated into this area of education.

Lexico-grammar across speech and writing

Firstly, while '[o]fficial policy usually equates educational knowledge with the written mode and commonsense knowledge with the spoken' (Halliday, 1987, p. 80), the respective properties of spoken and written language are beginning to be perceived in ways that are more nuanced than those suggested by the National Curriculum. As suggested above, it is often assumed that a very limited set of word families 'provide the lexical resources to engage in everyday spoken English discourse …'. However, corpus studies challenge this assumption, and '… suggest that more vocabulary is necessary in order to engage in everyday spoken discourse than was previously thought' (Adolphs and Schmitt, 2003, p. 425). Also easily overlooked is the complexity of speech. As Halliday points out, the products of writing conceal the planning, hesitations

and revisions that have gone into their production, and thus the similarities of each mode as 'highly organized, regular, and productive of coherent discourse' (1987, pp. 69–70). Where they do differ, he maintains, is 'in their preferred patterns of lexicogrammatical organization' (p. 71). This view of the lexis–syntax relationship contrasts with the more conventional approach, where '[t]he lexicon … provides the content for syntax and the instantiation of syntactic rules' (Clark, 1993, p. 259). The recognition that 'vocabulary' may instead best be viewed not as distinct from 'grammar', but as part of a single lexicogrammatical system, has far-reaching implications, including the finding that:

> … grammatical patterns and lexical items are **co-selected**, and … it is impossible to look at one independently of the other. Particular grammatical patterns tend to co-occur with particular lexical items, and – the other side of the coin – lexical items seem to occur in only a limited range of patterns. The interdependence of grammar and lexis is such that they are ultimately inseparable, working together in the making of meaning. (Clear, Fox, Francis, Krishnamurthy and Moon, 1996, p. 311)

It may be that, in time, this insight will go some way towards counteracting the bias towards writing as the basis for language description (Carter and McCarthy, 1995; Carter, 2003). The identification of patterns such as those discussed by Hunston and Francis (2000) in their 'corpus-driven approach to the lexical grammar of English' make it possible to see 'words' less as separate entities and more as the somewhat artificial constructs of textual practices:

> … speech consists of a continuous stream of sounds, while the concept of 'word' requires the segmentation of the stream of speech into meaningful units of a particular configuration. It seems that alphabetic, word-segmented scripts engender the notion that utterances are decomposable into lexical constituents called words. (Watson and Olson, 1987, p. 335)

Of course, corpus methods themselves rely on a version of print technology, and import the concept of the 'word' into investigations and calculations. As Gardner points out, the traditional ways of demarcating words, and of grouping them into 'families', can lead to unquestioned, and perhaps misleading, assumptions, including those about the relationships between variant forms of a lemma:

> … the case of the irregulars poses serious quandaries relating to the psychological validity of such family relationships – namely, that the opaque spelling and phonological connections between the lemma headword and the family members will surely cause more and different

> learning problems than their more transparent counterparts. (Gardner, 2007, p. 244; see also Read and Nation, 2004)

While a study by Tyler and Nagy (1990) 'offer[s] support for morphologically organized models of the lexicon' (p. 17), Biemiller and Slonim (2001, pp. 498–9) point out that young children may understand many derived or inflected words '… directly without any observable process of derivation. … two thirds of the derivatives of the root words used in the studies [they surveyed] are reported as "known" … at the same or younger age level as the level at which the root was identified'.

Differential behaviour of morphologically related forms

The present study made use of a corpus, extracted from the British National Corpus (BNC), of texts categorized as having been written for a child audience, which were predominantly imaginative prose. This was termed the 'CLLIP' corpus, from the title of the ESRC-funded project, 'An investigation into *C*orpus-based *L*earning about *L*anguage *I*n the *P*rimary-school'.[1] As well as providing a teaching resource, the corpus was investigated to identify various linguistic features of this kind of writing. The NLS prescribes lists of words which pupils must know by certain stages, and a comparison of the vocabulary of the CLLIP corpus with these lists revealed some interesting discrepancies in forms of the same lemma (Sealey and Thompson, 2006). For example, the lemma *follow* appears in the Year 4–5 list as 'follow(ing)', but neither of these forms is in the CLLIP list of 1000 most frequent words. However, the form *followed* appears just over halfway down at 586, with 132 occurrences. From the lemma *try*, the NLS lists the form *tries*, but this is not among the top 1000 words in the CLLIP corpus, while *tried*, at position 296, is much more frequent (as are both *trying*, at 315 and *try*, at 382). Likewise, *looked*, at position 77 in CLLIP, is significantly more frequent than *look*, at 104, but only the latter is included in the NLS lists. Similarly, the NLS lists include some singular forms of nouns (*balloon, number*) and some plurals (*animals, friends*); our analysis showed that *animals* is indeed more frequent in CLLIP than *animal* and *friends* than *friend*. However, the choice for *year* in the list is the singular form, but it is the plural *years* which is much more frequent in the corpus (the two forms occur at positions 735 and 333 respectively). These findings are of course not definitive, but they do indicate how corpus analysis might be used to inform policy that is concerned with word frequencies.

In addition to differential frequencies, as Hoey (2003) points out, 'there are differences in the *behaviour* of the word depending on whether it is singular or plural' (emphasis added). This can be illustrated with data from the CLLIP corpus, where, for example, *friend* and *friends*, or *year* and *years*, are typically found in differently patterned phrases. The word *friends* is

frequently part of the patterns BE + *friends* + *with* and MAKE + *friends* + *with*, whereas *friend* often occurs in the patterns *a* + *friend* + *of* and the related *friend* + *of* + *mine*. Similarly, *year* occurs in *of* + *the* + *year*, *time* + *of* + *year* and *year* + *after* + *year*, whereas *years* is often found in (ENUMERATOR +) *years* + *ago*, and ENUMERATOR + *years* + *old*, *all* + DEMONSTRATIVE DETERMINER + *years*, and *years* + *and* + *years*. While the usefulness of such corpus evidence has been demonstrated for learners of English as an additional language, to a native speaker, who is likely to be very familiar with this kind of patterning, it may seem trivial even to note it. However, it is analyses like these that have led to a renewed interest in, and development of, the concept of 'colligation', defined as 'the grammatical company a word or word sequence keeps', the 'grammatical functions preferred or avoided by the group in which the word or word sequence participates' and 'the place in a sequence that a word or word sequence prefers (or avoids)' (Hoey, 2005, p. 43). A traditional grammatical approach would not distinguish between two of the patternings of *year* and *years* in the CLLIP corpus which could be analysed as NOUN + PREP + NOUN [= *year(s)*]. When the 'preposition' is *of*, however (and Sinclair, 1991, throws doubt on this classification, since *of* is so unlike other prepositions in the structures in which it occurs), two quite distinct patterns emerge. The string *of* + *years* is invariably preceded by an enumerator (*couple*, *hundreds*, *thousands*). The apparently near-identical string *of* + *year* is preceded exclusively by a noun phrase headed by *time* ('this time of year', 'a different time of year', 'the wrong time of year').

It has long been recognized that an important component of learners' vocabulary knowledge may well be familiarity with the patterns associated with collocation and colligation (see, for example, Cowie, 1999, for a discussion of the influence on EFL dictionaries of the work of Palmer and Hornby in the 1930s). The absence of such understanding is partly responsible for the situation so familiar to teachers where non-native, classroom-taught learners produce grammatical, but unidiomatic, language (see Wray, 1999, p. 227). Some studies with young L1 learners have demonstrated that classroom dictionary exercises may be similarly unsuccessful in leading to idiomatic production. For example, Miller and Gildea (1987, cited in Pressley, Disney and Anderson, 2007) report on the 'mystifying' sentences that children produced when given the task of consulting dictionaries for definitions of unfamiliar words and trying to use them in their own sentences. Such examples suggest that individual words and their lemmas, with which vocabulary learning researchers and syllabus designers are often preoccupied, may be necessary but not sufficient for vocabulary extension. Single word forms are relatively easy items for syllabus designers to specify, and for vocabulary test designers to measure, but they may not be sufficient for learners to build more sophisticated meanings. 'The atomistic approach to meaning taken by vocabulary researchers', claims Oakey (2005, p. 169),

'obscures some lexico-grammatical aspects of … discourse'. He goes on to suggest:

> that students' difficulties in reading arose less often from encounters with unfamiliar words, but more often from unfamiliar combinations of familiar words. The pedagogical implications of the study [of academic discourse in an EAP context] are that phraseological information needs presenting alongside more traditional vocabulary items.

Even for young L1 learners, it seems highly likely that authentic 'vocabulary extension' will involve the internalization of phrases, collocations and colligations and not simply the acquisition of new collections of root forms.

Delexicalized words

The NLS requires learners to be taught about word classes, and some of the activities we devised in our research project were linked with these objectives. Our corpus was part of speech (POS)-tagged automatically, since it was a subset of the BNC, and this enabled us to generate concordance lines that were colour-coded in respect of basic word classes. As reported elsewhere (Sealey and Thompson, 2004), discussion suggested that some of the children we worked with had a rudimentary concept of the distinction between lexical and grammatical (or content and function) words. When trying to work out what the words appearing in black had in common – they included pronouns, articles and the 'to' infinitive marker – one boy suggested that 'they might just be like dull words kind of thing', adding that words such as *the* and *they* were 'not like that exciting'.

Less apparent from traditional part-of-speech classifications, however, is that many common words fall into a category that is neither fully lexical nor fully grammatical – words that are 'delexicalized', 'semantically depleted' or 'desemanticized' (Stubbs, 2001). For example, in the CLLIP corpus, *take* occurs 711 times. Stubbs maintains that the 'dictionary meaning' of *take* – 'grasp with the hand' or 'transport' – is much less frequent than might be thought, and in the corpus investigation he carried out, most examples occurred in expressions 'where almost all the meaning is carried by the noun' – such as 'take an interest in', 'take a deep breath', 'take a decision', and so on (2001, p. 32). The CLLIP corpus evidence is comparable.

In the CLLIP corpus, *take* occurs in the three-word strings *to take the* and *to take a* 27 times and 21 times respectively. Sometimes the meaning is indeed 'grasp with the hand' or 'transport', as in 'he decided to take the bicycle with him'; 'reaching to take the knotted hand that his father surrendered'. Other examples belong to the pattern *take* + NOUN PHRASE + PARTICLE, including 'went to the kitchen window to take the blackout

frame down'; 'soldiers hurried up to take the gardeners away'; 'I really was going to take the spell off straight away'. In addition are the more idiomatic occurrences such as 'she smiled to take the sting out of her words'; 'he was not yet ready to take the shabby magician fully into his confidence'; 'I've got to take the risk', and so on. When followed by *a* in the CLLIP corpus, *to take* seems to have an even lower incidence of literal meaning, with examples including 'to take a bath'; 'This is going to take a bit of brainwork'; 'to take a job'; 'to take a look'; 'to take a moment'; 'to take a risk' and so on.

As discussed elsewhere (Thompson and Sealey, 2007), one of the contrasts between writing for a child audience and writing aimed at adults seems to be that the more literal, less figurative meanings of the same words and phrases are more evident in the former. A potential pedagogical implication of this kind of analysis is that progression in vocabulary would be indicated by children's ability not only to make 'adventurous choices' but also to deploy frequent words in less literal contexts.

'Words' versus 'lexical units'

Closely related to these issues, but of perhaps greater significance still, is the evidence from corpus analysis that '… often the unit of meaning is [not an individual word but] a longer phrase or collocation' (Stubbs, 2001, p. 49). A radical position on this is adopted by some corpus linguists, such as Sinclair (2004, p. 281), who maintains that the referent of the term 'lexical item' should be extended to 'one or more words that together make up a unit of meaning'. Because of differences in both theoretical and analytical perspectives, there is no single term for these various kinds of 'longer phrase', and different writers are concerned, variously, with 'multi-word items', 'formulaic sequences', 'formulaic language', 'lexical bundles', 'lexical phrases', 'prefabs', 'chunks' and so on. Nevertheless, as research into such phenomena develops, it is quickly becoming apparent that these longer-than-a-single-word strings account for a significant proportion of the language (Erman and Warren, 2000; Foster, 2001a).

All of these findings about the nature of the (English) language itself must surely have implications for language learning and teaching. As noted above, these implications seem to have been more thoroughly explored in relation to second (or foreign, additional) language development than to learners for whom English is a first language. However, there is one area of potential overlap that does receive attention in the literature, and that is the possibility that the mental lexicon is organized – at least in part – in ways that reflect more nearly the descriptions that are now becoming available from corpus-based analysis. Hoey, for example, considers this possibility in his discussions of 'lexical priming', whereby both individual words and regular word sequences are 'primed for use in discourse as a result of the cumulative effects of an individual's encounters with [them]'

(2005, p. 13). Corpus data provide external evidence for the way speakers experience priming as they acquire language, so that words/lexical items/ 'stretches of sound stream' 'become imbued, by means of nesting, with a rich and complex web of socially embedded, genre-sensitive collocations, semantic associations, colligations and text colligations …' (Hoey, 2005, p. 160). As Schmitt, Grandage and Adolphs (2004, p. 128) suggest, 'It is not unnatural … to assume that the data drawn from corpus analyses reflects the psycholinguistic reality of how language is processed and produced.' This 'reality' can perhaps be shown to contrast with the more traditional accounts which saw grammar and lexis as separate systems, where information for learners was packaged in either the dictionary, or the thesaurus, or the grammar book.

Processing the lexico-grammar

Hoey (2003) exemplifies this with reference to the influence of the traditional demarcation between lexis and grammar on the 'pre-corpus dictionary'. Such a source, he suggests, typically: provides a definition; for nouns, gives brief information on the count/uncount status; indicates how the word is pronounced; and might also provide information on the use of the word in specific fixed phrases. It does *not*, however: provide information on the most common collocations; identify the semi-fixed expressions in which it participates; provide any clue about the different behaviour of the singular and plural forms; give any indication of the semantic associations of the word; nor any clue as to its colligations. While extensive consideration of the psycholinguistics of the lexicon is beyond the scope of this chapter, it seems plausible that, for the language user, information is not stored and retrieved in the linear arrangements favoured by much language description and many textbooks, and cognitive and psycholinguistic research does indeed shed some light on this. In her discussion of formulaic sequences, Wray (2002, p. 9) suggests that language users both store and retrieve 'whole from memory at the time of use' some components that are larger than a single word. There seems to be a clear processing advantage to managing some language units in 'chunks' (Kuiper, 2004; Schmitt and Carter, 2004), and there is some empirical evidence that this does indeed happen. For example, a study of eye movement in reading '… provide[d] evidence for the position that formulaic sequences are stored and processed holistically' (Underwood, Schmitt and Galpin, 2004, p. 167). Moon (1997, p. 43), likewise, claims that

> language is processed – and often acquired – in chunks or groups of words, rather than on a word-by-word basis. This … has important repercussions with respect to vocabulary learning and teaching, since words are again and again shown not to operate as independent and interchangeable parts of the lexicon, but as parts of a lexical system.

What are the implications of these developments for the young learner in the literacy classroom? As Meara (1997, p. 121) observes, '[t]he effect of language models … filters down to the classroom. The way we think about language … governs the type of textbooks which get published and the types of methodology advocated by teacher-training institutions.' As indicated above, current policy in this area rests on a concept of vocabulary largely uninfluenced by research in either corpus linguistics or the psycholinguistics of formulaic language. The alternative approach deployed in the study described below sought to identify, in a very initial exploration, some of the possibilities offered by corpus approaches.

A corpus in the classroom

The CLLIP research aimed to explore the contributions that could be made by a corpus and a modified version of the associated software in helping young monolingual English children to learn *about* their first language. The questions guiding the research included: How do primary school pupils respond to corpus-based teaching and learning activities? and What kinds of metalinguistic knowledge, understanding or misconceptions are the children prompted to articulate by the presentation of texts in a corpus format (such as concordance lines)?

As explained above, the corpus comprised texts extracted from the BNC, starting with all those classified as having been written for a child audience. While it would have been potentially very fruitful to include some transcribed spoken language (see, for example, Mauranen, 2004), the additional challenges presented by this kind of corpus data, particularly for these pupils, as novice users of the approach, made this impractical. Furthermore, access to English classrooms to conduct research has become increasingly difficult in recent years, and the schools we contacted were all under pressure to provide the daily 'literacy hour' within which our proposed work with subgroups of pupils had to be scheduled. The focus of the corpus-based activities was thus necessarily literacy-related. The research was conducted in two phases in each of two schools, with groups of six children from classes in either Year 4 or Year 5 (i.e. aged between eight and ten) whose levels of literacy made them suitable for participation in the study. Ethical procedures agreed within the University of Reading were adhered to, and the children and their parents gave informed consent to their involvement, as did the class teachers. Detailed recordings were made of six 40-minute sessions with each group in the first phase and of three 50-minute sessions, with some different children, in the second phase.

Our corpus consisted of approximately 800,000 words from 40 texts, including stories, history books, a *Brownie* annual, and so on. To investigate patterns in the English language, we used the concordancing program, WordSmith Tools. The project involved some linguistic analyses of the corpus itself, but its main aim was to explore corpus-based teaching with

young L1 learners. Mindful of the constraints on teachers noted above, the researchers devised activities which complied with educational policies in England, particularly the NLS. We reviewed those objectives prescribed for the two age groups with which we were working (Years 4 and 5), and identified instances where corpus data and analysis might be relevant. We then designed small group teaching sessions which dealt with these objectives. The findings presented here relate to those where the topic of vocabulary is most relevant. Thus, for example, one objective prescribed for Year 5, Term 1 is 'to use adverbs to qualify verbs in writing dialogue, e.g. *timidly, gruffly, excitedly*, using a thesaurus to extend vocabulary'. Five such adverbs were chosen, and from the corpus three concordance lines in which each one occurred were printed out on to a paper worksheet, but with the target adverbs omitted. The children's task was to decide which of the five candidate words best fitted each set of three lines. An example of a worksheet completed by one of the groups is displayed in Figure 3.1.

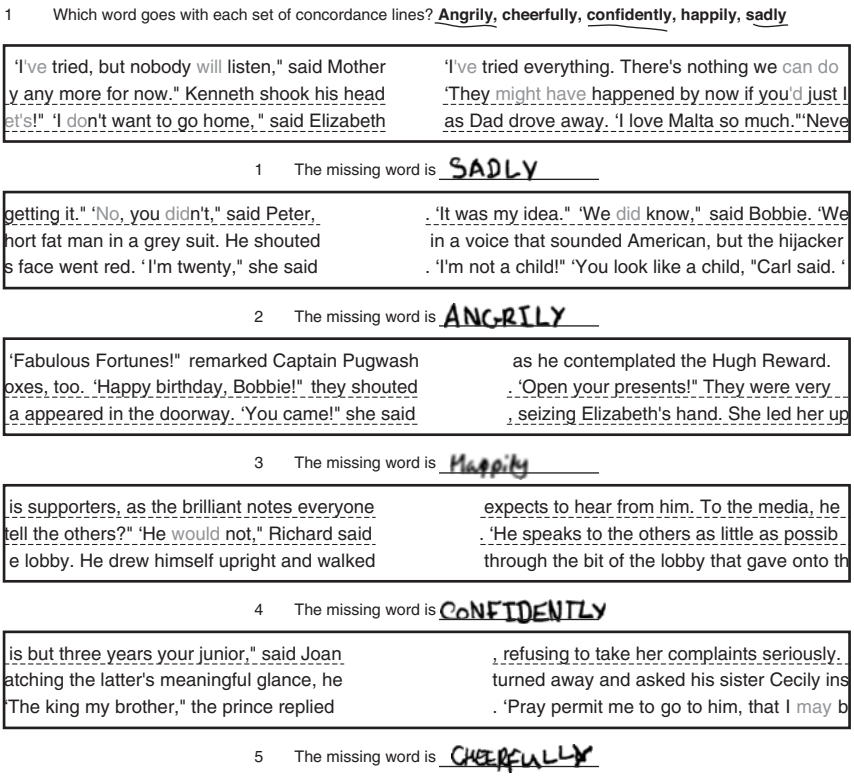


Figure 3.1 Completed worksheet of adverbs modifying quotative verbs, from the CLLIP corpus; originally colour coded, but presented here in greyscale

The investigation was conducted by the two researchers who withdrew participating children to an area suitable for small group work and recording. The recorded data thus consist of the talk generated by activities involving the corpus, concordancing and the interface, including prompted reflection immediately after completing these, and also some summative interviews. Throughout the project, the children were encouraged to reflect on and evaluate their experience of using concordance lines (particularly as paper printouts) and the corpus (particularly on computer), and the written work they produced was also included in the data set.

As will be apparent, this was conceived as an exploratory study, with no attempt made to 'test' the children's metalinguistic knowledge, or to evaluate this approach to teaching in comparison with others. This left us able to explore a wide range of issues as they were raised by the children, rather than focusing on predetermined categories.

Talk about synonyms

When the children were working in groups without the participation of the researcher, they were instructed to discuss and come to agreement about their ideas, although the talk was often much more directed to completing the task – finding the 'right' answer, completing the worksheet – than to discussion of reasons why one answer rather than another was correct. Talk with the researchers demonstrates more clearly how activities based on concordance lines both draw on the children's existing metalinguistic knowledge and provide opportunities for the extension of that knowledge.

The NLS requires children to learn about synonyms at Key Stage 2. As previously noted, one objective is 'to use alternative words and expressions which are more accurate or interesting than the common choices, for example, *got, nice, good, then*' (Y4T2W: Year 4, Term 2, Writing). Another is 'to explain the differences between synonyms, e.g. *angry, irritated, frustrated, upset*; collect, classify and order sets of words to identify shades of meaning' (Y5T1W). An introductory discussion about the activities we had devised to support learning in this area suggested that the children were fairly familiar, from previous teaching, with the relevant concepts. (In the transcripts, speaker identities are made up of a first letter to indicate sex (B = boy, G = girl), a second letter indicating which school they attend (A or B) and a number to distinguish them within the group. Thus 'GB2' is Girl 2 at School B, etc. A filled pause: 'erm'/'um' etc. is transcribed as '#'. AS is the researcher.)

AS:	we're looking at words that mean similar sorts of things but are slightly different in meaning yeah [Name]	1
GB2:	# it's not homophones but it's something like that	2
AS:	synonyms	3

GB3:	synonyms	4
BB1:	synonyms	5
GB2:	synonyms that's it I just put them together and don't know why	6
AS:	yeah these aren't quite synonyms because # they do mean something slightly different from each other some of them but	7
GB1:	which are synonyms again?	8
AS:	words that mean nearly the same thing	9
GB1:	oh yeah they've got	10
BB2:	just like different letters	11
AS:	well they no they they actually are different words	12

B6/67–B6/77 (Extracts, numbered sequentially within each recorded session, are indicated by school, session number and turn number.)

This kind of discussion was fairly frequent in the data, where the children demonstrated familiarity with both concepts and terminology, but were not always confident about which label belonged with which idea. (Similar evidence, of both knowledge and confusion, is present in some of the other discussions, particularly about grammatical categories.) BB2's contribution (l. 11) demonstrates this concept–label confusion, when he offers words with 'different letters' as a gloss for 'synonyms', and GB2 makes the problem explicit as she searches for the label 'synonym', aware that a cognate label 'homophone' is not right, but that she 'just put[s] them together and do[es]n't know why' (l. 6). Perhaps classroom work on vocabulary might encourage children to explore which words 'go together' for them, and develop the idea of semantic networks which accounts for the phenomenon (e.g. Aitchison, 2003).

The discussion continued with further demonstrations of the children's understanding of the concept of synonymy. In the transcripts, 'citations', including reading aloud from text on screen or paper, are indicated by italics, false starts by '-' and incomplete words or utterances by '+':

GB3:	like *huge* and *gigantic*	1
GB2:	like shades of meaning	2
AS:	like shades of meaning very good yeah	3
GB1:	they're spelt the same but they mean different	4
AS:	no no that's the that's the homophones	5
GB2:	that's	6
GB1:	no	7
GB2:	shades of meanings are sim-, are things like #	8
GB1:	*fat* and	9
GB2:	*hot* and *boiling*	10

B6/78–B6/88

GB1 continues to be distracted by spelling (l. 4), but GB3, by means of an example (l. 1), and GB2, who offers a definition (l. 2) and an example (l. 10), establish what is meant by the term in question. GB2's definition, incidentally, is the one used in the NLS and cited above.

The discussion then moves on to the kinds of differences in meaning which synonyms offer, and several lines of thought are now contributed. One idea implicit in the discussion is that synonyms are necessary when a 'better' word is needed. This is indicated in BB1's contribution: '*fat* isn't like a really good word isn't it/and so like *hu-*, *humungous*' [B6/92]. He expands on the idea of *fat* as 'not a good' word, after several turns from other speakers: 'so # *fat* isn't like really describing it but like *humungous* describes it a bit more/*fat* is kind of like saying it's medium-sized/ *humungous* is saying it's massive' [B6/100]. Two problems with *fat* seem to be identified by BB1 here. One is imprecision – it 'isn't … really describing it', while the other is degree – *fat* as 'medium-sized' and *humungous* as 'massive'. This reminds the children of some work done in class, when, as GB2 says, 'our teacher Mrs [Name] gave us two sentences/one was "the chubby baby who lived in the vast desert"' [B6/104], continuing 'and we saw what it was like when we swapped them round', as BB1 illustrates, '"the vast baby the chubby desert" [laughter]'. Prompted by the researcher's question as to what this tells them about synonyms, three children contribute their ideas. BB2 observes, firstly, 'they don't go together/well they they it has to be in the right type of sentence to make sense' [B6/118] and later '"the vast baby" you could get that but then you couldn't get "a chubby desert"' [B6/122]. In between, BB1 offers, 'they do sort of go/come in the same category as the word fat but +' [B6/119]. The first point being made here seems to be about appropriateness (or even, perhaps, collocation, although this term is not used in the NLS and is not apparently in the children's vocabulary). It could also be thought of as relevant to the syntagmatic dimension of linguistic patterns. The second point, which uses the word 'category', is a more paradigmatic organization: *fat, chubby, vast* and *humungous* are all adjectives which can be used to describe nouns such as *baby* or *desert*. A third point is introduced, however, when GB3 contributes ''cause it's a bit rude' [B6/128]. This is a more pragmatic aspect of word choice: it may well be quite accurate to describe someone or something (the non-specific 'it' of turn B6/100) as *fat*, but the word is rejected in this case because of the **social** judgements associated with either the state of being fat or the word – or both.

What does a corpus-based approach have to offer to such discussions? One obvious answer is 'empirical evidence' to use for an exploration of collocations, and for clues, from the attested language of the corpus, to determine which words in a category contribute to 'the right sort of sentence' and 'make sense' in context. For example, how could the children decide whether or not 'a chubby desert' is a likely occurrence in authentic text? Although it might be assumed that native speakers would think

this unlikely, grammatically the construction is possible, and the transcript reveals that, as the 'vast desert/chubby baby' constructions are presented, one pupil (GB1) says 'you probably could get a chubby desert' [B6/124]. Without access to empirical evidence, reliance must be placed on intuition to decide, but, as this conversation shows, intuitions may well differ. Consulting a corpus would be one way of resolving such disagreements – and indeed 'a chubby desert' is not found in the CLLIP corpus. But this is only the start of the useful activities and explorations that having access to a corpus facilitates.

The pupils seemed well aware of the NLS objective which exhorts writers 'to use alternative words … which are more accurate or interesting than the common choices' (Y4T2W), and in this case *fat*, being core vocabulary, would appear to fall into the category of a 'common choice'. In discussion, however, several of the contributions cited indicate that *fat* has some negative connotations for these children, linked variously with its lack of specificity, its unsuitability for describing anything really large, and its 'rudeness'. The sorts of questions that arise, then, include:

- Is *fat* the most frequent of the near synonyms these pupils suggest?
- What sorts of things are generally referred to as *fat* and how big are they?
- Is *fat* typically used negatively? Is it always a 'rude' word?
- Does *fat* have less obvious – less 'common' – meanings?
- Are the alternatives *fat*, *vast* and *chubby* used:
  – to describe the same kind of things and/or other things?
  – in the same way and/or in different ways?

Pupils could be asked to address these questions by consulting the corpus and the sort of evidence they would find is as follows.

In statistical terms, *fat* occurs as an adjective 47 times in the CLLIP corpus of approximately 800,000 words, *vast* occurs 30 times, and *chubby* 4. *Fat*, then, is indeed the most common. *Fat* is used to describe: body parts (3) ('cheeks', 'face' and 'stomach'); food (2), where the connotations are positive, connoting plenty ('berries' and 'drumsticks'); manufactured objects (3) ('hat', 'chequebooks', and a 'club', i.e. a weapon). Eight of the occurrences describe animals, including 'a cow', 'cattle' and 'snails', which demonstrate that the actual gradation of scale seems less salient in the choice of adjective than a relational quality, so that creatures as small as snails can be described as 'too fat to get back into their shells'.

By far the largest category of occurrences in these concordance lines (25) is of *fat* used to describe a human being, however. When used in dialogue, as one character addresses another (3), *fat* is clearly an insult, accompanied by another negatively evaluated adjective, such as 'lazy' and 'horrible'. Descriptions of characters by narrators as *fat* often include other attributes,

which may themselves be negative – 'over-dressed', 'run to seed', 'fierce', but need not be – 'a tall, fat woman', 'this fat wee lass', 'the small one, who was fat and bald'. *Fat* also features as a component of characters' names – 'Fat Frank', 'Fat Iain' and 'the Fat Boy in the Pickwick Papers'. The children's sense that 'fat' may be 'a bit rude' when used to describe people is not misplaced, but a perception on the part of a child character or narrator that an adult is 'fat' may also indicate a wariness of the contrast in size between people of different ages – 'Mrs Bean was fat but not fierce'.

Finally, *fat* occurs five times in this corpus in an idiomatic expression where a synonym is particularly unlikely to be appropriate. The figurative sense of *fat* drawn on is the positive one of plenty, but as it is usually used ironically, it also mobilizes its negative connotations. These expressions are variations on 'a fat lot of good', including 'Fat lot I care!', 'the job he had no fat chance of getting' and so on.

Vast is used to describe some things that are similar to those modified by *fat* (see Figure 3.2). Seven of the entities modified by *vast* are in categories such as food (2) ('a thin slice of the vast cake'); body parts (2) ('a vast pink tongue'); an individual person (2) ('a vast figure'); and people (1) ('the vast crowd'). Three could be classified as abstractions ('vast energy', 'his vast

```
nd disbelief at the sight of a vast creature staring down at her with
           he stopped. Here, a vast decaying tree whose roots were ea
x and Ray Shepherd seemed at a vast distance from them.
,     lying at the bottom of a vast expanse of immaculate lawn and su
      under a shroud of ice, a vast figure began to loom: a he
      still, clear air, hung a vast flag, red, with a gold
 icer scratched furiously at a vast form, a flush, as omin
          she added. She had a vast helping of shepherd's pie,
, and Little Billy could see a vast lake of water, gloriously blue
         avy as an oak limb. A vast pink tongue was hanging out of th
nd was sometimes seen clearing vast hazards on television.
k reluctantly followed Derek's vast bulk until they were behind the
          . My uncle, with his vast professional knowledge of the saleyard
es &mdash; so big they're like vast flapping sails. On windy
ky while the shallows were one vast gleaming mirror. The fort
    s thick as a mooring rope; vast pink tongue hanging out between a
ressed men and women, shedding vast overcoats, ordering wine, food
       partments in the Store. Vast screens and complicated-looking panels c
           . Merging into that vast, sweeping seascape of sky,
       looking rapt, under the vast banner of the G.D.R.
ame back from the walls of the vast, bare concrete building of an
    to cut a thin slice of the vast cake. Then he levered
 the forest, and of course the vast cloud of smoke came hurtling
  Mildred felt very shy as the vast crowd fell silent and every pair
staring out to the west at the vast industrial complex upstream of the
    heaped round the sink, the vast, open view through the window
mell of damp earth, seeing the vast pearly spread of the estuary far
      home of a Maharajah. The vast rooms were crammed with a rich
eginning to get hot, as though vast energy was building up somewhere
```

Figure 3.2 Concordance lines for *vast* in the CLLIP corpus

professional knowledge'), and three more as natural objects ('vast cloud of smoke', 'vast decaying tree').

The next largest group denotes space in some way, consistent with the string, quoted by the children, of 'vast desert'. These six occurrences include 'a vast lake of water' and 'a vast expanse of immaculate lawn', and a connection is noticeable between the entity described as 'vast' and the vantage point from which it is thus perceived ('the vast, open view through the window'). The other lines assigned to this category are '**seemed** at a vast distance **from them**'; 'that vast sweeping **seascape** of sky, waves and sand' (where these natural elements are grouped by perception into a 'seascape'), 'the vast pearly spread of the estuary **far away below him**' (emphasis added in all cases).

This tendency is also apparent in the largest category of instances, classified as 'manufactured objects', which accounts for 10 of the 30 occurrences. Words associated with perception occur close to the entity described as 'vast' in several cases ('**staring** out to', 'complicated-**looking**'), and even where this is not the case the quality of vastness is usually **in relation to** a perceiver. Thus one 'vast' object is a banner above a crowd, and another a form being completed by a 'Duty Officer'.

Like *fatness, vastness* is often a relational quality, relational, that is, to a perceiver, or to other entities to which a particular one is being compared. The 'pink tongue' or the 'overcoats' both described as 'vast' are not as large as a desert, but the adjective is applicable because they seem excessively large to a perceiver in context.

The four occurrences of *chubby* in the CLLIP corpus confirm the intuition that this word is associated with the very young. However, in only one instance is the adjective used of a whole person: 'a chubby six-year-old'. The other three entities described as 'chubby' are body parts: 'face' (2) and 'fingers'.

This example, then, illustrates how applying simple concordance procedures to an appropriate corpus yields a wealth of information which children can use to guide their discussion about words, to inform and encourage the choosing of alternatives and to extend their vocabulary. In particular, the exploration of a corpus allows this last – 'extend their vocabulary' – not just to be in terms of learning more words and their meanings, but also to encompass learning more meanings for words already known. 'Vocabulary extension' then includes becoming more aware of how apparent synonyms are actually used differently, and of the interconnections between the meanings of words and the company they keep.

Obviously, *fat, vast* and *chubby* (along with others such as *stout, portly*, etc.) could form the substance of similar paper exercises to those described above and illustrated in Figure 3.1. Pupils' familiarity and practice with such preset activities can soon lead to explorations beyond one specific corpus. For example, a Google search reveals that the expression 'chubby

desert' is in fact found, albeit in the longer strings 'chubby desert quail', 'chubby desert plant' and 'chubby desert crawler'. Children could be set the challenges first of conducting such a search and then of discussing whether or not what they find is 'fair evidence' that you can have a 'chubby desert'; the educational value of such experience seems obvious.

Conclusion

The aim of this chapter has been to indicate some of the potential for corpus-informed approaches to the teaching and learning of vocabulary in the context of the primary English L1 classroom. The use of a corpus – by teachers to supplement their intuitions about vocabulary and to prepare teaching materials and opportunities, or by pupils, as a resource for a wide range of explorations – is no panacea, and will not oust other kinds of resource and approach. Crucially important are extensive opportunities for children to enjoy whole texts, free from the obligation to mine them for metalinguistic information. Nevertheless, when vocabulary development is a pedagogic goal, the evidence discussed here, from the literature and the CLLIP project, points to some promising areas for development.

Firstly, policy and practice could benefit from a recognition of the normative nature of the discourse often used about the vocabulary of native speakers. This might help to dislodge some commonsense misconceptions about the supposed deficiencies of the spoken language and the assumption that a 'bigger' vocabulary is necessarily 'better' in any straightforward way. When the children we worked with were able to access the corpus directly, they fairly readily explored its features and the opportunity it offered to look for patterns and investigate frequencies. At their suggestion, we constructed small corpora of their own stories, some of whose features they compared with the CLLIP corpus. Such activities bring the empirical investigation of language under learners' control in ways that are more difficult to accomplish when adult intuition – or even reference texts such as dictionaries and thesauri – are the only available sources of evidence.

Secondly, while research into language description throws up ever-increasing evidence of the interrelationship of lexis and grammar, which is in turn incorporated into teaching for learners of additional languages, it would seem perverse to deny this awareness to native speakers in the literacy classroom. Investigations of the 'pattern grammar' and lexical patterns referred to above are well within the scope of primary school children, especially when, unlike ESOL (English for speakers of other languages) learners, their familiarity with English has developed from infancy. The different genres of texts that they encounter and sometimes attempt to produce would be a good starting point from which to develop these learners' sensitivity to patternings associated with contrasting registers, for example,

where the challenge may be associated not only with unfamiliar words, but also with 'unfamiliar combinations of familiar words' (Oakey, 2005).

Finally, evidence from a comparison of the CLLIP corpus with two corpora of writing for adults (Thompson and Sealey, 2007) seems to point to a gradual shift in vocabulary choice from more literal to more figurative uses of frequent items. The concordance line quoted above which makes use of the phrase 'to take someone into one's confidence' suggests maturity on the writer's part, yet it contains no particularly unusual vocabulary. Those responsible for education policy and practice may wish to consider this as our understanding develops of what it is that constitutes progress in the achievement of young learners in respect of their knowledge, use and metalinguistic awareness of vocabulary.

Note

1. The study was conducted jointly by Paul Thompson and myself, and Mike Scott also contributed with modifications to WordSmith suggested by our classroom experiences. We are grateful for the support of the ESRC, who funded the research: 'An investigation into corpus-based learning about language in the primary school' R000223900. I should also like to thank Rosamund Moon, David Oakey, Crayton Walker, Brian Richards and an anonymous reviewer for helpful comments on an earlier draft of this chapter.

4
Vocabulary, Education and Diversity

Viv Edwards
University of Reading, UK

Introduction

In the multilingual classrooms that are a feature of increasing numbers of schools in North America, Western Europe and Australia, vocabulary is seldom the sole focus of attention; more often it occurs as part of broader discussion. This chapter explores the key debates in which vocabulary has played a role. A description of the nature and extent of linguistic diversity forms the backcloth for a review of four issues which affect second language learners and/or speakers of non-standard dialects: the ongoing attempts to explain differences in educational outcomes in terms either of linguistic deficit or difference; the central role of extensive reading in the acquisition of the vocabulary and structures required for understanding and producing academic texts; the ways in which language awareness activities can deepen second language learners' understanding of how language works; and translation issues in providing multilingual resources for children.

Linguistic diversity

Linguistic diversity is not, of course, a new phenomenon. Nation building in Europe over many centuries reduced languages such as Welsh and Basque to minority status (Wright, 2004); the colonial adventures of Europeans achieved a similar fate for indigenous languages such as Maori and Arapaho (Edwards, 2004). In recent times, however, globalization has helped to create unprecedented levels of multilingualism. Cross-border flows of investments, services and ideas are mirrored in a similar transnational flow of languages. The 2001 Canadian census showed that one in every six people reported a home language other than English or French; more than 100 different languages were recorded. The 2001 Australian census reported that, in addition to Aboriginal languages, 142 languages were spoken by just over one in six of the population. According to the 2000 US census, the proportion of people speaking a home language other

than English was even higher: one in five. In the UK, over 300 different languages were recorded in a survey of London schoolchildren (Eversley and Baker, 2000).

Linguistic diversity also encompasses dialect variation. Dialects are usually considered to be a subset of a language. They form continua with only slight modifications from one location to the next. While speakers on adjacent points can understand each other, those at the far ends of the continua may well experience problems of mutual intelligibility. A single dialect continuum often gives rise to several languages and decisions about the boundaries can be contentious. Danish, Swedish and Norwegian, for instance, all belong to the same Nordic dialect continuum; their speakers consider them to be separate languages, mirroring the political autonomy of the three nations. Interestingly, the linguistic distance between these three 'languages' is far less than the differences associated with the 'dialects' of Chinese or Arabic: here, the political imperative is to emphasize commonalities between different peoples. Examples of this kind demonstrate very clearly that decisions about languages and dialects are determined more by politics than linguistics. In the words of a Yiddish saying: 'A language is a dialect with an army and a navy.'

Government policies not only address which languages will be used for official purposes, but which varieties of these languages. Power relationships invariably prevail. In England, standard English was based on the East Midlands variety spoken in the triangle enclosed by the seats of learning, Oxford and Cambridge, and London, the commercial heartland of the nation. In France, standard French is similarly associated with Parisian speech and the seat of power. Standard Modern Greek is based on the variety spoken around Athens. The use of standard varieties protects the interests and influence of a small but powerful elite and marginalizes other groups in the process.

In modern times, issues concerning standardization extend beyond national boundaries. In the UK, for instance, most Greek-speaking children speak the Cypriot variety rather than standard Modern Greek; most children of Italian heritage speak a southern variety; most Bangladeshi children speak Sylheti and not Bengali, the national language; and many Muslim Panjabi-speaking children will choose to read and write in Urdu, the language of religion and high culture in Pakistan, rather than Panjabi, the language of the home. In learning to read and write in minority languages, the issue thus becomes which variety should be used.

Deficit versus difference

One of the recurring debates around linguistic diversity in education over the last four decades centres on whether the underperformance of economically disadvantaged children is best explained in terms of

the social, cognitive and linguistic deficits of certain groups of children (Bereiter, Engelmann, Osborn and Reidford, 1966; Jensen, 1969; Hart and Risley, 1995) – or in terms of differences between the language of the home and school (the difference position). While this debate is seldom framed exclusively in terms of vocabulary, the issue often assumes considerable prominence.

The work of Basil Bernstein (1973) was particularly influential in advancing the deficit position. He attributed the development of two language codes with different grammar and vocabulary to different patterns of socialization. The restricted code with its more limited grammatical possibilities and smaller vocabulary was associated with working-class speakers who were held to make greater use of non-verbal channels of communication. The main function of the elaborated code associated with middle-class speakers, in contrast, was considered to put across relatively explicit meaning verbally. While users of the restricted code were seen as being able to access context-bound, 'particularistic' orders of meaning, it was suggested that the elaborated code was required for the universalistic orders of meaning associated with formal schooling.

The fierce debate which raged in the 1960s and 1970s subsided in the face of trenchant criticism of the deficit position, only to re-emerge in the 1990s with the publication of an equally controversial study by Hart and Risley (1995). Whereas, previously, the focus was on both grammar and vocabulary, these researchers homed in more specifically on the relationship between vocabulary learning and intergenerational poverty. The study is based on the interactions of 42 children in upper- and middle-class, working-class and welfare families from the time they first began to say words at about the age of one until they were about three years old. The analysis draws attention to the larger vocabularies of children from the upper social classes, which the researchers attribute to both the greater volume of talk to which children were exposed and the quality of the interactions. These differences persisted after children started school and were strongly predictive of vocabulary development and reading comprehension in the third grade. As was the case for Bernstein, different language practices are linked to cultural differences: professional parents are seen as preparing their children for participation in a culture which foregrounds symbols and analytic problem solving; the greater use of imperatives and prohibitions in welfare families, in contrast, reflects concerns with established norms.

Critics of the deficit position, however, maintain that all languages and dialects are rule-governed systems capable of expressing the communicative needs of their speakers; and, for this reason, the debate needs to be framed in terms of difference rather than deficit. The arguments marshalled against the deficit position are wide-ranging. In the earlier debate, the lack of evidence is a recurrent theme (Jackson, 1974; Labov, 1973; Trudgill, 1975). The tendency of Bernstein to use examples of speech which were

invented or elicited in artificial test situations stands in marked contrast to the empirical data used by researchers such as Tizard and Hughes (1984). Based on extensive recordings of four-year-old children at home and in school, these researchers were able to demonstrate that differences in language use between working- and middle-class children in the home were very small or absent. The main differences were in fact between the school and home, where conversations between adults and children were longer and more equally balanced; mothers in the study played much more with the children, talked to them much more and answered many more questions than teachers. These findings were therefore a serious challenge to the kinds of compensatory education programmes which language deficit theorists were proposing to improve educational outcomes for children from poor families.

Hart and Risley base their conclusions not on invented data but, like Tizard and Hughes (1984), on recorded interactions; their conclusions are, however, very different. One of the main criticisms levelled in this instance concerns methodology. Dudley-Marling (2007), for instance, draws attention to sampling issues:

> Assertions about the language and culture of the poor, based on a sample of six welfare families, all of whom were Black, and thirteen professional families, twelve of whom were white, puts a Black face on poverty, reinforcing pernicious racial stereotypes. The reality is that only 25% of the 33 million Americans living below the poverty line are Black. Forty-six percent are white (not Hispanic). (US Census Bureau, 2003)

Both earlier and later deficit theorists, however, have been attacked for their class bias. Edwards (1979) provides copious evidence of this tendency in a comprehensive overview of the earlier literature on language and disadvantage. Similarly, Dudley-Marling (2007) draws attention to the middle-class assumptions of Hart and Risley who view the greater use of indirect request forms by professional families as both polite and supportive of problem solving; the greater use of direct request forms by poor families, in contrast, is seen as negative and restrictive. Yet this is an assumption on their part: they make no attempt to support their position by exploring how poor and working-class families actually experience more direct forms of request.

In spite of these criticisms, the influence of both early and later deficit theorists cannot be underestimated. In the UK, the (1975) Bullock Report advocates that health visitors should urge parents to bathe children in language. Herbstein (1980, p. 12) describes a project where speech therapists and social workers distributed 'Mum, talk to me stickers' to mothers with children in supermarkets on the grounds that 'inner city children are simply not being spoken to enough by their parents in their vital

early years'. In the US, attempts of educators and linguists to provide an informed perspective on the confused public debate around a programme which used black English to help African-American children to read and write in standard English were scorned or ignored by the extremely hostile media (Perry and Delpit, 1997). By the same token, the Hart and Risley (1995) study has been used regularly by the popular press to justify early intervention programmes and was recently recommended as essential reading in a report on the teaching of reading for future teachers (Walsh, Glaser and Wilcox, 2006).

The overwhelming weight of opinion in academic circles remains in favour of the difference position. Bourdieu's (1991, 1997) sociology of power has exerted increasing influence over the last decade. He explains patterns of student performance in terms of the uneven distribution of three kinds of 'capital' – economic, social and cultural. *Economic* capital can take the form, for instance, of paying for private schooling. *Social* capital, in the form of membership of a particular club or community, can facilitate access to privileged educational pathways. *Cultural* capital takes the form of favoured 'ways of knowing and being'. Families with the right kind of capital are able to achieve the best outcomes for their children. The greater synergy between the experience of children from middle-class families and the norms and values of the school means that they experience far less disjuncture than children with different values and assumptions.

Corson's (1984, 1985, 1995) study of vocabulary also takes a position consistent with Bourdieu's analysis. He draws attention to highly significant differences in the active use of specialist words of Graeco-Latin origin between working- and middle-class children at 15 years of age, the point at which he suggests Graeco-Latin words predominate in school texts almost to the exclusion of Anglo-Saxon words. He proposes a 'lexical bar' which makes access to the lexis of semantic categories essential for success in education more difficult for users of certain social dialects. Unlike some earlier writers, however, he explains observed differences in terms of difference not deficit, stressing 'how schools value one kind of vocabulary as high status linguistic capital, although that vocabulary is unequally used by people from different sociocultural positions and is not at all regarded as high status cultural capital in many important human contexts' (Corson, 1995, p. 3).

The initial focus for Bourdieu was social class. His framework has, however, been extended to include ethnicity. Blackledge's (2001) study of Bangladeshi women, for instance, shows how mothers with better developed skills in English who are more knowledgeable about British education have easier access to the 'dominant market': they receive more advice from teachers on how to support their children's reading than mothers less fluent in English and less familiar with the system. In acting in this way, the teachers were, of course, in danger of increasing inequality by excluding families in greatest need of support.

The central role of reading

Other attempts to explain different patterns of performance point to the role of reading in this process. Although framed in terms of second language learners rather than speakers of different social dialects, the work of Cummins (1979, 2003) has served as an important catalyst for discussion in this area. Like many of the writers in the deficit–difference debate, he draws attention to the different demands of conversational and academic language. Conversational proficiency develops rapidly over a period of one to two years through face-to-face interaction where there is plenty of contextual support for understanding in the form, for instance, of non-verbal cues. Academic language proficiency, in contrast, is associated with academically demanding subject matter where, typically, there is a great deal less contextual support and acquisition is a much longer process – estimates vary from five to nine years. Children are often offered additional help in school only until they have developed conversational fluency. Yet, in classroom activities, such as synthesis, analysis and evaluation, which demand higher-order thinking skills, the absence of contextual support is likely to place students operating in a second or third language at a disadvantage.

The key issue, however, is how children – irrespective of social class or language background – acquire academic language. The main sources of academic language are, of course, written texts; the most important route to academic language is therefore, logically, through extensive reading. Reading for pleasure has been linked with high levels of reading attainment, writing ability, text comprehension, breadth of vocabulary and greater self-confidence in both first and second language readers (Clark and Rumbold, 2006; Häggsblom, 2006). Identified as more important for children's educational success than the socio-economic status of their family (OECD, 2002), reading for pleasure is seen by some as an important factor in combating social exclusion. Considerable attention has been paid to the reluctance of certain students – especially boys – who *can* read but choose not to (Burgess, McConnell, Propper and Wilson, 2003; Munns et al., 2006). Another issue raised in relation to this group is the question of access to books among poor populations (Krashen, 2007).

Research in second language learning has also demonstrated the strong relationship between the development of vocabulary knowledge and the amount of target language reading undertaken by learners (Dupuy, Tse and Cook, 1996; Elley, 1991; Krashen, 1993; Mason and Krashen, 1997; Postlethwaite and Ross, 1992). Given the weight of evidence, Cummins (1998) explores why greater use is not made of extensive reading and identifies two main issues. The first relates to the difficulty of finding linguistically accessible reading selections which are also interesting and cognitively appropriate; the second concerns the disruption to the flow of meaning caused by needing to look up unknown words.

New technologies offer ways of overcoming these problems: the multimedia 'e-Lective Language Learning' program currently in development seeks to address the problems of low-frequency vocabulary which makes up to as much as 25 per cent of the text by allowing any text in electronic form to be imported and used as authentic input for target language learning, and by supporting readers with a variety of L1 and L2 dictionary and learning strategy supports. Three principles are guiding its development. The first is that access to sufficient comprehensible input in the target language is a necessary condition for language acquisition. The second is that formal second language teaching is relatively unsuccessful for a significant number of learners, primarily as a result of impoverished input in the target language, both with respect to quality and quantity. The third is that the more target language text learners read and comprehend, the more of the target language they learn.

Language awareness

Although the main focus for the vast majority of teachers is on supporting children to acquire the official language, there is growing evidence of a willingness to explore ways of incorporating other languages. Language awareness is one such example, defined by the Association for Language Awareness (undated) as: 'explicit knowledge about language, and conscious perception and sensitivity in language learning, language teaching and language use'. In a classroom context, language awareness involves creating openings for new and culturally diverse experiences which alert children to the role of language in everyday life (Tulaswicz and Zajda, 1998).

There has been a great deal of discussion of implicit knowledge (or knowledge *of* language) and explicit knowledge (or knowledge *about* language) in second language teaching. Ellis (1993, 1994), for instance, proposes that explicit L2 knowledge facilitates implicit L2 knowledge by helping learners notice linguistic properties of the input which might otherwise be overlooked, and by allowing them to make comparisons which confirm or refute hypotheses in their implicit knowledge. The findings of researchers such as Chamot and O'Malley (1994), Fielding and Pearson (1994) and Postlethwaite and Ross (1992) also provide support for this position. Given the potentially important role of explicit knowledge in the development of implicit knowledge, writers such as Fotos and Ellis (1991) point to the usefulness of consciousness-raising (CR) tasks: learners are required to perform an operation on L2 data which will lead to an explicit understanding of the target language.

Whereas the discussion thus far has focused on dialect speakers and the English language learning of speakers of 'new minority' languages (Edwards, 2004), the next example concerns opportunities for learning the lesser-used languages of Europe, most of which are in serious decline. Fabula, a European

Commission-funded project, aimed to develop bilingual, multimedia story books in Basque, Catalan, Frisian, Irish and Welsh which could be used to support language learning. An existing picture book for children was repurposed as a bilingual, multimedia storybook which offered a range of CR tasks involving the comparison of L1 and L2. User testing with children in Welsh primary schools (Edwards, Monaghan and Knight, 2000) revealed children's hypotheses about how languages work and pointed to ways in which teachers could make sensitive and appropriate interventions in the learning process.

In the absence of evidence to the contrary, children understandably assumed that the rules of English applied equally to Welsh. On some occasions, this strategy served them well. Kirsty and her partner Ceri, for instance, tried to identify elements in one text by matching them with equivalents which they recognized in the other. They also drew on the language of classroom instruction – *dwy law* ***ar ei ben*** (hands on heads) – to work out the meaning of a Welsh phrase in the story, *gyda lwmpyn mawr* ***ar ei ben*** (with a big lump on his head). Daniel and Emily also drew on previous experience of going swimming to identify the Welsh word for swimmer in the text. These are positive examples of the children activating their prior knowledge to navigate their way through the text.

Children also used the display of both languages to good effect in identifying cognates. Rachel and Craig were able to identify the Welsh word for coconuts (*coconyts*) 'because it's nearly spelt the same'. There were instances, however, when the children appeared to recognize a cognate but were in fact drawing on a different, unreliable strategy, as in the analysis of the text in Figure 4.1.

Welsh *fwltur* [vYltir] and English *vulture* sound very similar, although visually they are quite different. When asked the meaning of the Welsh *fwltur,* Craig confidently answered 'vulture', suggesting that he recognized the words as cognates. However, on questioning, he explained that he had counted the words and matched them. When asked if he could see any similarities between the two words he said no. This raises interesting questions as to the point at which the written form of cognates become recognizable as such.

In the case of Lee, the idea of word-for-word correspondence was so strong that it overrode pre-existing knowledge and alternative strategies when reading the screen illustrated in Figure 4.2.

She clearly knew that the Welsh word *dyn* meant *man* in English, as she had identified it earlier. Yet she was prepared to jettison this knowledge in favour of a position-matching strategy. Even when encouraged to identify cognates by being asked to consider the surface similarity between English *happy* and Welsh *hapus,* she persisted with a counting strategy. Significantly, when it was pointed out that the match was incorrect, the pair went on to use a process of elimination, tracking which word in the English text co-occurred with *dyn* across different pages, until finally arriving at a correct analysis. In a discussion afterwards, the children were able to recognize that their original strategy

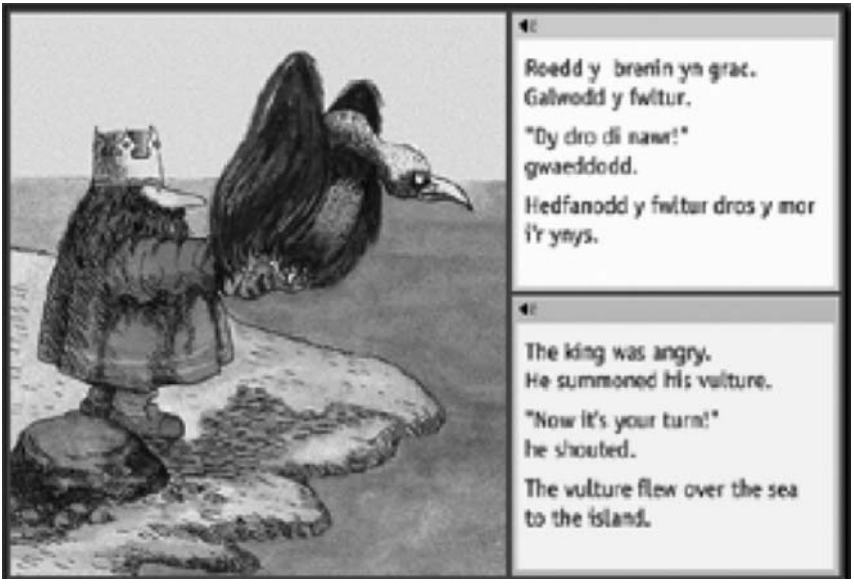


Figure 4.1 Recognizing cognates

Figure 4.2 Word order in different languages

did not work and – crucially from a language awareness perspective – that word order and sentence length can vary between languages.

Daniel used a reverse model of this word-for-word matching strategy, declaring 'Everything in Welsh is backwards.' He then proceeded to count inwards from opposite ends of the English and Welsh sentences until he reached the words he was trying to match. The notion that Welsh was written in reverse order to English was shared by his classmates, Lawrence and Lee, who employed a similar strategy, commenting, 'in Welsh it goes backwards'. A strikingly similar observation was made by classmate Craig who voiced the view that in Welsh, 'They swaps the words around.' This might suggest that some children had overgeneralized an observation made by a teacher on the order of adjectives and nouns in Welsh being the reverse of English.

Comparisons, however, can help correct false assumptions. For example, Lee was convinced that there would always be more words in a Welsh sentence than its English equivalent: 'In Welsh, there's loads of words when you want to say it.' When presented with a counter-example, she was willing to accept that her hypothesis was false.

Translation issues

One of the earliest and most visible responses to the presence of new minority languages in the classroom was the development of dual language picture books for children. Dual language books usually take the form of children's picture books where the illustrations are accompanied by texts in two languages – one 'official', the other a minority language. Writers such as Feuverger (1994) in Canada and the Multilingual Resources for Children Project (MRC) (1995) in the UK have drawn attention to their usefulness in, on the one hand, supporting children's literacy development in two languages and, on the other hand, raising the cultural and linguistic awareness of *all* children. They were welcomed enthusiastically by many monolingual English-speaking teachers as something new and different, though minority language-speaking teachers often felt more cautious, drawing attention to the fact that, as long as there was an English text, bilingual children, for whom English rapidly becomes the dominant language, would have little motivation to read the other language. In addition, there were issues of status: because of difficulties with non-roman scripts before the advent of multilingual desktop publishing, the second language was sometimes handwritten, and looked very much the poor relation alongside the typeset English text. Order was another challenge: which language comes first? In a book where the orientation of one language is left to right and the other right to left, where does the book begin? Designers have also struggled to find inventive solutions to problems such as how to position text in both languages on a page without suggesting that one is more important than

the other. There is evidence, however, that children notice details of this kind, which influence their attitudes towards the languages in question, often negatively (MRC, 1995).

A separate issue of greater relevance for the discussion of vocabulary concerns the variable quality of the translation of texts which are usually written first in English. As we have already seen in the discussion of the Fabula bilingual multimedia stories, the juxtaposition of languages opens up interesting pedagogical possibilities for even very young children to develop their metalinguistic awareness: the fact that word order differs from one language to another; that different languages are sometimes written in different directions; that equivalent words in different languages often bear no obvious relationship to each other. Children are also encouraged to think about the nature of translation: are they aiming for word-for-word meaning or trying to communicate the sense of the first text in composing the second?

Questions of this kind assume, of course, that the quality of the translation is satisfactory. The experience of the MRC project, which scrutinized dual texts in Bengali, Chinese, Gujarati, Panjabi and Urdu, however, would suggest that this is not always the case. Approximately 50 bilingual adults from the linguistic minority communities in question were consulted: experienced teachers who made up the working group that shaped the progress of the project, teachers in the mainstream schools and community schools which hosted the fieldwork for the project, and the teachers and translators who took part in the dissemination of the findings of the project. The aim was to identify and find examples of areas of difficulty and, wherever possible, to suggest possible solutions.

Various issues emerged as we started to look closely at translated books in other languages. The first of these concerned the shortage of professional translators in non-European languages, particularly translators with experience of children's literature. The second centred on what makes a good translation.

When native speakers were presented with a wide selection of translated texts, the consensus was that the quality of translation varied greatly. Many people were irritated by obvious departures from the original text. The *Fox and the Crane* (Hounslow Bilingual Support Project, 1990), for instance, tells how the crane struggles to eat soup from a plate. In the Gujarati translation, the crane struggles to both 'eat soup' and 'drink soup' in the course of the same sentence. On other occasions, the translator has made what appears to be a careless mistake: the Panjabi version of *Soma goes to market* (Cole, 1986) translates, 'He lives with his mother and daddy and little sister Mina in a big *town*' as 'He lives with his mother and daddy and little sister Mina in a big *room*.'

Many inaccuracies of this kind can be explained by the fact that the support structures enjoyed by writers working in English are not always

available for translators. A high quality of proofreading is essential for any publication, irrespective of language. However, translated texts present additional challenges. Monolingual mainstream publishers – like monolingual teachers – often have little sensitivity to multilingual matters. In the absence of an experienced editor and proofreader, the translator is likely to be the sole arbiter of the end product and the only person to check the proofs. Under these circumstances, it is not surprising that serious errors should occur. A more frequent criticism from native speakers is that, whereas translations are grammatically correct, they are stylistically flawed. Discussions of translated texts all too often provoke comments such as 'It doesn't sound good' or 'It sounds disjointed.' Sometimes erudite vocabulary and structures associated with more formal adult texts are used, making the translation far more difficult for children than the corresponding English text. This problem is often related to the sociolinguistic position of minority communities. As already mentioned, many Muslim children speak Panjabi at home but study Urdu as the language of high culture; and most Bangladeshi children speak Sylheti but study Bengali. For this reason, it is not unusual to find that a translator has chosen a word from the standard rather than the everyday language of the child: the use, for instance, of *bhojan* for dinner, rather than the more familiar *roti*, in the Panjabi translation of *school dinners* (Heaslip, 1978).

In books for beginners, the choice of more formal vocabulary of this kind is not too serious: it can sound very formal and stilted, but is unlikely to interfere with a child's understanding of the text, particularly where there are also visual cues. But, in books designed for more experienced readers, the decision to use words outside the experience of young bilingual British readers can interrupt the flow. In Figure 4.3, a page from *Amar's last wish* (Akhtar, 1985), there are four words, expressions or structures in Gujarati which native speakers believe might give rise to difficulties for British-born children.

The use of highly literary and formal language of this kind makes the Gujarati translation far more difficult to read than the original English. Many parents clearly wish their children to learn the standard language, the language of high culture. However, many community language teachers share the view that a more realistic educational goal would be to promote the language of the home as a bridge to the acquisition of the standard or official language (Alladina and Edwards, 1991).

The examples discussed above emerged during a research project undertaken in the mid-1990s. In the intervening years, considerably more expertise has been accumulated in relation to the translation of Chinese and the South Asian languages which formed the focus for this particular project. The 'superdiversity' or 'diversification of diversity' (Vertovec, 2006) which is the hallmark of the twenty-first century will ensure that the same issues will need to be addressed in relation to a range of new minority languages

It was the first day of the holidays. Amar and his parents and Charlie the dog were walking along the beach. Amar and Charlie felt adventurous and ran on ahead of the grownups.
Suddenly Charlie picked up an odd-looking bottle and brought it to Amar. When he pulled out the stopper, a sprite emerged.
Amar was astonished and a bit scared. "Don't be frightened," said the sprite, "I can grant you three wishes."
"Ooh!" said Amar, "My first wish is to have an adventure!"

તે રજાઓનો પહેલો દિવસ હતો. અમર તેના માતા-પિતા અને ચાલી (કૂતરો) દરિયાકિનારા ઉપર ચાલતા હતા. અમર અને ચાલી સાહસિકતા અનુભવવા લાગ્યા અને મોટેરાંઓની આગળ ચાલવા લાગ્યા. અચાનક ચાલીએ કઢંગી દેખાતી એક બોટલ ઉંચકી લીધી અને અમર પાસે લઈ આવ્યો. જ્યારે અમરે બોટલનું ઢાંકણું ખોલી નાંખ્યુ ત્યારે તેમાંથી એક ભૂત ઉપસ્થિત થયું. અમર ભયસહિત આશ્ચર્ય પામી ગયો. "ગભરાઈશ નહી" ભૂતે કહયું, "હું તારી ત્રણ ઈચ્છા પુરી કરીશ." "ઓહ!" અમરે કહયું, "મારી પ્રથમ ઈચ્છા કોઈક સાહસ કરવાની છે."

Figure 4.3 An example of a translation which is more difficult than the English

in major 'receiving countries' such as the US, Canada, the UK and Australia. They also affect countries where indigenous languages are being developed for use in education. There is, for instance, a serious shortage in South Africa, a country with 11 official languages, of translators with experience of children's literature; here, too, the 'endless mistakes' in the small body of material available in African languages have been identified as a serious problem (Edwards, 2008, pp. 20–2).

Conclusion

The various vocabulary-related issues highlighted in this chapter are intended to give a flavour of the challenges facing teachers and children in the linguistically diverse classrooms of North America, Western Europe and Australia. Of these, the longest-standing issue concerns conflicting theories of the role of language in educational achievement. Language deficit theorists attribute the blame for the underperformance of poor families to patterns of family socialization which encourage language practices ill-suited to the needs of formal classroom learning; those who take the difference position attempt to uncover ways in which schools

and teachers systematically privilege certain ways of talking and being over others. Criticism of the deficit position focuses on two main areas. The first is methodology: the work of earlier writers such as Bernstein relied heavily on invented data; the sampling procedures used by later writers like Hart and Risley (1995) raise doubt about the generalizability of their findings. The second area of concern is the bias which leads researchers to interpret language behaviour through the lens of their own predominantly white, middle-class experiences. While the overwhelming weight of academic opinion remains in favour of the difference position, the work of deficit theorists continues to be influential, particularly with the media.

The observable differences between academic language and face-to face communication which served as a catalyst for the deficit–difference debate have implications not only for dialect speakers but also for second language learners. Academic language is strongly associated with the written word; the main route to acquisition is therefore through reading. Extensive reading has been associated with various measures of educational success, including breadth of vocabulary, for both first and second language learners, though the issues for the two groups are sometimes different. In first language speakers, for instance, attention has been paid to motivational issues, such as the reluctance of boys to read for pleasure, and to problems of access to books on the part of children from poorer families. With second language learners, the challenges are to find linguistically accessible and cognitively appropriate reading selections; and the frustration of needing to look up unknown words.

Although education has traditionally been a monolingual habitus (Bourdieu, 1991), there is growing evidence of a willingness on the part of schools and teachers to explore ways of incorporating other languages. One such example is the development of language awareness activities which alert children to the role of language in everyday life and which also fulfil a useful function in language learning. As demonstrated by the example of the Fabula project, L2 knowledge allows children to make comparisons with the L1 which confirm or refute hypotheses about how language works. Consciousness-raising tasks requiring learners to perform an operation on L2 data lead to an explicit understanding of the target language. In the case of the Welsh learners this entailed an awareness that word order is different, that cognates can give important clues to understanding and that meaning is conveyed by different numbers of words in different languages. These understandings are likely to be very helpful in children's future attempts to engage with the Welsh language.

The focus on the use of dual texts in the Fabula project was language awareness. While (paper) dual language books have a similar potential, the opportunities for interactivity are clearly more limited than in the case of electronic resources. For paper books, an issue which has attracted

greater interest is the variable quality of translation. In the early days of dual language publishing, the dearth of authors with experience of writing for children in new minority languages – together with the limited editorial support – often resulted either in mistranslations or a very formal and stilted style which was likely to interfere with young readers' understanding of the text.

Vocabulary-related issues, then, engage the attention of teachers working with several different groups of students: speakers of non-standard dialects; students learning English as an additional language; and children using language awareness activities as part of the study of lesser-used languages. New pedagogies are opening up exciting possibilities, as is the case for bilingual multimedia stories and electronic dictionaries, though it should be noted that developments of this kind require both substantial investment and considerable time to reach successful completion. Finally, some issues are recurrent. Thus while the challenges associated with the minority languages widely spoken since the 1980s have receded as writers and publishers have gained experience, the superdiversity of the new millennium means that similar issues will need to be addressed for some time to come. By the same token, the failure to grasp the essentially political nature of attempts to explain different patterns of educational performance has led researchers to revisit the deficit–difference debate and may well continue to do so.

5

Language Dominance and Lexical Diversity: How Bilinguals and L2 Learners Differ in their Knowledge and Use of French Lexical and Functional Items[1]

Jeanine Treffers-Daller
University of the West of England, Bristol

Introduction

From the psycholinguistic literature we know that monolinguals and bilinguals differ from each other in how they process language and that bilinguals can therefore not be seen as two monolinguals in one person (Grosjean, 1997, p. 167). We also know that perfect bilinguals are extremely rare and that most bilinguals are dominant in one or the other language (Fishman, 1971; Grosjean, 1997; Romaine, 1995). Therefore, there are probably important differences between bilinguals in the command they have of their languages, depending on the frequency with which they use each language, and the purposes for which they need them. As Grosjean (1998) has pointed out, there is a lot of confusion around the concept of bilinguals, and researchers use widely differing operationalizations of this concept. Few researchers attempt to assess the knowledge bilinguals have of either language in any detail, although it is legitimate to question how one can differentiate between different types of bilinguals or between bilinguals and second language learners. Some researchers are reluctant to engage in precise assessments of bilinguals' proficiency profiles because this often leads to negative views of bilinguals or L2 users (see Cook, 1997a on the monolingual bias that is built into second language acquisition (SLA) research). Obtaining precise information about the proficiency of bilinguals is, however, important because language proficiency has an impact on language processing and thus it affects bilinguals' performance on lexical decision tasks or any other tasks that involve informants' language-processing mechanisms.

According to Kroll, Bobb and Wodnieczka (2006, p. 128) we do not yet have a comprehensive overview of how language proficiency and relative

language dominance affect the processes engaged during the planning of spoken utterances, but they point out that this is an important variable that researchers need to take seriously. Many researchers have shown that bilinguals are slower in picture-naming tasks or lexical decision tasks, probably because using two languages has the consequence of lowering the functional frequency of each (Kroll et al., 2006). The bilinguals' disadvantage may, however, disappear if one controls for vocabulary size. Bialystok, Craik and Luk (2008) have recently shown that bilinguals whose lexical knowledge is matched to that of monolinguals outperform monolinguals on a task of letter fluency and word naming, because bilinguals have an advantage over monolinguals in tasks that involve executive control. Their study illustrates the importance of obtaining precise measurements of informants' vocabulary knowledge: instead of reinforcing existing negative views of bilinguals, such measurements can contribute to the discovery of exciting new information about the advantages of being bilingual.

These results also illustrate that it is very important to get a better understanding of the notion of language dominance. Most bilinguals are dominant in one or the other language, but most researchers use the term *language dominance* without providing any measurements of their subjects' knowledge of either language. It therefore remains unclear what language dominance means in linguistic terms, that is to say, whether this mainly affects the lexicon or whether other areas of the language system are also more developed in one language than in the other of the bilinguals under study.

This chapter reports a follow-up to an earlier study in which the language dominance among different groups of Turkish-German bilinguals was investigated, with a particular focus on lexical richness (Daller, Van Hout and Treffers-Daller, 2003). In that study we showed that the proficiency profiles of Turkish-German bilinguals differ significantly from each other depending on whether they lived in Germany or in Turkey. The Turkish-German bilinguals in Germany were clearly dominant in German in that they obtained higher scores on various measures of lexical richness in German but lower scores in Turkish, and the opposite was true for Turkish-dominant bilinguals who had returned to Turkey eight years prior to the recording. Further analyses of the use of Turkish syntactic embeddings among all groups showed that German-dominant bilinguals used simpler syntactic embeddings than Turkish-dominant bilinguals (Treffers-Daller, Özsoy and Van Hout, 2007). These studies demonstrate that it is possible to measure language dominance in bilinguals using different syntactic and lexical variables. The current study aims to contribute further to our understanding of variation in lexical knowledge and use among different groups of bilinguals and how these groups differ from L2 learners in this respect.

For a number of reasons it is particularly important to focus on lexical issues. First of all, because the lexicon plays a central role in the latest versions of generative grammar (e.g. minimalism) and in psycholinguistic models such as Levelt's (1989) speech production model. Most models are lexically driven, that is to say, the grammar, morphology and phonology are determined by the lexical items selected by the speaker. Under this view, vocabulary is the key to learning (Bialystok, 2001, p. 48). Bates and Goodman (1997) even argue that the emergence of grammar depends directly on vocabulary size. Second, psycholinguistic research often focuses on lexical access in production or reception, and much less on syntactic structures. Third, it is reasonable to assume that there is important variability in the number of words individuals (monolinguals) know and the knowledge they have about these words, as lexical knowledge is clearly dependent on a range of sociolinguistic variables, in particular education. Achieving full grammatical competence is normal for individuals, at least in L1 acquisition, but it is difficult to define what full competence means in relation to the lexicon. Monolinguals as well as bilinguals are likely to vary considerably in their knowledge and use of lexical items, but because the latter use their two languages for different purposes, the variability in lexical knowledge among bilinguals is probably even greater than among monolinguals.

Bialystok (2001) summarizes the evidence concerning the existence of variation in lexical knowledge among children but few researchers have attempted to measure variation in adult bilinguals' knowledge and use of lexical items in any detail. In those studies which do consider vocabulary, the focus is most often on receptive knowledge of vocabulary, in particular vocabulary size, as measured with the Peabody Picture Vocabulary Test (Dunn and Dunn, 1959/2006) or the Mill Hill Vocabulary Scale (Raven, 1960). Studies which make use of these tests often show that bilinguals obtain lower scores than comparable monolinguals (Craik and Bialystok, 2006), but we know little about bilinguals' use of vocabulary in productive, more naturalistic tasks.

The aim of the current chapter is to obtain a clearer picture of variability in adult bilinguals' knowledge and use of vocabulary and how they differ from L2 learners. The focus is in particular on lexical diversity as measured with different tools that have recently been proposed in the literature and that are available under CLAN, the computerized data analysis tools developed by MacWhinney and colleagues (MacWhinney, 2000). The main hypothesis of the study is that indices of lexical diversity are excellent tools to measure the lexical proficiency of bilinguals and L2 learners, and to reveal the existence of differences in their use of lexical items. However, only detailed qualitative analyses can reveal the subtle differences in the ways in which Dutch-dominant and French-dominant bilinguals use functional items.

Measuring lexical richness: lexical items and function words

As Nation (2001, p. 27) has shown, vocabulary knowledge is multidimensional and therefore most researchers will agree with Richards and Malvern (2007, p. 82) that no 'single index can represent competence or performance in relation to vocabulary, or for that matter, any other linguistic domain'. Attempting to characterize the vocabulary used by learners with the help of a single measure of lexical richness is therefore necessarily a simplification, and it will be useful to complement this with additional analyses, which can give insights into qualitative aspects of vocabulary knowledge and use.

Previous studies have demonstrated that generic measures such as the index of Guiraud (Guiraud, 1954) and *D* (Malvern and Richards, 1997; Malvern, Richards, Chipere and Durán, 2004) give a good overall impression of the differences in lexical diversity between texts from different sources, including learner language (see Van Hout and Vermeer, 2007 for an overview and a critical discussion of the different measures). These measures do not, however, reveal what the relative contribution of lexical and functional categories is to the lexical diversity of texts. In addition, further analyses need to be carried out if one wants to obtain qualitative information about the lexical knowledge of informants, for example whether they differ in their knowledge of lexical items or function words, or whether there are any particular issues with the ways in which these words are being used. As is well known, learners often overuse particular words or structures that are simpler (Ellis, 1997) or avoid those that they are less familiar with (Schachter, 1974), but the above-mentioned generic measures cannot reveal this.

In order to address those issues, I have carried out analyses of the diversity of lexical categories, in particular nouns and verbs, although adjectives will also be discussed briefly. As nouns and verbs are the main lexical categories in French corpora (Gendner and Adda-Decker, 2002), one might expect that they contribute most to the variability of texts. According to Laudanna and Voghera (2002, p. 8), the frequency of nouns and verbs in English corpora depends on the amount of dialogue and the amount of planning, in that nouns are generally more frequent in monologues and planned texts, while verbs are more frequent in dialogues and spontaneous texts (see also Biber, 1995; Biber, Johansson, Leech, Conrad and Finegan, 1999). It will be interesting to see what the proportion of nouns and verbs is in our French corpus, and what these parts of speech contribute to the lexical diversity of the texts.

After having studied lexical items, we focus our attention on the ways in which learners and bilinguals differ from each other in their use of function words. Relativizers were chosen because their usage is relatively complex: L2 learners of French need to acquire many different forms, some of which (*lequel/laquelle/lesquels/lesquelles*) agree in gender and number with their

antecedent, distinguish between different syntactic functions of these forms and learn how to use them for different purposes in discourse. In addition, relative clauses can be embedded in a variety of ways into sentences, which adds to their complexity.

The literature on the L1 acquisition of French relative clauses is rather limited but the available evidence suggests that subject relative clauses are relatively early acquired and used frequently, but for a limited number of functions (Jisa and Kern, 1998). In addition, Jisa and Kern show that *que* is used much less frequently than *qui* by children as well as adults. In his study of the L2 acquisition of French relativizer morphology, Hawkins (1989) shows that the subject form *qui* is easier than the object form *que* because the former is closer to its extraction site (indicated with a _____) in the examples below, that is, the site from which the WH-word has been moved to COMP, as (1) and (2) illustrate. The form *dont* (which is used for genitive relative clauses) is the most difficult one because it is furthest away from its extraction site (see 3).

1. L'homme *qui* ____ connaît Pierre
   'The man who knows Pierre.' (Hawkins, 1989, p. 162)
2. L'homme *que* Pierre connaît _____
   'The man who Pierre knows.' (Hawkins, 1989, p. 162)
3. Le visiteur *dont* j'avais oublié le nom _____
   'The visitor whose name I had forgotten.' (Hawkins, 1989, p. 163)

It is the relative proximity of the relativizer to its extraction site that explains why first- and second-year students who are studying French for their degree course make more errors with *que* than with *qui* and most errors with *dont* (Hawkins, 1989). These findings form an excellent point of comparison for the use of relativizers by our three groups.

If our hypothesis is correct, measures of lexical diversity should be able to reveal interesting differences between French-dominant and Dutch-dominant bilinguals, as well as between bilinguals and L2 learners. Quantitative analyses cannot, however, uncover more subtle differences between French-dominant and Dutch-dominant bilinguals in their use of functional items. Qualitative analyses are therefore needed if we want to obtain a better understanding of the ways in which bilinguals differ from each other in their use of function words.

Methods

Three groups of subjects participated in the study. The first group consists of 25 adult bilinguals from Brussels who have always lived in Anderlecht, the south-western part of the Brussels agglomeration, in which a relatively large proportion of speakers of Dutch can be found. Participants are all

speakers of Brussels Dutch, the local variety of Dutch, and Brussels French, and some of them, but not all, also speak the standard varieties of either language. From the interviews held with participants in 2006 it is clear that most of these speakers are dominant in Dutch but they use French on a daily basis in everyday life as is common in Brussels which has a predominantly French-speaking population.

The second group are 25 eighteen-year-old Flemish students of French from Aalst, who were recorded by a team of researchers led by Housen in the framework of a project on the simultaneous learning of two foreign languages (French and English). The data for this project are available on the website of the French Learner Language Oral Corpora (FLLOC): http://www.flloc.soton.ac.uk/brussels.php.

The third group consists of French–English bilingual students from a business school in Paris, who grew up with French only but learnt English (and other languages) at secondary school. One student indicated to have spoken Spanish in addition to French in early childhood. These students were taught in Paris through the medium of English and they were enrolled in an English course at the University of the West of England, Bristol, in 2006 when they took part in this study. This group is clearly French-dominant, as is obvious from their language history, even though they use English on a daily basis for all subjects of their studies.

A controlled productive task was chosen rather than a free productive task to ensure the comparability of the content across the three groups, which is particularly important in studies which focus on lexical items. Mayer's (1969) storybook *Frog where are you?* was used to elicit semi-spontaneous speech from all individuals. This story has frequently been used to study language use of monolinguals and bilinguals (e.g. Berman and Slobin, 1994), which makes it relatively easy to obtain comparable data sets, such as the Brussels corpus on the FLLOC database. Because the Brussels bilinguals regularly use French in conversation but are not necessarily biliterate, written language tests were not considered appropriate for the target group. All informants were given some preparation time before telling their story individually to the researcher, either in their own home (the participants from Brussels) or in the school/university they attended. The bilinguals from Brussels also told another Frog story (*Frog goes to dinner*: Mayer, 1974) in Dutch. The Parisian students told this story in English, but these stories are not being analysed for this chapter, which focuses on variation in French. Some Flemish students were offered help by their interlocutor if they did not know a particular word, but all words that students learned from the researchers were discarded from the analysis. Two students who received a disproportionate amount of feedback from their interlocutor were excluded from the study altogether.

All data were transcribed in CHAT format (MacWhinney, 2000), and subsequently a morphosyntactic coding tier (the mor tier) was added to the

transcripts with the help of the MOR and POST commands under CLAN. Any remaining ambiguities, errors or inconsistencies in the resulting mor tier were corrected by hand. In addition, all proper names, filled pauses and other hesitation markers, exclamations as well as words from other languages (mainly Dutch or English) were excluded from the analysis.

For several reasons, using the mor tier for analyses of lexical richness is particularly useful. First, because this tier makes it possible to distinguish between homophones (e.g. *tu* 'you' as a personal pronoun and *tu* 'was silent' as the past participle of the verb *se taire* 'to be silent') which is only possible on the main tier by adding disambiguation codes by hand. Second, on the mor tier stems and inflections on nouns, verbs and adjectives are coded separately, which considerably simplifies the task of making calculations on lemmatized data sets (see Appendix 5.1 for an example of a transcript with a mor tier). Thus, the coding on the mor tier makes it possible to consider all inflected forms of the same verb, noun or adjective as one type, which is particularly important in a language such as French in which lexical types can take a wide variety of inflections. The different forms of function words, such as the definite article (*le*, *la*, *les*) and the possessives (*son*, *sa*, *ses*) are however coded as different types on the mor tier. The total number of French types in the current data set was reduced from 11,125 to 8981, that is a reduction of 19.3 per cent.

In a previous study, we lemmatized the data on the main tier in a way that is described in detail in Tidball and Treffers-Daller (2007), but if all researchers who work on French lemmatize their data in slightly different ways, this reduces the comparability of results significantly (see also the discussion in Richards and Malvern, 2007, on the effects of different lemmatization strategies on their measure *D*). The mor tier offers a standard that can be used by everyone. Third, new switches that can be used with the frequency command FREQ have recently become available under CLAN. These make type/token analyses of individual syntactic categories on the mor tier possible, which is extremely useful for studies of lexical richness.

The only problem encountered using the mor tier in CLAN is that the French mor tier distinguishes different subcategories of verb forms (infinitival, participial, progressive and other forms) in the information to the left of the pipe separator (|) which separates the syntactic category information from the word itself. Thus, the verb *trouver* 'to find' is categorized on the mor tier in many different ways: v:pp|trouver; v:inf|trouver; v:prog|trouver and v|trouver (see Appendix 5.1 for examples). This means that FREQ counts these different forms of *trouver* as different types rather than as different tokens of the type *trouver*, which results in inflated indices of lexical richness. This problem also exists, but to a lesser extent, for other syntactic categories such as pronouns. Using switches such as +s"*-% %", which tell CLAN to ignore form variants, does not solve the problem, because these switches only look at information *after* the pipe separator,

not before. For the purposes of this chapter I therefore decided to erase the above-mentioned subcategories of verbs with the help of the change string command (CHSTRING), leaving only the codes for subcategories of modal verbs (v:mdl| and v:mdllex) and auxiliairies (v:aux) in place, as the distinction between lexical uses of verbs such as *avoir* (*il a un livre* 'he has a book') and auxiliary uses of this verb (*il a acheté un livre* 'he bought a book') are obviously important for analyses of lexical richness. If the information about different verb forms could be coded after the pipe separator on the French mor tier, together with other morphological information regarding person and tense, this problem would not exist.

Results

Generic measures of lexical richness: the index of Guiraud and *D*

In the first instance, the differences between the three groups were investigated by calculating two generic measures of lexical richness, the index of Guiraud and *D*. As Table 5.1 shows, there are significant differences between the groups for both measures, in that the French-dominant bilinguals obtain the highest scores and the L2 learners the lowest scores, whereas the scores of the Dutch-dominant bilinguals fall between those of the other two groups. The results of the ANOVA and Tukey post hoc tests show that all groups are significantly different from each other for the index of Guiraud (ANOVA, F (2, 64) = 50.58, $p < .001$) as well as for D (ANOVA, F (2, 64) = 56.9, $p < .001$), but D is a bit more powerful in that it discriminates slightly better between the groups, as can be seen from the eta squared values. Group 3 also produces significantly fewer types and tokens than Groups 1 and 2, but Groups 1 and 2 do not differ significantly from each other in their use of tokens, and only marginally in their use of types. Therefore more sensitive measures such as D or the index of Guiraud are needed to demonstrate the existence of differences between the groups. Both measures correlate very strongly and significantly with each other (r = .951; N = 69; $p < .01$), which gives a clear indication that they are measuring similar aspects of lexical richness.

Table 5.1 Descriptive statistics for tokens, types, *D* and Guiraud for each group, and effect sizes for differences between groups (η^2)

Group	*n*	*Tokens*	*Types*	*D (SD)*	*Guiraud (SD)*
1. Business students, Paris	19	571	175	68.7 (13.5)	7.4 (.85)
2. Bilinguals from Brussels	25	500	143	50.3 (15.3)	6.5 (1.0)
3. Flemish L2 learners	25	283	83	28.7 (7.3)	4.9 (.58)
Total (mean scores)	69	441	130	47.5 (20.2)	6.2 (1.3)
η^2				.633	.612

The *D*-values in Table 5.1 are relatively high in comparison with those reported for French in Tidball and Treffers-Daller (2007), in which first-year students obtained mean scores of 18.78, final-year students 26.46 and French native speakers from the same Parisian business school obtained mean values of 34.87 for oral descriptions of cartoon strips. The results for the index of Guiraud are, however, only marginally higher than in our previous study in which the two student groups and the Parisian business students obtained scores of, respectively, 4.30, 5.25 and 6.27. There are several potential explanations for these differences, but it is most likely that the main reason for the differences between the two studies should be sought in the fact that different elicitation materials were used. It is possible that the relatively complex storyline of the frog story invites informants to produce more detailed narratives than the father-and-son comic strips used in the earlier research. In the former there is a wide range of activities involving many different participants, whereas the latter revolves around a small number of actions involving two protagonists with one or two additional characters. Evidence for this explanation can be found in the relatively large number of types (130) and tokens (441) the subjects in the current study produced, in comparison with the students and the native speakers in Tidball and Treffers-Daller (2007), who produced 97 types and 327 tokens on average in the father-and-son storytelling task. Second, lemmatization was done on the main tier in a slightly different way in our previous study, whereas the mor tier was used for this purpose in the current study. The mor tier distinguishes between different uses of function words such as *qui* 'who', which can either be an interrogative pronoun or a relativizer. The same applies to function words such as *le/la/les* 'the/him/it/her/them', which function not only as determiners but also as object pronouns. CLAN programs consider the different uses of these words as different types, which results in slightly higher *D*-values and slightly higher scores on the index of Guiraud, if these measures are calculated on the mor tier.

Given the differences in the elicitation task and the lemmatization issues mentioned above, it is remarkable that the values of the index of Guiraud are relatively similar in both studies. This could be an indication that this measure is slightly more robust in that it is less sensitive to task effects or lemmatization strategies. A comparison of absolute *D*-values or scores on the index of Guiraud remains, however, very difficult if the elicitation materials are not the same across studies and if there is no standard way to lemmatize French (see also David, 2008, who makes a similar point). Using the mor tier for measurements of lexical richness could offer a solution to the latter problem.

Figure 5.1 illustrates why the standard deviations (given in Table 5.1) are higher for the bilingual group than for the other two groups even though the interquartile range (that is the distance between the lower and the upper quartile) is higher for the Paris group. The higher standard

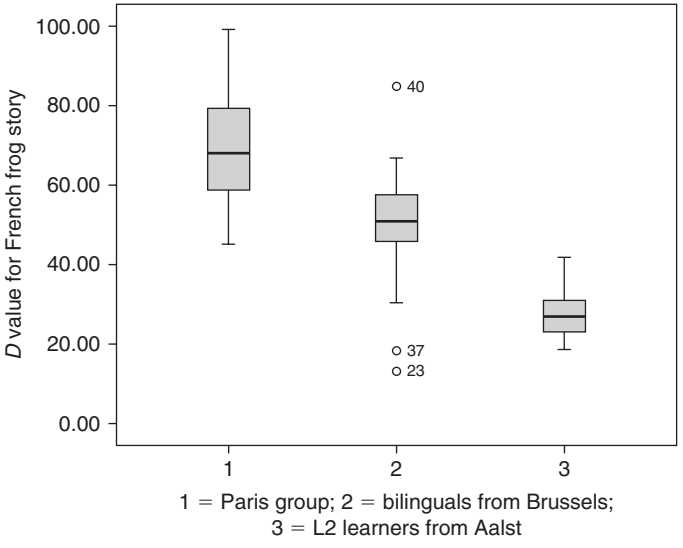


Figure 5.1 *D* scores for all three groups

deviations in the Brussels group are probably due to the exceptional scores of the three outliers. The existence of outliers is, however, to be expected as bilinguals inevitably vary considerably in the amount of use they make of their two languages, with some using French on a daily basis for a range of purposes whereas others make use of French in much more limited ways. The presence of three outliers in the bilingual group forms a good illustration of the variability in vocabulary knowledge and use among bilinguals.

Lexical diversity of nouns and verbs

In the second part of this study the focus is on lexical diversity in two lexical categories and one functional category. While it would have been interesting to compare the results for *D* and the index of Guiraud in this part of the study too, this turned out to be impossible because *D* can only be calculated if a minimum of 50 tokens is available. Hardly any of the L2 learners from Aalst and only half of the bilinguals from Brussels produced a sufficient number of verb tokens. A similar problem exists for the nouns and the relativizers (of which no informant produces more than 25 tokens). For this reason, *D* and its derivative the Limiting Relative Diversity index (Malvern et al., 2004, pp. 147–51) could not be calculated for these individual syntactic categories, and only the index of Guiraud was used.

With the help of new switches under CLAN, which are available with FREQ but not with VOCD, it is possible to obtain lemmatized frequency

Table 5.2 Mean scores for noun types, noun tokens and the index of Guiraud

Group	*Noun types*	*Noun tokens*	*Guiraud nouns 1 (noun types/ √ noun tokens)*	*Guiraud nouns 2 (noun types/ √all tokens)*
1. Business students, Paris	49.8	104.2	4.93	2.1
2. Bilinguals from Brussels	38.8	98.1	3.93	1.75
3. Flemish L2 learners	20.8	52.7	2.86	1.23
F-value (2, 66)	–	–	31.0 ($p < .001$)	30.3 ($p < .001$)
η^2	–	–	.484	.479

(All groups differ significantly from each other: Tukey post hoc analysis, $p < .01$.)

Table 5.3 Mean scores for verb types, verb tokens and the index of Guiraud

Group	*Verb types*	*Verb tokens*	*Guiraud verbs 1 (verb types/√ verb tokens)*	*Guiraud verbs 2 (verb types/√ all tokens)*
1. Business students, Paris	42.2	70.6	4.98	1.77
2. Bilinguals from Brussels	31.9	58.9	4.12	1.42
3. Flemish L2 learners	16.6	29.7	3.04	.98
F-value	(2,66)		30.8 ($p < .001$)	35.54 ($p < .001$)
η^2			.483	.510

(All groups significantly different: Tukey post hoc analysis, $p < .01$.)

lists per syntactic category (see Appendix 5.2 for an example). This allowed us first of all to establish that there are more nouns (2045 types and 4662 tokens) than verbs (1677 types and 2954 tokens) in the current corpus, which provides evidence for Laudanna and Voghera's (2002) claim that nouns are generally more frequent than verbs in monologues and planned texts. On the basis of the output of FREQ, the diversity of nouns and lexical verbs was calculated in two different ways: first the ratio of noun types over the square root of noun tokens (Guiraud nouns 1), and then the ratio of noun types over the square root of all tokens (Guiraud nouns 2). The same procedure was followed for the verbs. The two calculations of Guiraud differ only marginally from each other but the second may be preferable, as the same denominator is used for all calculations (nouns and verbs).

The three groups differ in predictable ways from each other in their use of nouns as well as verbs: the business students from Paris obtain the highest scores and the L2 learners the lowest, and the scores of the bilinguals from Brussels fall in between those two (see Tables 5.2 and 5.3). It is

interesting to see that eta squared for the verbs is higher than for nouns, which is an indication that the diversity of verbs as measured with Guiraud discriminates between the groups to a greater extent than the same measure for nouns. In order to find out whether verbs contribute more to the diversity of the texts than nouns, a paired *t*-test was carried out on the pooled data in which the mean values for Guiraud nouns 1 and Guiraud verbs 1 were compared. The differences between the mean Guiraud for the verbs (3.96) and the mean Guiraud for the nouns (3.82) are significant in the predicted direction with a one-sided *t*-test ($t = 1.7$; df = 68, $p < .05$). Thus, verbs may indeed contribute somewhat more to the diversity of the texts than nouns in this data set.

If noun and verb types are counted together, the calculation of Guiraud (verb types + noun types/$\sqrt{}$all tokens) discriminates even better between the groups (ANOVA, F (2, 66) = 41.2, $p < .001$; $\eta^2 = .555$). This result can be improved only slightly by adding adjective types to the calculation (ANOVA, F (2,66) = 42.1, $p < .001$; $\eta^2 = .560$). As these effect sizes are very close to those obtained by *D* (.633) and the index of Guiraud (.613), which are based on *all* types and tokens, words belonging to categories other than nouns or verbs contribute probably very little to the between-group differences.

Differences in the frequency of relativizers

As there are very few different relativizers (*qui*, *que*, *dont*, *où* and *lequel/laquelle/lesquels/lesquelles*), calculating the index of Guiraud for relativizers is not very meaningful because informants differ very little in the types they use. The number of tokens does, however, vary considerably per individual. Therefore a calculation of the proportion of all tokens that are relativizers can give interesting information about the differences between the groups.

As Table 5.4 shows, there are no significant differences between the bilinguals from Paris and the bilinguals from Brussels in their use of relativizers, but the L2 learners use significantly fewer relativizers than the two other groups. Contrary to expectations, Dutch-dominant bilinguals from Brussels obtained slightly higher scores than the French-dominant bilinguals from Paris. Even though this result was not significant and thus not generalizable to the wider population, I wanted to explore this finding in this particular sample, to see if there was any indication of overuse of particular structures by the bilinguals from Brussels.

This analysis revealed that the bilinguals in Brussels use the relativizer *qui* 'who/which' very frequently in combination with *il y a* 'there is', in utterances such as (4), whereas the other groups do not do this:

4. Allez et en une fois y a un hibou qui sort
   'Well and all of a sudden there is an owl which comes out.' (Bilingual informant JEA from Brussels)

Table 5.4 Percentage of relativizers in all three groups

Groups	*(All relativizer tokens/ all tokens) × 100 (SD)*	*(All relativizer tokens not triggered by 'il y a'/ all tokens) × 100 (SD)*
1. Business students, Paris	1.18 (.50)	1.18 (.49)
2. Bilinguals from Brussels	1.49 (.95)	1.10 (.84)
3. Flemish L2 learners	.43 (.49)	.41 (.49)
F-values	F (2, 66) = 15.4 ($p < .001$)	F (2, 66) = 9.97 ($p <. 001$)
Tukey post hoc analyses		
1 and 2	ns	ns
1 and 3	*	*
2 and 3	*	*

With the help of the COMBO[2] command under CLAN it was established that among the Brussels group, 57 of the 184 uses of *qui* as a relativizer occur in structures such as (4). The students from Paris, however, used the structure *il y a un X qui Verb* only once (out of 120 uses of *qui* as a relativizer) and the Flemish L2 learners employed it only once out of 43 uses of the relativizer *qui*. If the relativizers which are triggered by the occurrence of *il y a* are excluded from the calculations, the unexpectedly high frequency of relativizers disappears (see Table 5.4, final column). This does not, however, affect the overall results: the differences between the two groups of bilinguals in their use of relativizers are not significant.

Guillot's (2005) detailed comparative analyses of this structure across a range of written and oral sources can help to throw new light on its frequency in the data. Guillot shows that the occurrence of the prefabricated formula *il y a NP relative clause* is not only frequent in L2 learners' spoken and written language but also in unplanned native speaker speech and it is thus not an indication of non-nativeness (Guillot, 2005, p. 120). The fact that the L2 learners in the current study were not exposed as much to spoken French as the bilinguals from Brussels can probably explain why they did not use this structure frequently. The students from Paris, however, who were in daily contact with French, did not use this structure frequently either, which is somewhat puzzling. Jisa and Kern's (1998) analysis of the functions of relative clauses can help to throw light on this issue. They show that children use relative clauses more for general discourse functions (mainly to establish and introduce new referents) whereas adults use these for a much wider variety of functions. Although a detailed analysis of the functions for which the bilinguals from Brussels use relative

Table 5.5 Frequency of each relativizer in each of the three groups

	Qui (subject) [a]	*Que*	*Où*	*Dont*	*Lequel (+ form variants)*	*Total (100%)*
1. Business students from Paris	120 (80%)	18 (12%)	3 (2%)	2 (1.3%)	7 (4.7%)	150
2. Bilinguals from Brussels	184 (82.5%)	18 (8.1%)	17 (7.6%)	2 (0.9%)	2 (0.9%)	223
3 L2 learners from Aalst	43 (100%)	0	0	0	0	43

[a] There were no occurrences of oblique uses of *qui* (i.e. *qui* following a preposition) in the data.

clauses is beyond the scope of this chapter, bilinguals frequently use relative clauses to introduce new referents, as example (4) illustrates. This usage is very similar to the examples discussed in Jisa and Kern (1998). French-dominant bilinguals, however, hardly make use of this strategy to introduce new referents.

Table 5.5 gives further details of the qualitative differences in the uses of relativizers by the three groups. The L2 learners use only the subject relativizer *qui*, but the two other groups also use the object pronoun *que* and a small number of other relativizers. As the two main types of relativizer are used in roughly the same proportion, this is another indication that the groups from Paris and from Brussels do not differ significantly from each other on this point, but the stories of the L2 learners display less diversity on this variable. The data thus confirm the findings of Hawkins (1989) and Jisa and Kern (1998) that subject relative clauses are the most common (and probably the easiest), followed by object relative clauses, whilst other types are less frequent.

Conclusion

In this chapter we have seen that there are important differences in the lexical diversity of stories told by bilinguals and L2 learners, and that *D* and the index of Guiraud are excellent tools in demonstrating the existence of those differences. *D* proved to be somewhat more powerful than the index of Guiraud, in that the former discriminated more strongly between the groups than the latter.

As one of the aims of the study was to find out which syntactic categories contribute most to the diversity of the stories, separate analyses were carried out of the diversity of two lexical categories (nouns and verbs) and one functional category (relativizers) with the help of tools that have recently

become available under CLAN. The index of Guiraud was employed for the analysis of nouns and verbs, because *D* could not be used for reasons explained above. As nouns and verbs are the word categories which have most members it is not surprising that we found that these two categories contribute most to the total between-groups variance in the data. The eta squared values obtained for analyses based on nouns and verbs approached those based on all the words in the stories. Adding adjectives to the computation contributed very little to this result.

There were also significant differences between the L2 learners and the bilinguals in their use of relativizers, in that the L2 learners used fewer and a more limited range (only subject relativizers) than the bilinguals. Although there were no significant *quantitative* differences between Dutch-dominant and French-dominant bilinguals in their use of relativizers, a detailed *qualitative* analysis demonstrated that the Dutch-dominant group overused relativizers in prefabricated formulae to introduce new referents in the story. These subtle differences in the bilinguals' use of functional items could not be revealed with the help of generic measures of lexical diversity.

The main differences between the Dutch-dominant bilinguals from Brussels and the French-dominant bilinguals from Paris resided however in the diversity of the lexical items they used, in particular nouns and verbs, and not in differences in their use of the functional items studied here. The L2 learners in our study, on the other hand, differed significantly from the two groups of bilinguals in the diversity of lexical as well as functional items they used in the stories.

It is of course possible that language dominance manifests itself in some bilinguals in their use of lexical as well as functional items (see for example Treffers-Daller et al., 2007). Therefore we need further insight into the ways in which bilingual competence can vary in individuals, in other words, we need a typology of bilingual competence and an operationalization of the notion of language dominance in terms of the different language levels. The main contribution of the current study to our understanding of these issues is perhaps that it has shown that key aspects of language dominance can be measured with the help of indices of lexical diversity. These need to be complemented, however, with qualitative analyses of the ways in which functional items are being used if one wants to reach an in-depth understanding of language dominance.

Notes

1. I am very grateful to Alex Housen for making the L2 learner data from Aalst available on the FLLOC database, to Xu Ziyan for collecting the data from the Paris group, to John and Françoise Tidball for transcribing the data of the Paris group, to the Research Committee of Faculty of HLSS for sponsoring my sabbatical leave

and the costs of the data collection in Brussels in 2006, to the Research Fund of School of LLAS for sponsoring the transcription of the Paris data set, to Florence Myles and Annabelle David for giving me the French POST programme and to Michael H. Daller and Brian Richards for comments on earlier versions and advice on statistical issues.

2. The command used was: combo +s"y^*^qui", which tells CLAN to look for an occurrence of *y*, followed immediately or eventually by *qui*. The output then needs to be checked to see whether or not these occur within the same clause.

Appendix 5.1

Excerpt of a transcript of the frog story as told by one of the bilingual informants from Brussels:

@Begin
@Languages: fr
@Participants: DEM 003 Informant, JTD Jeanine Investigator
@ID: fr|AND|DEM|||||Informant||
@ID: fr|AND|JTD|||||Investigator||
@Date: 06-APR-2006
@Coder: JTD
*DEM: ça c' est le garçon qui avec son chien regarde le la grenouille dans un bocal .
%mor: pro:dem|ça pro|ce/ces&SING v:exist|être&PRES&3SV det|le&MASC&SING n|garçon&_MASC pro:rel|qui prep|avec det:poss|son&MASC&SING n|chien&_MASC v|regarder-PRES&_3SV det|le&MASC&SING det|la&FEM&SING n|grenouille&_FEM prep|dans det|un&MASC&SING n|bocal&_MASC&_SING .
*DEM: ici le garçon est en train de dormir et son pantalon reste droit et la grenouille sort du bocal et se dirige vers les pantoufles .
%mor: adv:place|ici det|le&MASC&SING n|garçon&_MASC v:exist|être&PRES&3SV prep:art|en n|train&_MASC prep|de v:inf|dormir&INTRANS conj|et det:poss|son&MASC&SING n|pantalon&_MASC v|rester-PRES&_3SV adj|droit&MASC conj|et det|la&FEM&SING n|grenouille&_FEM v|sortir&PRES&3SV det|du&MASC&SING n|bocal&_MASC&_SING conj|et pro:refl|se&3SP v|diriger-PRES&_3SV prep|vers det|les&PL n|pantoufle&_FEM-_PL.
*DEM: &oh dit le garçon et son chien où est passé notre grenouille ?
%mor: v:pp|dire&_MASC&_SING det|le&MASC&SING n|garçon&_MASC conj|et det:poss|son&MASC&SING n|chien&_MASC pro:int|où v:aux|être&PRES&3SV v:pp|passer&_MASC&_SING det:poss|notre&_SING n|grenouille&_FEM?

(transcript continues)

Appendix 5.2

Example of command used to extract verbs from the transcript of one of the students of the Business School in Paris:

```
freq +t%mor -t* +s@r-*,|-v,o-% +f
Thu Aug 07 15:22:45 2008
freq (09-Jul-2008) is conducting analyses on:
 ONLY dependent tiers matching: %MOR;
From file <c:\DOCUMENTS AND SETTINGS\
J-TREFFERSDALLER\MY DOCUMENTS\JEANINE\VOCABULARY
STUDIES IN L1 AND L2 ACQUISITION\FINAL POST 18 JULY 08
NEGOCIA\NEGOCIA VERBS SIMPLIFIED\F03.mor.pst.str.str.cex> to
file <C:\DOCUMENTS AND SETTINGS\J-TREFFERSDALLER\MY
DOCUMENTS\JEANINE\vocabulary studies in L1 and L2
acquisition\final post 18 July 08 Negocia\Negocia verbs simplified\verb
freq\F03.mor.pst.str.str.frq.cex>
****************************************
3 v|amuser                  1 v|profiter
1 v|apercevoir              1 v|préoccuper
2 v|appeler                 1 v|préparer
1 v|arriver                 1 v|rechercher
1 v|assurer                 3 v|regarder
4 v|attaquer                2 v|repartir
2 v|attraper                1 v|retenir
1 v|avertir                 2 v|revenir
1 v|blesser                 1 v|réussir
1 v|cacher                  1 v|réveiller
1 v|contrarier              1 v|sortir
1 v|coucher                 2 v|soulever
2 v|courir                  4 v|tenter
1 v|dire                    4 v|tomber
1 v|disparaître             4 v|voir
2 v|dormir                  1 v|échapper
2 v|découvrir               2 v|éviter
1 v|entourer                ------------------------------
2 v|essayer                 43 Total number of
2 v|grimper                 different word types used
1 v|immiscer                71 Total number of
1 v|manquer                 words (tokens)
1 v|parler                  0.606 Type/Token ratio
1 v|partager
1 v|perdre
2 v|prendre
```

6

Lexical Diversity and Native-Like Selection: the Bonus of Studying Abroad[1]

Pauline Foster
St Mary's University College, UK

Part of a child's task in becoming a competent language user is to gain control of a system of rules (i.e. a grammar) which has prodigious generative power, allowing the creation of a potentially infinite number of utterances, and conferring the ability to distinguish word combinations that are grammatical from those which are not. This generative model of language (Chomsky, 1965) has been widely accepted for some time although, as has been pointed out by many (e.g. Becker, 1975; Pawley and Syder, 1983; Widdowson, 1989) it cannot fully account for language use. No first language speaker uses the creative power of grammar to anything remotely like its full extent, and out of the infinite variety of utterances which could be generated by a grammar, only a limited variety are attested in speech. As extensive corpus analysis shows, samples of authentic language are characterized not by infinite originality, but by the pervasive recycling of common word combinations (Sinclair, 1991). These can be simple collocates (e.g. *Happy Birthday! worst-case scenario, fish and chips*), whole utterances (e.g. *It's not beyond the realms of possibility, I wouldn't worry about it if I were you*) or complex syntactic frameworks needing a few additions (NP *is the sort of person who goes around* V-*ing* NP). Such fixed or partially fixed word combinations are named 'native-like selections' by Pawley and Syder (1983). They are familiar to native speakers because they have been encountered many times before, whereas their paraphrased grammatical equivalents sound odd because they have not been encountered before: *Enjoyable Birthday! chips and fish, window-breaking while going around is done by this sort of person.*

The fact that all members of a speech community are apt to select the same word combinations indicates these are an important part of every speaker's acquired knowledge. As Partington (1998, p. 16) observes, this knowledge is both active and passive; a native speaker's communicative competence includes not just the ability to produce native-like selections, but also the ability to detect unnative-like ones.

The question of how speakers get to know which combinations of words are allowed and which are not, must therefore be concerned not just with

what is grammatical, but also with what is natural. While a Chomskyan approach to language acquisition accounts only for the former (the child unlocks the grammatical structures of its first language(s) by exploiting an innate universal grammar), emergentist and connectionist approaches can account for both grammaticality and native-likeness. In these approaches children are described as discovering the grammatical rules of their first language through recognizing the pervasive patterns in frequently encountered word combinations (e.g. Ellis, 1996; Hopper, 1988, 1998). The more often words are encountered in the syntactic company of other words, the more the child will store them as chunks, and the more the child will be able to infer about the recurring patterns (rules) that tie such chunks together. Grammatical competence thus arises from acquaintance with words in use as they are committed to memory and subjected to an implicit pattern analysis (Clark, 2003). Socio-pragmatic competence grows in the same way. As a young child goes about her daily life, she files in memory the details of when, where, how, by whom and in what combinations words are used. Pawley and Syder's (1983) puzzle of native-like selection (Why do we all say things the same way?) is thus accounted for.

Although Levelt's (1989, 1993) influential model of speech processing is concerned with how linguistic knowledge is organized for speech and not with how it is acquired in the first place, here also lexis rather than grammar is the engine driver. According to his model, the linguistic formulation of the preverbal message is initiated by lexical retrieval, and when a word is selected for use, its collocations and colligations are necessarily activated along with it. Hoey's (2005) work on lexical priming reflects a similar architecture of lexical storage: the mental lexicon is a complex of associative networks built from experience. A word embeds itself within its collocational and colligational habitat, and its associated words, grammatical structures and contexts become part of its unique signature in the speech community. In this view there is a good case to be made for describing both first language acquisition and first language use as mediated by a remarkably powerful and retentive memory.

The question arises of how far this can be true of second language acquisition (SLA). For Hoey (2005, pp. 83–4), when a word is learned in a second language (L2), it will inevitably gather to itself the colligations and collocations (primings) of its perceived equivalent in the first language (L1), but these are not likely to be helpful (and can be very unhelpful). Even if a word in the L2 has an exact semantic equivalent in the L1, it will not have parallel primings. The learner, however, may well assume this to be the case and as a consequence produce very unnative-like selections. For Hoey this is an especially likely outcome of classroom SLA where words might be presented 'stripped of their primings' in lists to be memorised, and where contextualized encounters with L2 vocabulary are restricted to what the textbooks or tasks present. Learners in this situation are not in a

rich enough environment to be able to build native-like primings for words, and will use an L2 vocabulary item as if it had the same lexical and grammatical relationships as its L1 translation. For learners who are living inside the target language environment, however, the situation is different. The language is encountered beyond the classroom in a multitude of contexts, allowing for at least the possibility that learners will be able to build more authentic networks of L2 word associations.

A number of studies have looked at the effect of learning context on various dimensions of SLA, and these have illuminated interesting differences in its impact. Collentine (2004), using morphosyntactic accuracy as a measure for SLA development, confirmed earlier findings by Möhle and Raupach (1983) and DeKeyser (1991) in finding little difference between Study Abroad (SA) learners with At Home (AH) learners. Collentine reports no advantage for SA learners in terms of vocabulary size, but the study did not look at vocabulary beyond the level of word recognition. Möhle and Raupach (1983) found that compared to AH learners, their SA learners developed greater fluency and ability to sound natural in the L2, and Regan (1998) doing a similar study and getting similar results, posits that such fluency gains for SA learners are, in part at least, down to their greater use of formulaic sequences. Milton and Meara (1995) showed that an SA context made a significant contribution to the size of a learner's L2 vocabulary while Ife, Vives Boix and Meara (2000) described the SA context as leading to an L2 vocabulary with a more native-like organization. Marriott (1995) and Siegal (1995) also showed greater use of formulaic sequences in SA learners and suggest that these, rather than greater syntactic knowledge, account for the observed higher levels of syntactic complexity in this group. Unusually for such studies, Towell, Hawkins and Bazergui (1996) used a within-subjects design to compare English L1 learners of French before and after an SA year in France. They reported significant gains in fluency due to the learners using a greater number of formulaic sequences of words. Regan (1995) showed no gain in morphosyntactic control for SA learners of French, though they had acquired a facility to delete the negative particle '*ne*' and thus sound more fluent and colloquial, something presumably picked up from SA exposure to spoken French where this deletion abounds. Broadly, these studies are pointing to lexical organization, especially of the formulaic language kind, as the main area of benefit for SA learners, and not morphosyntactic accuracy. This is an interesting finding, suggesting that while morphosyntactic development is amenable to the mature cognitive reasoning that adults (and not children) are able to bring to the learning task, lexical development (in adults and children) thrives more on wider and more varied exposure to the target language. It also suggests that the irresistible pattern analysis which children bring to bear on their growing acquaintance with words in use, and which results in implicit mastery of L1 grammar, does not work as automatically (or as successfully) in adults.

What these studies show is that learning context is a significant variable in the acquisition of lexical knowledge of an L2, though the measures used to gauge this knowledge were usually receptive. For example, participants were asked if they recognized a word as part of the L2 lexicon. The study reported in this chapter was designed to explore this from a different angle, by looking at productive lexical knowledge instead, not from structured interviews where the participants have a degree of individual freedom in what they will say in their answers, but by cartoon picture prompts requiring all the participants to describe the same thing. The lexical profile of the data is therefore likely to be more homogeneous and enable an analysis of how different speakers formulate the same preverbal message. The study contrasted two comparable groups of intermediate learners of English in different learning contexts. The 40 participants in the 'Study Abroad' environment of London were mostly female, were aged between 19 and 47, came from a wide variety of mother tongue backgrounds, and had been in the UK for at least a year. In an Oxford Placement Test they had all been scored as 'intermediate' (band 4). The 60 participants in the 'At Home' environment of Tehran were all female, aged between 19 and 45, and L1 speakers of Farsi. They had all studied English for at least three years and in the locally administered placement test of English proficiency they had achieved scores that put them in an 'intermediate' category. They also took the Oxford Placement Test. A Pearson product-moment correlation was run to compare their two scores. The correlation coefficient (r = .56, significant at the .01 level) was considered large enough to equate the results reliably. (To be even more confident of the equivalence of the two groups, all learners in both venues were given oral interviews to confirm their proficiency.) The study also included baseline data from a third group of participants. These were 40 London-based native speakers, aged between 18 and 60, all undergraduates studying literature or psychology at a university in London. All had learned English from early childhood, and none were proficient in any other language.

The 140 participants were asked to tell the story from two of four strip cartoon prompts, entitled *Picnic, Journey, Walkman* and *Football* (taken from Heaton, 1966, Jones, 1979, and Swan and Walter, 1990). They were carefully piloted to make sure they were all engaging, and not too demanding for intermediate learners to attempt. The storylines are briefly summarized in Table 6.1.

Each narrative comprised six frames and had two main characteristics: a loose or tight structure, and a simple or dual storyline. A loose structure means the sequence of frames can be changed without loss of narrative coherence, as for example in the *Walkman* and *Journey* narrative where the middle frames could appear in any order and the stories would be essentially the same. A tight structure does not allow this, as in the *Picnic* and *Football* narratives where only the given order makes any sense. *Picnic* and *Walkman*

Table 6.1 Synopses of the four narratives used in the study

Task	*Frame one*	*Frame two*	*Frame three*	*Frame four*	*Frame five*	*Frame six*
Picnic: tight structure, with background events	Two children and their mother are preparing a picnic. A puppy watches them	The mother shows the children a map, while the puppy jumps into the basket unnoticed	The children wave to their mother as they leave the house with the basket	They are climbing up a hill in the countryside	The puppy jumps out of the basket when they sit to eat	The puppy runs off and they see there is no food left
Football: Tight structure, without background events	Four boys are playing football in a park. One kicks the ball very high	The ball goes over the head of a boy and lands in a hole in the ground	He tries to reach it but is warned there is a snake coming. Another boy is thinking what to do	This boy runs off	The boy returns with a large tub of water	He pours the water into the hole, the ball floats up and they get it out
Walkman: Loose structure, with background events	A man has left his house and is walking along listening to music on his walkman	Two cars collide behind him, but he does not hear anything	A robber smashes the window of a jewellery shop behind him, but he hears nothing	Two men have robbed a bank and are shooting at the police, but he hears nothing	He is sitting reading in the park, still listening to his music. He does not see a tiger walking past	He arrives home. His wife asks him if he saw anything and he says no
Journey: Loose structure, without background events	A couple are cycling on a country road	A car passes them	They sit outside a pub drinking a glass of beer	The man is in the sea and waves to the woman who is on the beach	They knock at the door of a guest house	An older couple serve them a meal

	Inherent narrative structure	
Narrative events	***Loose***	***Tight***
+ ***foreground*** − ***background***	journey story *n* = 70	football story *n* = 70
+ ***foreground*** + ***background***	walkman story *n* = 70	picnic story *n* = 70

Total *N* = 140, comprising 60 in Tehran, 40 in London and 40 native speakers. Participants did *either* journey and football tasks, *or* walkman and picnic tasks. The task order was counterbalanced.

Figure 6.1 Research design

have both background and foreground events, and thus have dual storylines, while *Journey* and *Football* each have only foreground events and thus a single storyline. These characteristics of task design were chosen in order to investigate the effects on performance of greater or lesser narrative complexity, and are summarized in the research design shown in Figure 6.1.

The spoken narratives were recorded and transcribed, then coded for a variety of variables: fluency (measured through incidences of both repair and breakdown); syntactic complexity (ratio of subordinate to main clauses); accuracy (percentage of error-free clauses); and lexical diversity (Malvern and Richards' (2002) *D* measure; see also Skehan (Ch. 7) and Daller and Xue (Ch. 11), this volume). The larger study included research questions concerning the effects of narrative type on L1 and L2 performance as well as those concerning the impact of the learning environment on L2 performance. The results of the quantitative analyses are outlined here in brief, and are reported in full in Tavakoli and Foster (2008) and Foster and Tavakoli (in press). Narrative type was found to be a significant influence on language performance, with dual storylines leading both the native and non-native speakers into using more subordinated language, while a tighter narrative structure was associated with the non-native speakers producing language of greater accuracy. Fluency was not affected by either task characteristic. Concerning the impact of the learning environment, this was measured by comparing the mean scores of the London learners with those of the learners in Tehran. The result for fluency was mixed, with only the *Journey* task showing an effect; the learners in London were significantly more fluent than those in Tehran, but on one measure only (number of mid-clause pauses). The results for accuracy were clearer, and confirmed the initial placement testing; there was no significant difference between the London and Tehran learners in their levels of accuracy on any of the four tasks. However, the learners in London produced more subordinated language than the learners in Tehran on all tasks, and to a statistically significant degree on three of them (see Tavakoli and Foster, 2008),

Table 6.2 Means (and standard deviations) of *D* for learners in London (L), Tehran (T) and native speakers (NS) for all four narratives and result of the ANOVA conducted on each narrative

	Tehran (n = 60)	*London (n = 40)*	*NS (n = 40)*	*F*	*p*	*Significant differences*	η^2
Football	28.75 (11.20)	38.37 (11.18)	40.21 (7.89)	9.02	.001*	T vs L and NS	.212
Journey	25.82 (9.49)	36.11 (11.01)	38.75 (9.77)	11.84	.001*	T vs L and NS	.261
Picnic	27.76 (5.89)	36.59 (9.46)	39.90 (8.88)	15.87	.001*	T vs L and NS	.322
Walkman	33.62 (6.40)	43.37 (12.43)	45.67 (10.15)	11.61	.001*	T vs L and NS	.258

showing they were more ambitious in English if not more successful than their counterparts. The most intriguing result was obtained for lexical diversity. An analysis of variance compared the lexical diversity of the language produced by all three groups of participants, using Malvern and Richards' (2002) *D*. The results, shown in Table 6.2, indicate a significant effect of group. Furthermore, post hoc tests show that not only were the learners in London using vocabulary that was significantly more diverse than that of the learners in Tehran, they were using vocabulary that was no *less* diverse than that of the native speakers.

These results are in tune with the earlier studies reported above that compared SA learners with AH learners and found the benefits of an SA learning environment to be lexical rather than morphosyntactic in nature. Figure 6.2 shows a task effect on vocabulary, with the *Walkman* story prompting the most diverse language from all three groups, and *Journey* the least. The group scores remain in step across the tasks in a remarkably consistent pattern. No matter what the task, the learners in London are, in terms of diversity, lexically closer to the native speakers than they are to the learners in Tehran, and as we saw in Table 6.2, to a highly significant degree.

These intriguing results were further explored in a qualitative analysis of the transcripts. In order to illuminate how the learners in London were influenced by their SA learning environment, their lexical choices were compared with those of the native speakers and the learners in Tehran. To this end, the transcripts for each of the four cartoons were analysed frame by frame across the three participant groups, to lay out how each narrative event was handled by the 70 participants who described it. This analysis produced 24 lists, one for each frame and each with 70 entries: 20 from the native speakers, 20 from the learners in London, and 30 from the learners in Tehran. Each list showed 70 ways of describing the same thing

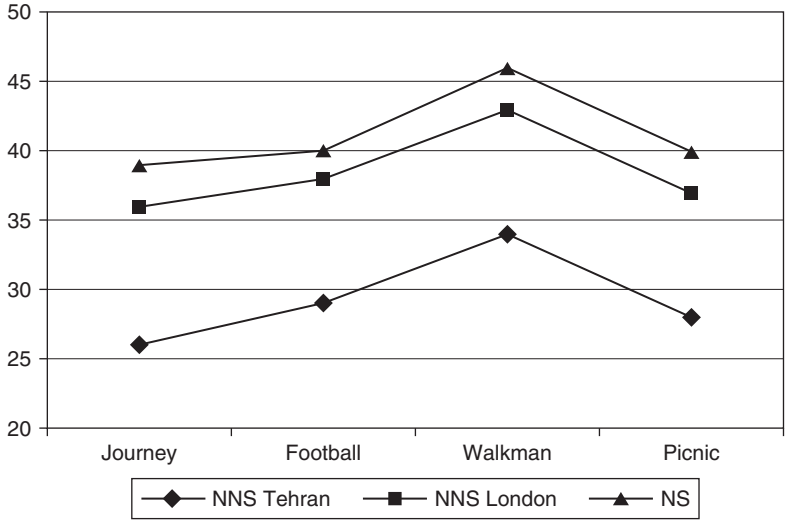


Figure 6.2 Patterns of lexical diversity across the three groups

(including not describing it at all). An example of such a list, abbreviated for reasons of space, is given in Table 6.3. It shows the three groups handling the action presented in frame one of the *Journey* story where two people are setting out for a bike ride. This list was analysed by carefully comparing the particular word choices, together with their collocations and colligations, that each of the three groups made.

Looking at this list, it is clear that the learners in Tehran do not use the more colloquial word *bike*, selecting *bicycle*, either as a direct object of a variety of verbs (*run, take, drive*) or in a prepositional phrase (*by, with, for*). In these phrases they are very apt to preface the word *bicycle* with the possessive pronoun *their*. The list has only one use of the noun *cycling,* and none of the verb *to cycle*. The Tehran learners do not use the word *ride* as a noun or a verb.

The native speakers by contrast use *bike* as much as *bicycle*. Some use *ride* as a verb, others choose the verbs *have* and *go*, as in *have a bike/bicycle ride* or *go bike riding/for a ride on a bike*. The great majority do not use the pronoun *their* with *bike/bicycle*. The most commonly selected preposition is *on* (*on their bikes, on a bike ride, on a bike, on a cycle run*). Several native speakers do not refer to the bicycle at all and use just the verb *cycle*.

Looking at the choices of the learners in London, it is apparent that they are using selections which are different from the choices of the Tehran learners and in several respects similar to those of the native speakers. Many use the colloquial word '*bike*', and only twice as the direct object of

Table 6.3 *Journey* task, frame one: riding the bicycles

Tehran learners	*London learners*	*Native speakers*
They run bicycle	They decided to go for trip by bike	They are cycling to the seaside
They take their bicycle and go out of their home	They're going cycling, their going with their bikes	They are cycling along
They are driving bicycles and in the near of the road	They going for picnic by bike	… cycling down a country lane
They go with their bicycle	They are doing biking	are riding their bikes through a park
They prepare to take their vacation by their bicycles	They cycling	on a cycle run
Two people are driving	They was in the bicycle	… rode their bicycles along the road
they decided that they go vacation for their bicycle	They go by bike	they are having a bike ride
they go with their bicycle	They use their bicycle to go outside	they started out by going for a ride in the country on their bikes
they want to travel by their bicycle	They are cycling	… on a bicycle ride
go with bicycles round sea	They decided to do bicycles	decided to go on a bike ride
… and have fun with their bicycle	They are going to small journeys on their bicycle	on bikes
… and for cycling	They take their bike	they often go bike-riding
	… with bicycle	for a nice country bicycle ride
… with a bicycle	he go by bike to a nice trip	for a bit of a ride on a bike
they decided to go by bicycle	… take by bike	
	… by bicycle in the countryside	
	they really enjoy cycling so they went for a trip	

a verb (but not the verb '*ride*'). Some use the verb *cycle*. In contrast to the Tehran data where many examples of the unnative-like *by/with/for their bicycle(s)* are found, the phrase *by bike/bicycle* crops up often in the London learners' data, and although this is not found in the native speaker data in this study, it is a common enough collocation in the British National

Corpus (where, by the way, it never occurs with any possessive pronoun). Why the native speakers do not choose *by bike/bicycle* in this circumstance is intriguing; the British National Corpus (BNC) examples suggest it is for describing how you get somewhere, and not what you are doing. If this is true, the learners in London have picked up the native-like collocation '*by bike/bicycle*' without (yet) fully knowing its native-like boundaries of use.

A similar analysis was carried out for each narrative event in each story, resulting in an exhaustive and exhausting account of the contrasts and similarities between the two groups of learners, and between the learners and native speakers. To get a more manageable (and presentable) overview of the analysis, a few general observations were extracted, and are detailed below with some representative examples from the transcripts. These are understood better by reference to the narrative synopses in Table 6.1. For brevity, the Tehran data are referred to as T, the London data as L and the native-speaker data as NS.

'General purpose' verbs

Compared to native speaker use, the non-native speakers in both London and Tehran tend to rely on verbs such as *go, come, say, give, understand* and *see* to get their message across. But learners in London are likely to show examples of the more narrowly defined lexical choices that the native speakers make instead: *jump, hide, stop, tell, explain, realise, notice, pay attention*. Thus we find in the *Picnic* story, frames one and two:

(They) go to/for/on a picnic
Very frequent in L, one use in T, infrequent in NS

They prepare (lunch/food/sandwiches) for a picnic
One use in T, frequent in L, very frequent in NS

The dog went to the basket
Very frequent in T, quite frequent in L, not found in NS

The dog jumps/jumped into the basket
One use in T, quite frequent in L, very frequent in NS

And in the *Journey* story, frame three:

They go/went to a …
Very frequent in T, less frequent in L, very infrequent in NS

They stop/stopped (off) at a …
Very infrequent in T, frequent in L, very frequent in NS.

In the *Walkman* story, frame two:

He does/did not understand
Very frequent in T, only one use in L, no use in NS

He does/did not notice/realise/pay attention/care
Very infrequent in T, very frequent in both L and NS

In *Football* frame four, although a large number of the Tehran learners do not describe the action at all, those that do prefer 'go'. A smaller number of learners in London do the same, though some choose similar wording to the NSs:

(He) goes/went …
Fairly frequent in T and L, two uses only in NS

He ran off/away/home
Not found in T, fairly frequent in L, very frequent in NS

Delexicalized verbs

A common feature of informal spoken English (see also Sealey, this volume, on written language), these semantically depleted words are to be found much more often in the London and native speaker data than in the Tehran data. The NS are very likely to prefer the phrasing of *have something to eat or drink, have a swim* where the Tehran learners select *drink, eat* and *swim* as main verbs. The learners in London are more likely to make the NS selection. In *Journey*, frames three and six, for example:

They drink (something)
Very frequent in T, less frequent in L and not found in NS

They have a drink/have something to drink
Very infrequent in T, more frequent in L, very frequent in NS

They eat (something)
Very frequent in T, far less frequent in L and not found in NS

They have (a meal)/something to eat
Infrequent in T, frequent in L, very frequent in NS

Something similar is evident in *Journey*, frame four:

They swim/are swimming
Very frequent in T, infrequent in L, not found in NS

They have a swim
Not found in T, two uses in L, very frequent in NS

Collocate phrasing

As we have discussed above with the examples for bike riding given in Table 6.3, the learners in London are more likely to select the NS collocate prepositions and verb. In *Picnic*, frame one, for example, though all the learners know the word 'picnic', those in T use it mostly in a way not found in NS:

They go to (a/the) picnic
Very frequent in T, infrequent in L, not found in NS

They go for/on a picnic
One use in T, frequent in L, very frequent in NS

In *Picnic* frame three, the learners choose the word 'goodbye' or 'bye-bye', and all but one collocate it with 'say'. The NS prefer to use 'goodbye' as a kind of adverbial with the verb 'wave', and one learner in London does the same:

They say/said goodbye/bye-bye to their mum/mother
Very frequent in T and L, not found in NS

They wave/waved goodbye to their mum/mother
Not found in T, one use in L, very frequent in NS

Existential 'there is/are'

There is evidence that the learners in Tehran and (to a lesser extent) in London prefer a subject–verb construction to describe an event where the NS selection is to choose a noun phrase after *There is/was*. This is the case in *Walkman* scene four, with one example of a learner in L making the NS preferred selection:

(They) shoot at/to/towards/with them
Frequent in T, occasional in L, not found in NS

There is/was a shoot-up/shoot-out/gunfight
Not found in T, one use in L, very frequent in NS

This kind of thing happens also in *Walkman* scene two, where there are several selections in T and L of *crash* as a verb, and in T there are even examples of *accident* as a verb:

(Two) cars crash/ed
Frequent in T, and L, not found in NS

Two car accident with them
Several uses in T, none in L or NS

The commonest NS selection of *crash* and *accident* appears to be as a compound noun with *car* following *there is/was*, and again it is the learners in London who are most likely to select this:

(there was/is) a car crash/accident
One use in T, quite frequent in L and very frequent in NS

This frame-by-frame analysis puts flesh on the quantitative measure of *D* shown in Table 6.2, illustrating that the greater lexical diversity of the learners in London arises from their being able to choose between more narrowly defined words instead of something broader, that is to say, *notice* or *realise* for *understand; cycle* or *ride* for *go; puppy* for *dog*. The London data contain colloquial choices such as *stop off at* as well as *go, guy* as well as *man, kid* as well as *boy, chat* as well as *talk, have a bit of a swim* as well as *swim, best spot* as well as *best place*. These colloquialisms are either very rare or non-existent in the Tehran data, whereas the native speaker data have, unsurprisingly, plenty of such examples. The analysis also shows evidence that the learners in London are to some degree framing their sentences in more native-like ways.

One final analysis was undertaken to quantify the number of lexical phrases in the data. To do this a native speaker of English (the researcher) identified all the places in the entire corpus where combinations of two or more words could be regarded as prefabricated to some degree and stored in memory as single choices. Because this analysis was concerned with the effect of a target-language environment on SLA, only native-like lexical selections were looked for. It was not possible to identify what idiosyncratic single-choice combinations might have been stored in the lexicons of the non-native speakers, though these certainly exist.

Both fixed and partially fixed selections were counted, using Wray and Namba's (2003) criteria as a guide. For example, a sequence of words is counted as fixed to some degree if judged to be associated with a specific situation and/or register, or as commonly used to convey a given idea, or to have been encountered by the speaker before. To give some examples from the transcripts:

In the background
a day off
eat up
rummaging around
wave (someone) off
blissfully unaware
There was no (food) left

They got the (map) out
On a (picnic/bike-ride)
Not got a clue

Once identified, these were checked in the BNC and regarded as reliably identified if they occurred there more than five times (for the vast majority of such phrases the BNC recorded between 50 and several thousand occurrences). The incidence was then calculated for the native speaker and the non-native speaker transcripts. This method of quantifying native-like selections was developed for this project. In a previous study, Foster (2001a) had identified native-like lexical selections in the data using the combined intuitions of six native speakers, but this procedure was very time-consuming. By using just one native speaker's intuition, guided by Wray and Namba (2003) and checked against the 100 million word sample of the BNC, a faster (but no less valid) result could be obtained. These are presented in Table 6.4, as mean totals per participant in the three groups across the four tasks.

Though the London and native speaker groups each comprised 40 participants, the Tehran group comprised 60 which produced rather uneven word totals for each. The learners in Tehran produced an approximate total of 19,200, those in London 16,000 and the native speakers 13,700. To avoid manipulating the mean totals of lexical phrases further, these are not adjusted to account for the different size of the sub-corpora. But even bearing this in mind, it is very clear from Table 6.4 that lexical phrases are much more frequent in the native speaker corpus (a mean total of 21.1 per participant) than either of the two non-native speaker corpora, and also that they are much more frequently encountered in the London corpus (10.6 per participant) than the Tehran corpus (5 per participant), in spite of the latter being so much longer. It is clear that the learners in London are drawing on lexical resources which the learners in Tehran do not have, and are able to construct utterances from more prefabricated chunks of a native-like character.

Taken together, the analyses conducted on the language output of these learners show that those in London were not any better than their

Table 6.4 Mean lexical phrases per participant (counted as tokens, not types)

Narrative	*Learners in Tehran*	*Learners in London*	*Native speakers*
Football	0.8	1.7	5.6
Journey	1.4	3.0	6.1
Picnic	0.7	1.9	3.8
Walkman	2.1	4.1	5.7
Total mean	5.0	10.6	21.1

counterparts in Tehran in terms of the grammatical accuracy of their English, but they had incorporated into their lexicon more native-like selections of single words and strings of words. The vocabulary choices of the learners in Tehran were not necessarily wrong, of course, but more limited and less native-like. The learners in London had developed their English lexicons to such a degree that on these tasks it was statistically no less diverse than that of the native speakers, and that is a remarkable achievement. The analyses also showed that the learners in London attempted significantly more subordinated language than their Tehran counterparts, and this might be in part down to knowing (imperfectly perhaps) more syntactic frameworks. For example, one learner in London described the dog in the picnic story as *He is interesting seeing how they prepare the food*, which is syntactically complex, certainly native-like in inspiration, but let down by the grammar connecting the first to the second clause. Unlike the results obtained in Möhle and Raupach (1983) and Regan (1998), greater use of formulaic sequences was not associated with greater fluency for the learners in London, apart from the result of one measure which suggested they were less likely to pause in the middle of a clause than were the learners in Tehran. This suggests they were better able, to some degree, to plan their utterances as whole clauses as native speakers do. Again, this may be because they are drawing on more memorized clause-length phrases (Marriot, 1995; Siegal, 1995).

The results suggest that living inside the target language community, being exposed to the language on a daily basis in all manner of contexts, results in an enriched and networked lexicon which enables the learner to 'sound more natural' (Möhle and Raupach, 1983) even if they are not more grammatically accurate. By contrast, living outside the target language environment, and encountering the language only in the limited setting of a classroom, could mean learners overworking the words they know, and remaining unacquainted with their boundaries of use. In Hoey's (2005) terms, classroom-acquired vocabulary lacks the necessary primings which attend repeated encounters with words in different contexts, different situations and different speakers. The question arises how such a state of affairs should be taken into account in language testing. Most learners of English are not able to stay for an extended period of time in an SA environment, and idiomatic British English is not necessarily their goal. They may come across as less natural to a British ear and this might lead, for example, to a British native-speaker tester undervaluing their English, despite its level of grammatical accuracy. The opposite side of this coin is illustrated by the case of Wes, a Japanese immigrant to the USA studied by Schmidt (1983). His English was evaluated by his friends as having improved a lot over three years in Hawaii, but Schmidt's analyses shows Wes's grammatical ability had hardly improved at all. What did improve was his command of formulaic sequences which helped him come across as fluent and native-like.

One final benefit of studying abroad

Putting something into words depends ultimately on the learner having an appropriate entry in her L2 lexicon in the first place. In default of that, she will either avoid describing that part of the scene, or will resort to selecting the closest word she can find. This strategy may account for some of the learners' wayward choices, but an inappropriate word choice could equally arise from a failure to recognize what the picture shows. Learners who do not read the cultural signposts in the artwork are at a disadvantage which has little to do with linguistic proficiency. From analysing the transcripts it seems that some of the pictures presented more problems of interpretation to the learners in Tehran than to those in London. (The native speakers had no such problems.) For example, in the *Journey* story it is possible that many of the learners in Tehran did not recognize the pub sign as indicating a pub, calling it instead a restaurant or coffee shop. Unlike the learners in London they did not recognize the shape of the glass as indicating beer rather than coffee or juice. Most learners in Tehran also called the beach a swimming pool, perhaps again not recognizing that sandcastles and deck chairs are typical features of a UK beach, something which most learners in London appeared to know. In the final picture of this story, most learners in Tehran had failed to understand what meal this was and where it was taking place, perhaps not knowing the bed and breakfast tradition in the UK. In the *Football* narrative only a few in Tehran identify the game as football, unlike the London learners who mostly identified it correctly. The building in frame one may have looked like a house to many in Tehran, misleading them into describing the boys playing in a garden. For the learners in London, however, the building is recognized as a typical British sports pavilion, and so they concur with the native speakers in describing the boys as playing at school, or in a park. Overall, it is clear that the learners in London have an advantage over their Tehran counterparts because they are more familiar with British culture and are closer to the native speakers in their understanding of what the pictures are drawn to represent. As vehicles for language testing, such narratives could be setting booby traps for a learner studying outside the culture in which the pictures were created.

Note

1. The study on which this chapter is based was funded by a grant from the Economic and Social Research Council (RES-000-22-1155).

7

Lexical Performance by Native and Non-Native Speakers on Language-Learning Tasks[1]

Peter Skehan
Chinese University of Hong Kong

Introduction

The last 20 years or so have seen a vast increase in research into second language learning tasks. A series of articles has been published by this author and co-researchers taking a cognitive approach to task performance (Foster, 2001a; Foster and Skehan, 1996, 1999; Skehan and Foster, 1997, 1999, 2005, 2007). This chapter reports on a meta-analysis of these studies (see also Skehan and Foster, 2007), but it does so with two additional foci. First, most research with tasks has focused only on second language learners. As a result, it is difficult to disentangle whether performances which are reported are the result of the different variables which are being manipulated (e.g. task characteristics, task conditions) or simply the second language speakerness of the participants. One needs baseline native-speaker data, of the sort reported in Foster (2001a), to enable a better perspective on the results to be obtained.

A second shortcoming of the research is that it has used a restricted set of performance measures. These have been *complexity*, generally measured through an index of subordination which is based on analysis of speech (AS) units, roughly equivalent to clauses (Foster, Tonkyn and Wigglesworth, 2000); *accuracy*, measured usually as error-free clauses; and *fluency*, measured variously through pausing-based indices (e.g. Foster and Skehan, 1996), repair indices such as reformulation, false starts and so on (Foster and Skehan, 1996), speech rate (Tavakoli and Skehan, 2005), or length of run (Skehan and Foster, 2005). A major area of omission concerns the lexical aspects of task performance. There have been occasional attempts at measures here. Foster and Skehan (1996), for example, did explore measures of lexical variety, and Robinson (2001) reports values for what he terms the token-type ratio, but in the main the lexical area has not been well served.

A brief word is necessary in this section also on the meta-analytic nature of the research reported here. The research is based on a series of linked studies, six in total, which will be detailed below. The present research

therefore is an attempt to establish patterns which emerge across larger data sets. It is hoped that this approach will produce more robust and generalizable results (Norris and Ortega, 2006).

Measures of lexical performance

The literature on lexical performance generally distinguishes between text-internal and text-external measures (Daller, Van Hout and Treffers-Daller, 2003). The main text-internal measure which is widely used is the type-token ratio. However, the basic measure is extremely vulnerable to a text length effect (Malvern and Richards, 2002), and typical correlations between text length and type-token ratio are negative and in the order of −.70 (Foster, 2001b). A series of responses to this problem have been developed and these are reviewed in Tidball and Treffers-Daller (2007), Van Hout and Vermeer (2007) and Jarvis (2002). The different corrections for length have strengths and weaknesses, but for the present research, the measure which was used is *D*, obtained through the use of the VOCD sub-routine within CLAN (and CHILDES: MacWhinney, 2000). In a series of publications, Malvern and Richards (2002, Richards and Malvern, 2007) have demonstrated the reliability and validity of this measure, which is based on mathematical modelling. McCarthy and Jarvis (2007) propose that there are measurement-related flaws in the use of *D*. However, it is clear that the value that *D* delivers correlates very highly indeed with other measures which are proposed and so there seems no reason not to use it as the most effective lexical diversity measure available.

The next question, of course, is to ask what such a measure measures. At this point, things become a little less clear. At one level, the answer is simple: *D* provides an index of the extent to which the speaker avoids the recycling of the same set of words. If a text has a lower *D*, it suggests that the person producing the (spoken or written) text is more reliant on a set of words to which he or she returns often. This naturally raises the question as to which factors influence the values for *D*. The problem is that there are multiple possible factors involved here. These include:

- The development of greater vocabulary size and so the capacity to choose from a wider range of words where previously there was a smaller repertoire. One might predict therefore that age for first language learners, or proficiency level for second language learners, would be associated with higher values of *D*.
- The possession of a better organized lexicon, with the result that a greater range of words can be easily drawn on.
- Performance conditions, for example written versus spoken performance, would allow more time for lexical retrieval, generating higher values of *D*.

- A repetitive style, which might be an individual difference factor, could be important here. The contrast would be with a style which tries to achieve what might be termed elegant variation, where the speaker attempts to avoid recycling in order to convey an impression of composed, created language. (This influence will not be pursued here, since it does not connect with the present research design.)
- There may be task influences in that when topics in conversation change with regularity, this may lead to new 'sets' of words being accessed leading to lower opportunities for lexical recycling over the text as a whole.

Clearly the problem here is the existence of what is only a laundry list of influences, reflecting underlying lexicon, communication style, and task influences. The difficulty is disentangling which of these influences is most operative. The present study will begin to address these issues.

A contrasting class of lexical measures uses some external yardstick to evaluate a different construct of lexical variety. Essentially, a measure is computed of the extent to which the speaker draws upon more varied words, referenced by some external criterion. This has been termed 'lexical sophistication' (Read, 2000). Two issues are immediately apparent. First, there is the question of what 'varied words' might mean. Second, there is the problem of how an index is computed which reflects putative variety.

The standard approach to defining variety has been through word frequency. A performance is then judged in terms of its tendency to draw upon less frequent words. One of the most influential methods, the Levels Test (Laufer and Nation, 1999), uses word lists based on generalized written corpora, including specialist corpora for academic words. The test provides information on the number of words in a text drawn from the 1000 word level, the number drawn from the 2000 word level and so on, enabling a judgement to be made regarding the 'penetration' in the text of less frequent words. The ensuing judgement therefore is profile based and gives a complex but interesting perspective on the extent to which very frequent words are less relied upon.

An alternative measure also exists, though, which, like *D*, uses a mathematical modelling procedure. Meara and Bell (2001) have devised a procedure, P-Lex, which divides a text into ten-word chunks, and then computes the number of infrequent words in each ten-word chunk. For example, one might have the distribution shown in Table 7.1 for a 300-word text. There are 30 ten-word chunks to work with in Table 7.1 (hence the numbers in

Table 7.1 Distribution of ten-word chunks with infrequent words

No. of infrequent words per 10 words	0	1	2	3	4	5	6	7
No. of word chunks	9	9	6	4	1	1	0	0

the second row add up to 30). One can then explore how many ten-word chunks contain no infrequent words, how many contain just one, and so on. The distribution from the set of scores shown in the table suggests a text where ten-word chunks with no or only one infrequent word predominate, with nine of each. Intuitively, this (hypothetical) distribution suggests a text with mainly fairly frequent words. Meara and Bell (2001) demonstrate that distributions such as that shown in Table 7.1 can be modelled by the Poisson distribution, a distribution particularly appropriate for data with infrequent events. The method is to estimate the value, lambda, which generates a Poisson distribution which approximates the actual pattern of scores with most accuracy. P-Lex has been widely researched and it has been demonstrated (Bell, 2003) that it is an effective measure for texts which are longer than about 100 words. Table 7.2, where the examples are drawn from the data sets covered in this chapter, provides some examples of actual score distributions, and the associated lambda values.

Clearly, the first two speakers have more ten-word chunks which contain infrequent words, that is to say, the penetration of infrequent words goes further to the right in each set of scores, while Speaker 3 produces a preponderance of ten-word chunks with no infrequent words, or only a small number of such words. The lambda values reflect these distributions, and show that, the higher the lambda, the more infrequent words are being used.

The original computer program, P-Lex, needed some slight modifications for the data sets used in the present meta-analyses. The rewritten program was referenced from the British National Corpus spoken component, and so drew upon a corpus of 10 million words (Leech, Rayson and Wilson, 2001, and also the Lancaster corpus linguistics group website). The reference list was lemmatized (and in fact could be used to generate lambda values either in lemmatized or unlemmatized forms). Files of task-specific words were compiled to enable words to be temporarily defined as easy, adaptable for different runs of the program. Finally a cut-off value, using the lemmatized reference list, of fewer uses than 150 per million words was used as the basis for defining difficulty, or rarity, the central requirement of the P-Lex program (Meara and Bell, 2001; Bell, 2003). This value seemed to be most effective in producing a good range of discrimination. It might also be

Table 7.2 Example distributions and associated lambdas (λ)

No. of infrequent words per 10 words	0	1	2	3	4	5	6	7
(Native) speaker: personal task: λ = 1.50	4	6	2	2	2	0	0	0
(Non-native) speaker: narrative task: λ = 1.54	6	8	9	3	2	0	0	0
(Non-native) speaker: decision-making task: λ = 0.78	18	10	6	2	0	0	0	0

regarded as fairly 'generous' in making difficult decisions. However, spoken language tends to contain notably fewer infrequent words than does written language.

Assuming this provides a valid and reliable measurement option, we still need to discuss what the construct of lexical sophistication represents and what influences it. Earlier, for lexical diversity, a variety of influences were discussed. These were:

- development of vocabulary size and/or organization;
- performance conditions, such as modality, time pressure, planning opportunities;
- style, whether repetitive or variational;
- task influences.

Interestingly, all of these would also seem relevant for greater lexical sophistication. Greater size and/or organization of vocabulary should enable greater lexical sophistication. Similarly, favourable performance conditions such as planning versus no-planning should similarly be associated with a greater capacity to draw on less basic vocabulary. Style is difficult to comment on here, although perhaps this variable is less salient for lexical sophistication than for lexical diversity. Finally, task influences too might well have an impact on performance, although whether these are the same task influences as those which impact upon lexical diversity is an empirical issue. On the face of it, though, a similar set of influences may be operative, and so one might, again at first sight, expect lexical diversity and lexical sophistication to pattern similarly. Exploring their actual interrelationship will be one of the central themes of the present research.

The research database

Table 7.3 outlines the six studies which form the basis for the present meta-analysis. The individual studies drew on a range of task types and task characteristics, on the one hand, and task conditions, on the other. Tasks fell into one of three categories: personal information exchange (P); narratives, either based on picture series or on a video (and necessarily more monologic in nature) (N); and decision-making, where, through interaction, pairs or groups of students were required to make decisions (D). Examples of the tasks are as follows:

> *Personal Information Exchange*: 'You are at school and you have an important examination in ten minutes. But you suddenly remember that you have left the oven on in your flat. Ask your friend to help, and give them directions so that they can get to your home (which they have never visited) and then get into the kitchen and turn the oven off.'

Table 7.3 Overview of the studies

Study	*Focus*	*Results*	*Size in words*
1. Foster and Skehan (1996) (NNS)	P vs N vs D Planning	Strong planning effect Selective task effect	25K
2. Skehan and Foster (1997)	P vs N vs D Planning Post-task	Strong planning effect Selective task effect Partial post-task accuracy effect (decision-making task only)	36K
3. Skehan and Foster (2005)	Planning Mid-task surprise	Strong planning effect No effect of mid-task surprise information	18K
	Time (5 vs 10 mins)	Strong time effect (5 mins > 10 mins on all measures)	
4. Skehan and Foster (1999)	Degree of structure (narrative tasks)	Structured task was more fluent and sometimes more accurate	30K
	Processing load	Simultaneous processing is very difficult	
5. Foster and Skehan (under review)	N vs D Post-task condition	Clear accuracy effect of post-task on both tasks	30K
6. Foster (2001a) (NS, same design as Study 1)	P vs N vs D Planning	Strong planning effect with complexity and fluency	25K
	NS vs NNS	Native speakers less formulaic when planned, non-native speakers the reverse	

P = Personal task: N = Narrative task: D = Decision-making task: NS = Native speaker: NNS = Non-native speaker.

Narrative: a cartoon series from the work of the French cartoonist, Sempé, was presented. It showed a story of a woman going to the fortune teller's. While having her fortune told through cards, the fortune teller's telephone rings (situated directly behind the fortune teller). While the fortune teller's back was turned, the client turned up the cards, saw they were not to her liking, and rearranged them. When the fortune teller finished the call, she unsuspectingly turned back round and told the (glowing) fortune based on the rearranged cards.

Decision making: Participants were given letters supposedly written to a magazine Agony Aunt and were required to agree on appropriate advice. A typical letter (of three presented in total) would be: 'I'm 14 and I am madly in love with a boy of 21. My friends have told him how I feel and

he says that he likes me, but he won't take me out because he says I am too young. I'm upset. Age doesn't matter, does it?'

Table 7.3 provides an overview of the results of these studies. The dependent variables (cf. the earlier discussion) are always complexity, accuracy and fluency. Then a series of independent variables have been explored, including task characteristics, as well as pre-, during- and post-task conditions. Pre-task planning was generally operationalized through the provision of ten minutes' planning time; during-task operationalizations were either to introduce surprise new information while the task was being done or to vary the time pressure conditions; the post-task condition was either to have to redo a task, publicly, after the actual task was done, or to have to transcribe one's own performance, post-task. A very brief outline of the results for each study is shown, as is the corpus size for each study, in thousands of words.

For now, we can see that a series of generalizations can be made on the basis of the results reported in Table 7.3 (see Skehan and Foster, 2007, for extended coverage):

- planning has a consistent effect, strongly raising complexity and fluency, and raising accuracy to a lesser extent;
- a post-task condition, for example, a public performance of the same task, or the requirement to transcribe some of one's own performance after the task is completed, leads to raised accuracy, especially with the interactive decision-making task;
- personal tasks based on familiar, concrete information lead to higher levels of fluency and accuracy;
- decision-making tasks produce higher accuracy and complexity;
- narratives appear to be the most difficult task type, with lowest accuracy;
- tasks containing structure such as tasks based on a clear schema, like a restaurant schema, or alternatively a problem-solution schema (Hoey, 1983) lead to raised accuracy;
- tasks requiring the transformation, manipulation or integration of information lead to greater language complexity;
- there is a trade-off between the performance areas, with higher performance in one area often being at the expense of others.

Lexical performance on tasks

There are four basic questions to be considered in this section. First, fundamentally, we need to explore how native and non-native speakers differ in their performance. Second, and equally fundamentally, we need to consider how the two measures of lexical performance interrelate. Third, there is the general question as to how the lexical measures relate to other

measures – whether, for example, they relate to complexity, or accuracy, or neither. Finally, there is the issue of what influences the lexical measures. We have seen different patterns of influence on complexity and accuracy, as in Table 7.3. Now we need to explore the same question with the lexical measures.

Native versus non-native speakers

The first comparison to explore is that between native and non-native speakers. This comparison is only possible for Studies 6 (Foster, 2001a) and 1 (Foster and Skehan, 1996), since it was only in these two that participants did exactly the same tasks, and with essentially the same conditions (10 minutes planning versus no planning). The relevant results, based on between-subjects *t*-tests, are presented in Table 7.4.

The obvious (in both senses of clear, and also predictable) generalization here is that native speakers produce more impressive lexical performances than do non-native speakers. All significances are at the .001 level, and the differences are all very clear indeed. Native speakers, that is, when doing tasks, draw upon less frequent vocabulary, and also pack a greater variety of words into a text they produce. But there are also some interesting task–speaker interactions. Native speakers are most appreciably higher in their lambda scores with the personal and narrative tasks. The difference between the two groups is much less with the decision-making task, and in any case, the lambda scores here are clearly lower for both groups. The situation is different with *D*, since here the personal task shows the least difference. Once again the narrative task shows a large difference, making this the only task which is consistent in a large advantage for native speakers for both measures. But interestingly, the values for *D* for the decision-making task show a clear difference between native and non-natives (in contrast to the smaller difference for lambda). The interactive task leads to less recycling with the native speakers: non-native speakers are nothing like as impressive.

Table 7.4 A comparison of the lexical performance of native and non-native speakers

	λ			*D*		
	Personal	*Narrative*	*Decision-making*	*Personal*	*Narrative*	*Decision-making*
NS mean	1.38	1.68	0.87	45.6	75.2	90.6
($N = 31$)	(0.39)	(0.52)	(0.25)	(11.3)	(19.0)	(11.5)
NNS mean	1.02	1.14	0.65	36.1	46.9	52.9
($N = 29$)	(0.33)	(0.37)	(0.27)	(9.7)	(13.4)	(11.9)
Significance	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$

Standard deviations are given in parentheses.

We can assume that the two groups differ markedly in both the size and organization of their mental lexicons. So it seems that the possession of larger, better-organized lexicons does lead to the use of less frequent lexis *and* less recycling of a smaller set of words. In the first case, the larger vocabulary, linked to its greater accessibility, means that the fund of words that is drawn upon is considerably greater and native speakers can react to tasks effectively. In the latter case, it is also clear that native speakers are not reliant on limited word sets which they have to keep using because of the lack of others.

The relationship between *D* and lambda

The relationship between these two measures is shown in Table 7.5. The table is organized in terms of the different studies which were completed, and also shows the separate tasks where there was more than one task in a study. Each cell gives a correlation coefficient to indicate the nature of the relationship concerned. Significance and marginal significance are shown.

The *N* sizes in these studies are not large, and so it is difficult to achieve significance. But the basic conclusion is unavoidable – the level of relationship between these two measures is very low at best, and more probably, non-existent. The highest correlations would only account for very low levels of shared variance. This applies to native speakers and non-native speakers alike, and across personal, narrative and decision-making tasks. We thus have to draw the conclusion that lexical diversity and lexical sophistication are independent of one another. Earlier, it was speculated that possessing a

Table 7.5 Pearson correlations between *D* and lambda in different studies

	Personal	*Narrative*	*Decision-making*
Study 6: Native speakers ($N = 31$)	0.19	−.28	−.30
Study 1: Non-natives ($N = 29$)	−.27	0.08	0.32
Study 2 ($N = 40$)	0.35	0.06	−.11
Study 3 ($N = 64$)	n/a	n/a	0.29 ($p < .04$)
Study 4 ($N = 22$; $N = 24$)	n/a	−.33; −.06	n/a
Study 5 ($N = 45$)	n/a	−.23	0.31 ($p < .06$)

Study 6 is placed first, next to Study 1, since the same tasks were done in both, with the difference only being the native vs non-native speaker status. In addition, for Study 4, two values are shown, since there were two narrative tasks. Where 'n/a' is shown, this indicates that a relevant task type was not used in that study.

larger and better-organized lexicon might raise *D*, and one might also speculate that this would impact upon lambda also. The evidence is not consistent with this happening. So it may be that the salient influences upon *D* are not the same as those on lambda. We need to look elsewhere to try to tease out what these contrasting influences might be.

Relationships of lexical measures to complexity (and accuracy)

The three areas concerned here, lexis, structural complexity, and accuracy, are all part of the formal structure of language. It is interesting to explore, therefore, how they interrelate. Robinson (2001, Robinson and Gilabert, 2007), for example, proposes that accuracy and complexity should correlate, while Skehan (1998) suggests that limitations in attention means that usually they do not, as non-native speakers prioritize one performance area over the other. It now becomes interesting to throw lexical performance into the mix.

There are interesting differences here in the patterns of relationships between measures for native and non-native speakers. For the non-natives across the range of studies, lambda correlates consistently negatively with accuracy, that is to say the greater the lexical sophistication and use of infrequent words, the lower the accuracy. The relationship between lambda and complexity for this same group is not quite such a clear pattern, but the relationship here, too, is mainly negative. Less frequent words, for non-native speakers, are associated with lower complexity. In other words, more varied lexis seems to cause problems for non-native speakers and provokes more error while not driving forward complexity. There seems, in other words, to be something of a toll for those who mobilize less frequent lexical items, in that the syntactic implications of such words derail, rather than build, syntax.

It was considered inappropriate to use accuracy measures with the native speakers in Study 6 in the present data sets. But we can examine the relationship between lambda and syntactic complexity. This is positive, with the three correlations of .43 (personal, $p < .05$), .57 (narrative, $p < .001$), and .21 (decision-making, not significant), with *N* sizes of 28, 31 and 33 respectively. In other words, for native speakers, less frequent words seem to push speakers to use more complex language. Native speakers seem able to handle the consequences of lemma retrieval without disruption, presumably accessing information quickly and then acting upon its consequences in real time. Non-natives, in contrast, pay a penalty for more difficult lexical retrieval.

The relationships with *D* are different. For non-native speakers, lexical diversity tends to be positively related to accuracy: the less recycling of vocabulary there is, the higher the accuracy that is achieved. Possibly, greater recycling is associated with more within-clause repetition of lexical items as speakers are attempting to buy time to deal with the trouble that

they have encountered, while non-native speakers who are not experiencing trouble are able to avoid such clause-internal repetition and introduce more variation into their speech. Finally, again for non-native speakers, *D* correlates negatively with complexity in the majority of cases. In other words, speakers who recycle vocabulary most, nonetheless are able to achieve greater complexity. Drawing on the same lexical sets, in other words, seems to provide room for attention which can enable more complex language to be produced. Native speakers, in contrast, show no correlation between *D* and language complexity.

These results give us our first major insight into the nature of speech performance for the two groups. If we relate their performance to Levelt's model of speaking (1989), with its three major stages in speech production of Conceptualisation, Formulation and Articulation, it appears to be the case here that with native speakers, Conceptualisation delivers a preverbal message which makes demands upon the Formulator, but that the Formulator meets these demands very well, in that the lexical choices implied by the preverbal message then trigger effective use of syntactic frames. More demanding lexis leads to more complex syntax. With the non-native speakers, in contrast, this does not happen. More demanding lexis implied in the preverbal message creates difficulty for the Formulator and disrupts syntactic planning. Lexis does not drive syntax in the same way as with native speakers. (It does, though, need to be borne in mind that the non-native speakers here are at low intermediate level, and research is certainly needed with higher proficiency levels to explore whether increasing proficiency is associated with a greater correspondence between lexis and syntax.) The final point of interest here is the positive association between *D* and accuracy for the non-native speakers. Comfortable non-recycling seems to be a reflection of a non-native speaker being able to devote ongoing Formulator-linked attention to avoiding error. (The relevance of Levelt's model is covered in much greater depth in the Discussion section.)

Task influences on lexical measures

We have already had a glimpse of the influence of task types while comparing native and non-native speaker data from Studies 1 and 6. Regarding lambda, these two studies are representative of all the others. The average values across all the studies, now drawing on different examples of personal, narrative and decision-making tasks, are 1.23, 1.49 and 0.66 respectively. In other words, narrative tasks consistently produce the highest values, and so are provoking the greatest use of less frequent words. It would appear that the monologic nature of the narrative, coupled with its non-negotiability (i.e. the given story which has to be told, with its characters and elements), accounts for this pattern of lexical use. Lambda for personal tasks does not reach this level. Although not monologic, the personal task often did lead to monologic-type turns as people developed

a viewpoint, and one of the personal tasks used, the Oven task, was itself close to a narrative, with its unavoidable sequence of actions that has to be followed to give clear instructions to one's partner about getting to one's home. But very strikingly here, and for native as well as non-native speakers, the lowest values for lambda are found with the decision-making task. The previous two tasks are either very strongly input driven (narratives, the Oven personal task), or draw upon familiar, well-organized information. The decision-making tasks, in contrast, require a blend of basic cognitive activity where general principles have to be applied to particular cases, and also improvisation, as the interactive nature of the task is responded to. It may also be that speakers aim at a lower level of what might be termed 'idea density', and are more reliant on time-creating devices. Possibly also there is listener awareness because of the greater obviousness of interactivity. The consequence seems to be a lower tendency to use language which is lexis-driven or in which less frequent lexical elements are drawn on as necessary. The differences in the figures between tasks are striking and consistent.

The task effects on *D* are interestingly different. The personal tasks are inconsistent, but the figures for the narrative and decision-making are the reverse of those for lambda. For non-native speakers, the decision-making task is consistently higher for *D*. These differences are not as great as they were for lambda, but are statistically significant for Study 6, the native speakers, and for Studies 2 and 5, with all significance levels at .001 (paired subjects *t*-tests: *N* sizes respectively at 23, 23 and 36). The comparison for Study 1 approaches significance ($p < .08$), with the decision-making *D* score higher than that for the narrative. It appears that interactivity is associated with an avoidance of recycling, possibly because learners are using one another's words more, and so there is scope for 'on the fly' input, in contrast to the more monologic narratives where learners are more concerned to express their own ideas. In contrast, the focus in the narratives seems to be on selecting and retrieving the appropriate word even if it is more difficult to do so. There is also an important task influence in the decision-making tasks, in that these tasks require pairs of participants to discuss a series of things. In one case this was a series of putative crimes, and in another a series of letters to an agony aunt. This means that the topic within the interaction changed at quite regular intervals. It may be that this, too, has a significant effect, with the new topic causing participants to need to use new sets of words. This may then lead to the lower recycling and higher *D*.

Returning briefly to the native/non-native comparison, it is worth recalling from the last section that the difference between these two groups does not operate at a consistent level. In the main there seems a slightly greater difference between them with *D*, and this especially for the narrative and decision-making tasks. With lambda it is the narrative which generates the greatest difference, and there is surprisingly little difference for the

decision-making task. It would seem that narrative tasks provoke the most consistent difference in lexical performance between the two groups, since native speakers here are able to draw upon much less frequent lexis, and avoid recycling lexis. These effects do not appear jointly so strongly with the other two task types.

The effects of planning on lexical measures

In order to be able to compare native and non-native speakers, we will restrict discussion here to Studies 1 and 6 because these studies, with the same research design, explored the performance of the two groups. The relevant figures for *D* are presented in Table 7.6, while the figures for lambda are presented in Table 7.7, where the results of between-subjects *t*-tests are reported.

The difference between these two tables is striking. Lexical diversity does not seem affected by the opportunity to plan. It seems more an online processing issue, which must reflect Formulator operations on a second-by-second basis. In contrast, the values for lambda do show a planning influence, although not everywhere and not all the time. Arithmetically, all planning values are higher than the non-planning values, but significance is obtained for only one of the native speaker tasks, the narrative,

Table 7.6 Influence of planning on *D* for native and non-native speakers

	Personal	*Narrative*	*Decision-making*
Study 6: NS unplanned	46.9	79.8	92.2
Study 6: NS planned	44.3	67.2	89.0
Significance	ns	ns	ns
Study 1: NNS unplanned	37.0	44.7	53.6
Study 1: NNS planned	34.6	49.0	51.8
Significance	ns	ns	ns

Study 6: $N = 31$: Study 1: $N = 29$.

Table 7.7 Influence of planning on lambda for native and non-native speakers

	Personal	*Narrative*	*Decision-making*
Study 6: NS unplanned	1.27	1.46	0.80
Study 6: NS planned	1.48	1.95	0.93
Significance	ns	.01	ns
Study 1: NNS unplanned	0.94	1.10	0.54
Study 1: NNS planned	1.14	1.18	0.78
Significance	.05	ns	.01

Study 6: $N = 31$; Study 1: $N = 29$.

and for two of the non-native speaker tasks, the personal and the decision-making. The native speakers show an effect of planning only on the most monologic task, where the opportunity to plan seems to equip them to draw upon less frequent lexis. There is also the point that these two tasks, given their monologic nature, are inherently more predictable, since there is less scope for interaction to take the conversation in unforeseen directions. Planning, as a result, can have a more dependable impact. Yet the narrative is the one task that does not show a significant difference for the non-native speakers, whereas here the more interactive tasks, especially the decision-making task, do see raised performance. The complexity of the narrative retelling, despite perhaps its push towards specific lexis, seems to have defeated the non-native speakers, who seem to have allocated so much attention to wrestling with the ideas that they could not mobilize any less frequent words. This is a curious result. In contrast, they do seem to have been able to channel planning time to using less frequent lexis in the more interactive or more familiar tasks. It seems as if these tasks are within their abilities to a greater degree, and there is enough spare capacity available to enable them to retrieve less frequent vocabulary items.

Discussion

It is striking that the two lexical measures in this study do not correlate and are often affected by different things. The capacity to avoiding recycling vocabulary, and the capacity to inject vocabulary richness into performance seem to connect with different aspects of speaking. On this issue, as well, the congruence in results between native and non-native speakers is striking – the two measures do not relate for *either* group. One might think that factors like having a greater vocabulary stock which is more organized and more accessible ought to be a strong fundamental influence. Although this does seem to account for the performance differences between native and non-native speakers, that is as far as it goes. Elsewhere, different patterns for the two measures are more salient. These results are consistent with studies by Daller and Xue (2007) who report a correlation of .21 (non-significant) between the two measures for a group of 50 Chinese learners of English doing an oral picture description task; and by Daller and Phelan (2007) who report a correlation of .39 (again non-significant) for essays written in an EAP (English for academic purposes) context. In contrast, Malvern, Richards, Chipere and Durán (2004) do report a significant correlation (.42: $p < .001$) for a large sample of L1 British children writing narratives at Key Stages 1–3. It is possible, however, that written material may be associated with higher levels of correlation, although even here, .42 could not be regarded as a very strong level of relationship.

If then the influence of a larger mental lexicon, while important, does not account for many aspects of the results, principally the lack of

relationship between lambda and *D*, we need to ask what other factors are at play. First we have the issue of unavoidable lexis. We have seen that the narratives, in general, lead to the highest lambda scores. In narratives, the 'task' is strongly input-driven, and task fulfilment requires engaging with the material which is given. In a sense, therefore, what needs to be said is non-negotiable. This seems to push participants, native or non-native speaker, into retrieving the less common words which are implicated in the task. This influence does not seem to impact upon *D* to the same extent. Further, interactive tasks, although they make some lexical items salient, seem to allow participants freedom to express themselves without necessarily retrieving these key items if alternative means of expression can be found. There may also be the issue that interactivity can tolerate some degree of vagueness and generality, because speakers anticipate that, if necessary, further interaction can resolve misunderstandings. Narratives, in contrast, may put pressure on the speaker to be more precise and find more exact phrasing.

A second possible factor concerns a tension between interactivity and predictability. Interactive tasks produce higher values of *D* (i.e. less recycling of words). Clearly, within an interaction there is unpredictability, as a conversation takes the course that it takes. There is also the issue that turns are shorter, usually, and speakers may, as part of what they say, take account of interlocutor needs, including processing needs. As a result, their speech may be more involved and less detached, with the result that a speaker does not focus so much on their own contribution in isolation but may try to incorporate things said by their interlocutor. The result may be that their own speech draws on this interactive input, and as a result pushes up the values of *D*. In contrast, non-interactive tasks are more likely to put the speaker into a detached, long-turn, self-sufficient mode, leading to more recycling because there are not so many external influences.

There is a third influence which comes into play here, and unfortunately there is something of a confound involved. The monologic tasks tended to be about one thing, for example, narrate a story or describe how to get to your home. In contrast, the interactive tasks tended to have shifting topics, such as judgements on a series of crimes, advice for a series of letters to an agony aunt. It may well be the case that there is a strong topic effect on *D* scores. The arrival of a new 'crime' to discuss, or letter to advise on, may trigger the use of new sets of words. As a result, there is less recycling, but this is an artefact of topic change, rather than an inherent feature of interactive discourse (although of course much interactive discourse does have such topic change as an entirely natural component). So the result is that we cannot distinguish here between interactivity and topic change as possible influences upon the *D* scores which were obtained. Further, more focused research designs are needed to explore this issue.

We turn next to attempt to relate the results to a wider model of speaking, whether of first or second language. The Levelt (1989) model of speaking proposes three stages, the Conceptualizer (whose output is the preverbal message, and which essentially is concerned with the conceptual content and packaging of what will be said), the Formulator (which accepts the preverbal message, and which then engages in processes of lemma selection and consequent syntax-building processes), and the Articulator. Focusing on the first two of these processes is illuminating in relation to the present results. One can propose, from the results presented here, that lambda, and lexical sophistication, relate more to the Conceptualizer stage of the Levelt model, and to the nature of preverbal message implications for lemma retrieval. This applies particularly clearly to the native speakers whose lexical systems are richer and more organized, and who therefore can handle the processing implications delivered by the Conceptualizer and integrate lexis effectively to realize the demands that are being made on the Formulator. For them the correlation with (syntactic) complexity is another reflection of the way their syntactic performance can be effectively lexically driven. This is easier for native speakers, and problematic for non-native speakers. It is also interesting that for non-native speakers, more demanding Conceptualizer operations have bad implications for accuracy. Heavy lexical demands on Formulator operations impair parallel processing (in which the Formulator currently works on previous Conceptualizer operations while the Conceptualizer gets on with new work), since the difficulties experienced by the Formulator have attentional implications which spill over and influence the Conceptualizer. The result is a need for the second language speaker to engage, laboriously, in serial operations (Kormos, 2006).

In contrast, it is hypothesized that lexical diversity, as indexed by *D*, is more clearly a Formulator factor, perhaps shown by the correlation between *D* and accuracy. The Formulator is concerned with online, moment-by-moment decisions during speaking, but within certain parameters. Digging deep, and retrieving unusual lexical items is not the emphasis (since these are the province of the Conceptualizer and the preverbal message it delivers). Making surface-level choices is, and so the attention available seems to be concerned with using less demanding words more effectively. It is as if restricted sets of words prime one another, and once available, can be integrated more easily, and help avoid the need for wider, and more disruptive lexical retrieval.

A final thought concerns the applicability of the Levelt model to non-native as well as native speakers. Kormos (2006) argues that it is extendable, and at a general conceptual level, this must be so. All speakers need to organize thought and then marshal linguistic resources to express their thoughts. But the Levelt model, in the native speaker case, operates with certain assumptions. These are that what the Conceptualizer delivers to the

Formulator is 'fair game', in the sense that unreasonable demands, given the size and organization of the mental lexicon, are not being made, and so the Formulator can deal with these demands, in real time. Native language lexicons are rich, comprehensive and well organized, permitting lemmas to be accessed, and, when particular lemmas are problematic, enabling substitutions to be made (Pawley and Syder, 1983). Now it has to be said that the non-native speakers in all the studies that the present chapter is based on were intermediate level, and low rather than high intermediate. Any claims which are made have to be restricted to this group. But it is clear for these second language speakers that the sorts of relationship one finds for native speakers do not apply clearly, as in the case of the lambda–complexity correlations. The interpretation seems to be that the second language mental lexicons on which they draw are not as extensive or as organized, and that this has major implications for the transferability of the Levelt model to this case. These learners, one assumes, have a Conceptualizer stage which is potentially as effective as that of the native speakers, but the preverbal message it delivers to the Formulator makes demands that the more limited Formulator cannot meet. The speaker is then in a race against time, as s/he wishes to produce more language, but is still wrestling with the implications of the previous Conceptualizer preverbal message. As a result, strategies of communication become more salient, including avoidance and abandoning messages. It seems clear, therefore, that additional influences upon performance, such as task characteristics and also performance conditions, may be more important as they can ease the processing burden in ways which make the Conceptualizer–Formulator connection less troublesome.

Conclusions

There has been a strong exploratory quality to the research reported in this chapter, since there is relatively little published material on relationships between measures of lexical diversity and lexical sophistication; on lexical comparisons between native and non-native speakers; and on lexical measures related to a variety of task genres completed by the same participants. The emphasis therefore has been on the presentation of data on each of these points. But clearly now the need is pressing for there to be additional research which attempts to resolve some of the puzzles identified here.

First, there is a need to gather data, using measures such as *D* and lambda, with second language learners at different, and especially higher, proficiency levels. Such data can be very revealing about how the Levelt model becomes appropriate, without modification, as higher proficiency levels are achieved. Second, we need much more research with lexical measures specifically comparing different genres. Most research to this point has been based on only one type of task, rather than on a comparison of different

tasks as was the case in the present research. Third, we need to explore more systematically the variables identified post hoc such as topic change, unavoidable lexis, and interactivity. Such research will establish whether plausible interpretations are indeed convincing.

Note

1. The author would like to thank Brian Richards and an anonymous reviewer for their very helpful comments on a previous version of this chapter.

8

Can Differential Processing of L2 Vocabulary Inform the Debate on Teacher Code-Switching Behaviour? The Case of Chinese Learners of English

Ernesto Macaro, Tao Guo, Huili Chen and Lili Tian
University of Oxford, UK

Introduction

For a number of years second language acquisition researchers at the University of Oxford have been concerned with establishing 'optimal' (or 'principled') first language (L1) use by teachers in the second language (L2) classroom. This research is set against a background of a growing international interest in code-switching behaviour in second language classrooms in general and the extent to which this behaviour reflects, or is similar to, the well-documented code-switching behaviour in naturalistic environments. The latter type of code-switching occurs in non-instructed contexts where bilinguals, in a single conversation, communicate by using more than one language or dialect. Much of teacher code-switching behaviour appears to centre around unfamiliar or unknown L2 lexical items (Canagarajah, 1995). Although naturalistic code-switching encompasses far more than lexical switches, lexical items which are communicated in the 'embedded language' (Myers-Scotton, 1993), for a variety of reasons, are a strong feature of the switching patterns of bilinguals (Muysken, 2000).

In this chapter we first set the scene with regard to the teacher code-switching debate. We then move on to summarizing a number of studies which have been designed to inform that debate by exploring theories of vocabulary acquisition through the lens of teacher code-switching. In bringing these studies together we are proposing the notion of 'teacher as dictionary and dictionary designer'. We have chosen to contextualize these studies and this notion in the relationship between two languages, Chinese and English, two languages which are non-cognate and have different writing systems. The significance of these contextualizations will, we hope, become clear as the chapter progresses.

Origins of the 'code-switching in classrooms' debate

The debate as to whether teachers should use, or allow learners to use, the learners' L1 during L2 lessons has arisen as a result of three historical developments which it will be necessary to describe in order to understand the origins of the controversy.

Communicative language teaching

The first historical development is the phenomenon known as communicative language teaching (CLT). Space does not allow an in-depth presentation of CLT (for a thorough description see Richards and Rogers, 1986), but generally it is accepted that in CLT-type classrooms the emphasis is on communicating meaning through the L2 rather than examining the L2 by, among other things, making comparisons with the L1. As a consequence, for the teacher to use, or to allow students to use, the L1 seems to conflict with the basic premise of the CLT teaching approach. The argument goes like this. If a teacher wants to communicate information which is not related to the language being studied (e.g. 'those students who wish to avoid having their mobiles confiscated will they please switch them off now!') and does so in L1, s/he is: (a) missing the opportunity to develop listening skills and reducing opportunities for incidental vocabulary acquisition and; (b) not 'mirroring' the situation the students might face in an L2-only environment outside the classroom. The problem with this argument is: what is the teacher to do when communication breaks down? Is s/he to continue to attempt L2 communication by lengthy input modification (e.g. paraphrase, circumlocution) or provide L1 information of certain words/phrases until the breakdown is repaired?

An issue also arises when the teacher wants to communicate information which is directly related to the language being studied (e.g. 'we say, *I wonder if you wouldn't mind,* when we don't want to offend or upset people'). If a student fails to understand the target phrase, should a teacher draw attention to the functional and/or lexical equivalent in the L1 or arrive at its pragmatic force by inference? To do the former more likely ensures accurate comprehension. However, frequent reference to L1 information risks turning the interaction into a 'translation' activity, undermining the very principle of CLT. Therefore, the first contextualizing principle for the code-switching debate is that it has perforce to take place in reference to 'broadly communicative classrooms' not 'grammar-translation' classrooms.

The non-native speaker teacher

The second historical development that has influenced the debate is the way that researchers and practitioners have valued the native-speaker teacher (NST) versus the non-native speaker teacher (NNST). Throughout much of the 1970s and 1980s, partly through the promotional work of

the British Council, the NST was seen as the gold standard embodying all the advantages of the 'perfect command' of both target language and target culture. This value attribution has, in the past decade or more, changed considerably. Once again space does not allow a full discussion (see Llurda, 2005; Medgyes, 1999), but NNSTs are now recognized as having essential knowledge about the learners and their L1 which NSTs may not have. We say 'may not have' because the NST may undergo a period of change in terms of linguistic competence. Lengthy residence in the country of the learners (or language study prior to residence) both changes their grasp of their L1 and develops their knowledge of the learners' L1 thus providing them with insights into the learners' potential learning difficulties. Therefore, the second contextualizing principle for the code-switching debate is that the teacher must have sufficient 'access' to the learners' L1 for theorists to even begin to debate a possible optimal use of code-switching.

Naturalistic code-switching

The third historical development is the increase in the number of settings in which naturalistic code-switching occurs. Traditionally, in bilingual communities, code-switching has been considered the practice of alternating between two languages as a function of communication. There are a number of principles underlying this switching: it is normally accepted that one language is the dominant language (the 'matrix language', Myers-Scotton, 1993) and the other is the embedded language; that switching can take place intra-sententially or inter-sententially; and that the grammar of either language is rarely violated, or at least the switching is severely constrained by the grammatical properties of both languages (Muysken, 2000). In these communities code-switching is considered to be a bilingual competence, not a symptom of language deficiency (Poplack, 1980; Wei, 1994), and one of a series of communication strategies through which meaning can be expressed. However, naturalistic code-switching has increased as a result of widespread use of the Internet for communication, the globalization of manufacturing and services, and the hugely increased movement of peoples. As a result of these global trends the purposes and settings for which we need a second language have gone beyond the bilingual community and, increasingly, naturalistic conversations are being conducted in which code-switching plays an important part (Chan, 2004; Li, 2000; Wei, 1994). Thus, just as some might have argued in the past that the CLT classroom was a preparation for a monolingual environment, we would now argue that the CLT classroom of the twenty-first century is a preparation stage for multilingual environments (see also Cook, 1991, 2001 on 'multicompetence'). However, to what extent can the L2 classroom be said to be a preparation stage for this changing linguistic situation? Studies of L1 use in the classroom that have adopted this perspective are still very few. For example, Arnfast and Jørgensen (2003) demonstrated how code-switching patterns developed in

American learners of Danish. In the early stages of language proficiency, it was essentially a compensation-type strategy. As their proficiency increased, the patterns resembled much more the code-switching behaviour found in natualistic settings. This progression, from code-switching to compensate for language deficiency, to code-switching as a tool for optimizing communication, can serve as our third contextualizing principle. That is, optimal use of teacher code-switching should be considered not only in terms of helping students to learn more language but also in terms of helping them to use language more effectively in real contexts.

In the work that we have been carrying out at Oxford, we have so far only focused on how code-switching may help with the language learning process. Readers interested in classroom code-switching as a developing linguistic competence could also turn to the work of Liebscher and Dailey-O'Cain (2005).

Code-switching in the L2 classroom

A number of papers have appeared in journals and books in which the authors espouse contrasting positions on teacher use of L1 and to some extent these are based on the three contextualizing principles outlined above. See for example Chambers (1991), Halliwell and Jones (1991) and Krashen and Terrell (1988) for those arguing in favour of (near) exclusive use of L2, and Auerbach (1993), Butzkamm (1998), Celik (2003), Macaro (1997, 2000, 2005) and Turnbull (2006) for those advocating some principled use of L1. There have also been government guidelines issued about limiting L1 use or 'maximizing L2 use', for example in the UK (Department of Education and Science, 1990), in France (Ministère de l'éducation nationale, 1993) and in China (Ministry of Education, 2001). We now turn to a summary of *empirical* studies which have attempted to inform the debate.

Empirical studies on L1 use

Research on classroom code-switching behaviours (or L1 use more generally) has drawn essentially from two sources of data. The first is the reported beliefs of the participants (usually teachers, but occasionally both teachers and learners), these beliefs often being coupled with their self-estimated L1 use (e.g. Franklin, 1990; Kharma and Hajjaj, 1989; Levine, 2003; Liu, Ahn, Beak and Han, 2004; Macaro, 1995, 1997). In all of these studies (a number of which are reviewed in Macaro, 2000) the majority of the teachers surveyed believed some L1 was admissible or even necessary. The second source of data is quantity of L1 use by teachers through direct observation (Arnett, 2001; Duff and Polio, 1990; Guo, 2007; Kim and Elder, 2005; Liu and Jiang, 2004; Macaro, 1997, 2001; Macaro and Mutton, 2002; Rolin-Ianziti and Brownlie, 2002). The most important finding from this collection of studies is that there is a wide range of teacher L1 use both

inter- and intra-practitioner. Additional findings are that: practitioners are not necessarily influenced by principled beliefs or driven by theories in their choice of when and for what purposes to code-switch; different functions for L1 use have been identified across educational contexts; one of the most important (or at least most frequent) functions to which L1 use is put is to communicate the meaning of new or unfamiliar lexical items for the benefit of the learners.

There are two research implications, pertinent to this chapter, resulting from the above summary of research. The first is that the findings are not, in *most* cases, clearly set in a 'broadly communicative' teaching context. In fact, in most reports, we simply do not know whether teacher beliefs or (more importantly) quantity of L1 use were being investigated in the context of a communcative environment or simply the result of classes in which translation figured prominently as a pedagogical activity. Secondly, in very few of these reports is there any kind of examination of the effect of L1 use on L2 learning. Both Macaro (2001) and Guo (2007) report that there was little or no correlation between quantities of teacher L2 use and learner L2 use. However, despite these insights, we have virtually no hard evidence that teacher use of L1 has a positive effect, a negative effect, or even no effect at all on language learning. This situation has led to a number of researchers calling for more research into 'optimal' use of L1 (Macaro, 2001; Levine, 2003). That is, where limited code-switching, in broadly communicative classrooms, can enhance L2 acquisition and/or proficiency better than L2 exclusivity.

Before going on to examine some of the studies (at the University of Oxford) which have attempted to answer this call in the context of vocabulary acquisition, it is necessary to establish two things. Firstly, the notion of 'teacher as dictionary and teacher as dictionary designer' (see Macaro, 2005) and, secondly, some understanding of the differences between the Chinese writing system and the writing system of English. We will therefore need a considerable detour.

The teacher as dictionary and dictionary designer

It is well established in the vocabulary acquisition literature that learners learn new lexical items both incidentally and intentionally (e.g. Hulstijn, 2001). Such authors also provide fairly strong evidence that, when we build in the time it takes to acquire a range of items, intentional learning is more effective although not necessarily more enjoyable. A number of studies which have measured vocabulary growth as a result of reading have done so by either providing glosses somewhere near the text (Bensoussan, 1983; Davis, 1989; Jacobs, Dufon and Hong, 1994) or by allowing students to look up words in a dictionary (Knight, 1994; Laufer and Hadar, 1997). Few studies have compared whether L1 or L2 glosses, or whether monolingual or bilingual dictionaries, are better (but see Wingate, 2002). Thus, although

some questions remain unresolved, there is little doubt that how to 'focus on the form' of a lexical item (using Laufer's 2005 expression), when reading an L2 text, has been of interest to researchers. However, researchers have shown very little interest in 'how to focus on the form of a lexical item' in the *teacher's discourse*, where the teacher's discourse can be considered as text. By investigating focus on form in the teacher's discourse (as an activity akin to focus on form in the comprehension of other texts), we are led to explore how teachers operate as dictionaries.

In the case of NNSTs, teachers can operate both as monolingual and bilingual dictionaries. In fact, NNSTs can go further than this. Because of their knowledge of both languages, and (at least some) knowledge of their learners, they can make informed decisions about not only which language (L1, L2 or both) to provide the information in, but also what type of information to provide. In other words they can be considered as 'dictionary designers'.

So, what type of information is (theoretically) available to the NNST that might prove useful to the L2 learner? We would argue that, at the very least, it includes the information provided by the 'options' in Figure 8.1. A few words about the terminology used in Figure 8.1 may be helpful.

A definition is 'a statement of the exact meaning of a word' (*The New Oxford Dictionary of English*: Pearsall, 1998), whereas a paraphrase is 'to express the meaning of (something written or spoken) using different words especially to achieve greater clarity'. A circumlocution is 'the use of many words where fewer words would do, expecially in a deliberate attempt to be vague or evasive'. Circumlocution in SLA literature does not normally have this pejorative sense. Rather it is seen as a way of resolving problems, when putting across the meaning of a lexical item or concept, when simple definitions and paraphrases have failed, or when learners are experiencing a linguisic and/or cultural knowledge deficit. An exposition is 'a comprehensive description and explanation of an idea or theory'. Examples of 'use' or 'contextualizations' are illustrations of how a word can be used, usually in a sentence. To contextualize a word is not only to provide its use in a phrase or sentence but also to give that sentence a real world context. An example of a teacher providing a 'hierarchical exemplification' would be: 'reptiles? … you know … crocodiles, turtles, lizards, snakes?'

In Figure 8.1 we adopt the De Bot (1992) adaptation of the Levelt model (1989, p. 9) of conceptual and lexical stores where the Conceptualizer (the learner's schematic architecture) is non-language specific whilst the lexical stores are language specific. However, as De Bot points out, this is a steady-state model, not a developmental one. If a developmental dimension were to be introduced then we would have to recognize that the learner's Conceptualizer, whilst being non-language specific, would have been largely moulded via its contact with the L1 culture and personal experience, but that this might change over time. Thus an abstract concept which only

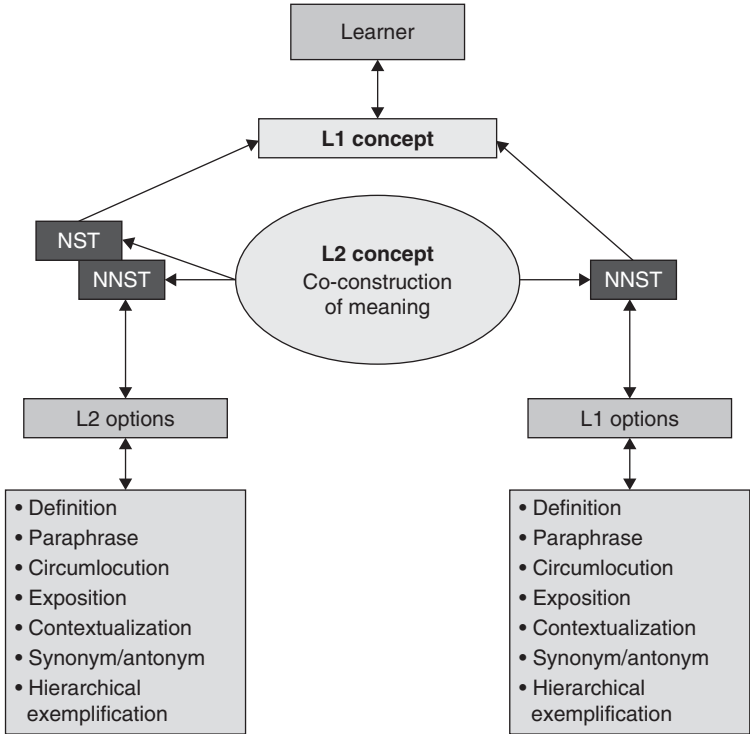


Figure 8.1 Teacher as dictionary

exists in an L2 culture has to be accommodated into the learner's schema through language rather than experience. In many classrooms it is the teacher who attempts to mediate this process of accommodation, as we hope to show in our data below. In other words meaning, and particularly the meaning of abstract concepts (or low-frequency concrete concepts), is arrived at through a process of 'co-construction'.

All the 'L2 options' in Figure 8.1 can be provided, theoretically, by both an NST and an NNST. However, an NNST can provide all the above information in the L1 as well! Moreover the NNST, as suggested in the diagram, can make a reference to cultural schemata as represented by the concept–lexeme mapping in both languages. The problem facing the 'teacher as dictionary and dictionary designer' is which and how many of these options to clarify a lexical item s/he should take up. Clearly the more information is given, the greater the likelihood of thorough understanding of the item (and possible retention). However, in CLT classrooms the pedagogical purpose of an activity is rarely limited to vocabulary understanding and

vocabulary learning, and time taken up with a focus on lexical form–meaning relationships is time taken away from other purposes.

The writing system of Chinese

The Chinese language uses a logographic writing system. That is, in theory, each symbol (or character) represents a concept. The concept can be either a 'content concept' (as in 'tree', 木) or a 'function concept' (as in 'more than one', 多). In terms of how the system operates to convey meaning, the closest equivalent of a Chinese character to English is, therefore, the morpheme. However, the equivalence is by no means a perfect one. Although a character always represents just one syllable, some characters contain 'bu shou' (部首), which literally means 'the parts of the character'. Some characters cannot be separated into different component parts. For example, 人 (pronounced 'ren' and meaning 'person'), of itself represents just one syllable and cannot be further divided. However, other characters are composed of several parts. The meanings of these different parts combine together to form the meaning of a concept. For example, the character 休 ('xiu') means 'having a rest'. It contains two parts: 人 ('ren') meaning 'person', and 木 (mu) meaning 'tree'. A person by the tree means 'taking a rest'. Therefore, as we can see, a character is not the smallest unit of meaning and it is not equal to the concept of morpheme in English. Nevertheless, we can loosely conceptualize Chinese as being structured around morphemic units and, as Bassetti (2005, p. 339) has argued, 'the central role of the hanzi [the character] obfuscates the role of the word for Chinese natives in probably the same way that the central role of the word obfuscates the role of the morpheme for English natives'.

A logographic system carries with it implications for absorbing new words, especially abstract words emanating from different languages and cultures. In order to convey the meaning of some words in English, Chinese has sometimes to create 'a phrase' which can consist of many characters. For example, although a word such as 'to commute' could nowadays be expressed with only two characters (通勤), in the past it was expressed as a phrase (乘汽车、火车或开车经常往返于工作与住所之间). This is because, although there might have been a 'commute' concept available to a Chinese person, the concept itself was very low frequency and the two-character combination was not helpful in providing the true meaning. Thus a more detailed explanation was preferred.

In Chinese, when there is no existing character combination to express the meaning of an English word, a translation of the word's *English* definition is often adopted. Later, if the popularity of the concept grows, a need may arise to create a new word using logographic criteria or to use the phonological translation or equivalent of an English word as one of many possible strategies to tackle the non-equivalence in the translation of the two languages.

As we can see from the above, the mediating role (mediating between new L2 word and learners' understanding of that word) played by the NNST is considerably increased in the case of the Chinese teacher in the English language classroom. The teacher may well have to mediate the concept–lexeme representations in order to overcome the obfuscating nature of one language's writing system. It is with this mediating role in mind that we have pursued a series of studies at Oxford.

Studies at Oxford that inform the code-switching debate

In the Meng (2005) study, carried out in mainland China, the aim was to explore whether teacher code-switching for communicating the meaning of unknown lexical items during reading comprehension led to better acquisition of those lexical items. Meng established via a pre-test that the target items were indeed not known to different classes of secondary school students and also established that their vocabulary knowledge and their general proficiency were not significantly different. He also ensured that the classrooms were broadly communicative in nature via prior observation. He then persuaded a teacher to teach differently to each class, thereby establishing three different conditions. In the first condition the teacher stopped the flow of her discussion about the text at appropriate moments, and provided L1 equivalents of a number of words including the target words (henceforth *L1 Equivalent Condition*). In the second condition, the same teacher stopped the flow of her discussion about the text and provided L2 definitions of a number of words, including the target words (henceforth *Definition Condition*). In the third condition, the same teacher stopped the flow of her discussion about the text and provided definitions, followed by L1 equivalents (henceforth *Definition Plus L1 Equivalent Condition*). In this condition, the teacher contextualized the word in a phrase, in the same way as a monolingual dictionary might. For example the teacher said , 'you know the word meadow? It's a field covered in grass. For example, this Sunday we are going to have a picnic in the *meadow* in front of the house.' Two different texts were used on two separate occasions. Immediate and delayed post-tests of vocabulary were administered to the students.

Surprisingly, results suggested that the different teacher approaches to these new lexical items had little or no differential effect on vocabulary learning. In the case of the first text (about the ubiquitous nature of sport), small but significant differences in favour of the Definition Condition were found at immediate post-test. These disappeared at the delayed post-test. In the case of the second text (about the life of Walt Disney), there were no significant differences for any of the conditions either at post- or delayed test. These results suggested two things. Firstly, limited teacher code-switching did not appear to be detrimental to the learner's potential for learning vocabulary whilst at the same time (probably) ensuring better

comprehension of a complex L2 text. Secondly, it was clear that the methodology used was not sensitive enough to answer fully Meng's research questions. This original finding led to a series of further studies, some of which are still ongoing.

As in the Meng study described above, the teaching approach in a study by Tao Guo (see Guo, 2007) was established before carrying out the study. The participants were two teachers and their students in two different Chinese universities. The Communicative Orientation of Language Teaching (COLT, see Fröhlich, Spada and Allen, 1985), an observation coding scheme, was used in order to ensure that the lessons of the two teachers were indeed broadly communicative. This study was 'naturally occurring' not 'researcher manipulated' as in the Meng study, but it similarly was centred around the comprehension of a variety of reading passages. Guo collected a large amount of data via video-recording – over 700 code-switching episodes over 16 lessons. In the study the most pertinent questions for the argument we are developing here were: what types of lexically related code-switches were there in the corpus of data, and what was the students' strategic reaction to these switches? In other words, when the teachers were providing different information types about new words, what was going through the minds of the students? The students' strategic reactions were obtained through stimulated recall, that is, by viewing a selection of the video-recordings and being asked to recall what they were thinking at the time of the teacher code-switch.

Let us now see if we can identify the notion of 'teacher as dictionary and dictionary designer' in examples taken from the Tao Guo code-switching corpus. In the corpus there is a clear majority of episodes where the teacher opts immediately for a code-switch without first taking up any L2 options. In Example 1, as in the majority of the examples below, the teacher is working his or her way through a written text where the focus is essentially on the general meaning of the text.

Example 1 ('assemble')
Teacher … Assemble in the crowd. Assemble means 聚集 [*tr: assemble*]. Give me another word for assemble. (looking around). You know this word. Of course … everybody does!
S1 Gather
Teacher Gather. g-a-t-h-e-r. (students repeat the spelling in chorus)

Here the target word is provided in Chinese almost immediately after its occurrence in the teacher's speech stream and, we may surmise, the teacher reassures the students that the L1 equivalent is a direct and precise one through the use of 'assemble means'. Considered against Figure 8.1 there is, apparently, complete overlap, in both languages, between the linguistic representations of the concept 'to assemble' and the underlying

concept itself. However, note two things. Firstly, the teacher attempts to elicit 'another word for *assemble*', perhaps with the purpose of exploring the mental map some students may already have. However, having triggered the form–meaning relationship in Chinese (聚集), the students are presented with four elements (i.e. the form–meaning relationship in *both* languages). Secondly, we should note that the equivalent is expressed in Chinese by two characters. We shall be returning to this last point later in the chapter.

There are many examples of code-switches where the episode is combined with a definition, or with another of the L2 options in Figure 8.1. These can occur prior to the code-switch or as in the case of Example 2, after the code-switch 'make up part'.

Example 2 ('constitute')
Teacher … Right, you know these three paragraphs constitute, constitute 构成 [*tr: constitute*] make up part two … so, what is the main idea expressed in these three paragraphs? …

Note also that in this example (as in many others) repetition of the target item occurs before the code-switch as a signal of a change of focus from meaning of the text to focus on the form–meaning relationship of a single lexical item – presumably to prepare them for the change of focus and/or the change of language. This assertion is given further credence by the students' reactions, as we shall see. Example 2 contains a more implicit way of making salient the lexical item being focused on than the explicit request to notice the phrase 'thanks to' (Example 3 below), where the teacher states clearly 'another phrase you need to pay attention to'.

Example 3 ('thanks to')
Teacher … another phrase you need to pay attention to is, thanks to. Thanks to means because of, due to good reasons, positive reasons, you know, in Chinese, 幸亏 [*tr: thanks to*]. Ok so, if you have some negative reasons, you should not use 'thanks to'…

In a quite different example (Example 4), the teacher consolidates any possible inferencing that the students may have attempted from the text ('in paragraph 1, one new word is career') by drawing a distinction between 'career' and 'job', thus focusing on the form–meaning relationship of the specific lexical item.

Example 4 ('career')
Teacher … In paragraph 1, one new word is career, so you should know that a career means a job or a profession that one is trained in and intended to have for a long time. So, a career is different from a job.

Ok? You have to engage in a career for a long time. They say now in China, there are more and more career women than before.

S1 职业女性 [*tr: career women*].

Teacher Ok, very good. So, career women are like the opposite of housewives, right? So, career women are like the opposite of housewives, right, 职业 女性, 职业女性 [*tr: career women, career women*]. Career means 事业或职业 [*tr: occupation or profession*] ok, 事业或职业 [*tr: occupation or profession*]. And the next phrase is …

However, rather than providing an immediate L1 equivalent she first provides a (partial) definition ('profession intended to have for a long time'), before giving a contextualizing phrase ('in China, there are more and more career women than before'). Immediately after, she accepts a student's provision of an L1 equivalent, before providing firstly an approximate antonym ('opposite of housewives') and then her own code-switch for career as 'occupation or profession' where the five characters contain a notion of 'length of time'. We do not know why this teacher here combines a number of options, nor the fact that she starts off with L2 options before consolidating with L1 options. This contrasts with information about 'via' (Example 5) which is provided by an immediate code-switch before a definition 'by way of'. It is possible that the greater attention given to 'career' is because of its semantic richness and a conceptual difference between the two languages.

Example 5 ('via')

S1 What does 'via'mean?

Teacher (spells the word) v-i-a, v-i-a, via means through, in Chinese, 通过 [*tr: through*] … via means by way of, by way of (She writes 'via' on the board) … v-i-a. Any other questions? …

Another interesting example is provided by clarification of the word 'ordeal' in Example 6. Here the teacher first attempts a definition ('ordeal means a hard task, or big problem').

Example 6 ('ordeal')

Teacher So, the first paragraph said … whether for a college or for a job, most people face this ordeal (referring to an interview). Ordeal means a hard task, or a big problem, ok? Ordeal, 严峻的考验或者是何种麻烦 [*tr: a serious challenge or problem*] (she resumes reading the text aloud) However, people who prepare for interview …

In Example 6, the teacher is indeed faced with a hard task. It is impossible to find a two-character equivalent for the abstract word 'ordeal'. A direct equivalent for 'ordeal' using more than two characters is difficult

to find and there would likely be disagreement among native speakers of Chinese as to the best multi-character rendering of the meaning of ordeal (we should note, in passing, that the English word derives from a Christian concept of a trial by pain as divine proof of innocence). The teacher offers a 12-character code-switch, perhaps because she does not feel confident that her L2 definition has led to real understanding of the concept. We should observe, however, that the code-switch is merely a definition attempt rather than a full-blown exposition.

A similar lengthy code-switch (24 characters) occurs in Example 7 when the teacher is trying to put across the meaning of 'within one's reach'. Here she first provides a Chinese contextualizing phrase (line 2). She then simply repeats the L2 phrase and then provides an L2 contextualizing phrase which is more or less a direct translation of the previous Chinese version. We should note that, within the 24 characters, the sequence of 11 characters 在我力所能及的范围之内 is a fairly close equivalent for 'within one's reach'. Thus, whereas in English we have three fairly high-frequency words (or five morphemes) making up a not very transparent collocation, in Chinese the concept is transparent in the sense that it is not metaphorical.

Example 7 ('within one's reach')

Teacher … ok, 那么这件事情我办不了，这个不在我力所能及的范围之内, 你可以说 [*tr: well, I can't help you on that as it is not within my reach, you may say*] … It is not within my reach, ok? I cannot help you because for this job … it simply is not within my reach …

There are some examples of lexical items which are taken beyond the provision of meaning via definition or paraphrase. These are where the items represent a complex idea which needs an exposition or elucidation. Such an example is number 8 where the collocation 'role model' appears in the text being read. The concept of role model is well matched in both languages and the meaning rendered easily by two characters.

Example 8 ('role model')

Teacher … And he was my dad and in my eyes he was my role model. (teacher finishes reading from the text). So, he respected him, he loved him and he even admired his dad. In a word, his dad was his hero. His dad was his role model. (looks round for a few seconds), role model? (writes 'role model' on the board), His dad was his role model, 榜样 [*tr: role model*] … so, we often say the parents are the first teachers of their children in childhood. That is very true, it's very true …

Here the teacher does not attempt a paraphrase but he provides an empathetic interpretation of what the original text was offering ('he respected

him, he loved him and he even admired his dad … was his hero … we often say the parents are the first teachers of their children …'). Here an idea or concept is being developed in L2 which is some way beyond the meaning of the original phrase although the development is sensitive to the context in which the original item was placed. Yet, he also provides a brief code-switch with the equivalent in Chinese perhaps to show that there is direct equivalence in both languages. From the transcription we note that the teacher looks around for signs of comprehension before writing the L2 item on the board and then providing the code-switch. Perhaps he felt that the exposition may have gone beyond the simple meaning of 'role model'. There appears to be an attempt here to remind the students that there is no cultural gap between the English concept behind the lexical item and the Chinese concept.

This is in contrast to Example 9 where the code-switch occurs before the target item is even uttered. Here we can interpret its function to be as close to a naturalistic code-switch as we have seen so far in the data.

Example 9 ('collective')

Teacher … so, you know Chinese culture, first, we talk about the nation, as a whole, right? Then we talk about 集体 [*tr: collective*], collective (some students read aloud the word 'collective' spontaneously) then we talk about individuals, right?

It is interesting that here the *Oxford English Dictionary* (which defines 'collective' as 'done by people acting as a group') does not provide the full meaning of 'collective' as intended in the transcribed text. This is because the two Chinese characters have a sociopolitical connotation not immediately triggered by its English equivalent. In English it is only when we collocate the word ('collective ownership') or use a derivation ('collectivism: the principle of giving a group priority over each individual in it') that we get anywhere near the cultural force behind the Chinese equivalent. The code-switch is thus naturalistic because it conveys illocutionary and cultural force denied the L2 equivalent. It expresses more precisely the teacher's meaning.

To summarize so far. The two NNSTs in the Guo study provide strong evidence that they are attempting to act as 'teacher as dictionary and dictionary designer'. They go beyond mere glossing of the meaning of the word, that is, providing a meaning specific to the text in which it occurs. They appear to be thinking about the information options available to them and in relation to: (a) their 'best guess' as to the students' potential for understanding; (b) the complex meaning–form relationships in the two languages; (c) broadening the students' understanding of potential meaning and future encounters with the word. When a decision is made to code-switch, we detect what appears to be a search for whether a direct L1

equivalent is available via two characters and, if not, whether paraphrases, contextualizations and expositions are going to best serve the understanding and (possibly) acquisition of the item. The availability of L1 equivalents is explored in the next study but first we should stay with the Tao Guo data and see what the students' strategic reaction to these switches is.

A number of themes emerge from the data collected via stimulated recall. One of these is the fact that most students were easily able to accommodate code-switches in the teacher's input.

Example 10 ('subtitles')
Teacher I strongly recommend you to watch English movies, English movies, in the original version, without, without seeing the subtitles in Chinese, 不要看中文字幕 [*tr: don't look at the Chinese subtitles*]. Ok, you can read the English subtitles …

Here the teacher provides an exact translation of a phrase regarding learning advice. Student 1, in reaction to the code-switch, remarked:

S1 Actually I can feel that she was slowing down her speech a bit when she was saying 'subtitles' in English … […] and my immediate feeling was she was about to repeat it in Chinese, because she knew we didn't know the word [……] … generally speaking, when teachers say one word slowly and with stress, they tend to repeat it in Chinese …

Thus, it appears that these brief lexical switches have become routine in the interaction. They are a recognized feature of the pedagogy, such that the students are not surprised by them. Rather, they recognize certain prosodic features of the teacher's L2 speech as a signal that a switch is about to occur, and their cognitive processing of the speech stream seems not to be disrupted. Although we are not aware of studies of code-switching in bilingual communities which have examined the *listener's* cognitive reaction to switches in the speaker's communications, we hypothesize that this kind of accommodation may well be the norm. However, accommodation is not what appears to be taking place when the form–meaning relationships in the two languages are complex, as we can see from Example 11.

Example 11 ('fruitful')
Teacher … Can you tell me the meaning of 'fruitful'?
S1 '成果' [*tr: ripened fruits; good outcome of an effort*]
Teacher Yeah, 硕果累累的有成果的 [*tr: abundantly {first 5 characters}, fruitful, fruitful*], ok, with research environment …

Before noting a student's strategic reaction to the code-switching related to 'fruitful' we should first consider its semantic properties in English where

its literal meaning (producing much fruit) is closely linked to its figurative meaning (producing good or helpful results). Reflecting back on the episode, Student 2 commented:

> *S2* When she was speaking English and translating her English into Chinese … I felt suddenly slightly confused … I wondered what the word meant … if she had spoken a Chinese word that was not linked in meaning to the preceding English word, I would not have been confused [……]. When … for example, she said '今天早上[*tr: this morning*]' in a flow of words and I thought about '今天早上[*tr: this morning*]'.

So what we have here is anything but accommodation resulting from the code-switch. Instead there is quite a substantial disruption or perturbation of the otherwise normal processing of the L2 input. Whether this leads to better opportunities for recall we do not know for sure but there is a suggestion that the student's strategic reaction is more intense and may have afforded deeper processing opportunities than an L2 definition alone. There is some evidence to support this assertion in that the student acknowledges being more confused when the L1 equivalent is provided than if a Chinese word or phrase had merely been inserted 'which was not linked to the preceding word'. Put differently, drawing attention to problems of form–meaning relationships may cause a kind of confusion which may result in deeper processing during the search for the meaning of the new word. If this is indeed the case, then what we have here is in effect an episode of focusing on form on the part of the student, even though the teacher has not broken off essentially from the communication of her more general message. She acknowledges the student's contribution in line 3 which provides both literal and figurative information in L1 but the teacher offers additional L1 information by providing two adjectives: 硕果累累的 and 有成果的 to match the English adjective 'fruitful'. The Chinese character '的' is an inflection for constructing adjectives. In addition, she confirms semantic equivalence for 'fruitful' in both languages, but adds the notion of 'abundance' (line 3). She provides evidence of some crossover in form–meaning connections in both languages.

We should now return to an earlier example ('ordeal') and to a new but linked example (Example 12). One student's reaction to the 'ordeal' code-switch is:

> *S3* I could understand her English explanation … [but] … I felt more sure about it when I heard her Chinese explanation.

Example 12 ('appearance')

> *Teacher* He knows how to influence and affect the audience by his words, his gestures, and his appearance, 表情啊, 语言啊 [*tr: appearance, words*], something like that, so we say …

To which a (different) student reacted as follows:

S4 Well, consciously or unconsciously … I just feel that when he's speaking English, I tend to focus on the meaning of his English … however, when he switches to Chinese, I can understand it all … as well as make quite a lot of associations in Chinese.

In both these examples there is ostensibly a positive cognitive reaction to the code-switches. Firstly, the L1 information provides reassurance about the meaning of low-frequency words such as 'ordeal' and 'appearance'. Secondly, it seems to trigger deeper processing of semantic links in working memory. In two linked examples (13 and 14), this type of cognitive reaction is again alluded to by three students' reactions, confirming that they themselves were converting the information into their L1.

Example 13 ('rationality')

Teacher So here intellectuals refer to people with what, with rationality 大家注意，西方人特别注重 reasoning … reasoning 带来的就是 rationality [*tr: attention everyone, Westerners are keen on* reasoning … reasoning *leads to* rationality] so what is rationality in Chinese? (looking around and then uttering) 理性 [*tr: rationality*]

Ss (repeating teacher in chorus) 理性 [*tr: rationality*]

S5 When he was speaking English, it probably took me a second to translate his English into my Chinese mother tongue. Then, when he was speaking Chinese, I didn't need to translate at all.

S6 This is the bit which I think the biggest differences between Chinese education and western education lies.

Example 14 ('rationality')

Teacher because rationality stresses intelligence, the ability to reason. 理性的东西更强调智慧 [*tr: rationality prioritises intelligence*]. So they value rationality and are interested in facts …

S7 At that time, I was listening to his English carefully before the switch … I wanted to know what the teacher was trying to say … I could only understand his English by translating it into Chinese.

S8 At that time I was probably thinking of matching the Chinese with the English … making a mental note of it, the meaning.

Note that the L1 information provided in these examples is not limited to a definition of 'rationality'. Rather it is expository, providing an argumentation about a (Western) concept. Further, note that the meaning of 'rationality' is offered only by an approximation of one of its constituents (reasoning) which itself is lexicalized through an 18-character exposition (in Example 13) followed by further 10-character exposition in

Example 14. In their reactions there is clear evidence of the students' need to accommodate the new information into existing schematic knowledge via L1 mediation and that mediation is being assisted by the teacher's selection of information about the target item. Rationality is a concept that may not be immediately evident to these university students and therefore not completely translatable by them without his assistance. Hence, the teacher's question, 'what is rationality in Chinese?', is not merely an elicit for an L1 equivalent. It appears to be a kind of 'comprehension check' in the deepest sense of the meaning of comprehension – adorned with cultural and philosophical information.

One last example of potential for increased processing is provided by Student 9 to a code-switch in Example 15. Here the literal meaning of 'scar' seems to have been understood by the student without difficulty but possibly not its metaphorical meaning as in the text.

Example 15 ('trauma')

Teacher Because he was hurt. The scar, the trauma 创伤 [*tr: trauma*], was left by the person whom he loved so much. So that is why sometimes love and hatred are just one step aside ….

S9 I understood 'scar' immediately, but the other I didn't understand at all. I was wondering what it meant, when Teacher said 'the trauma', then 创伤 [*tr: trauma*] … but I still couldn't figure out its spelling (laugh). Then I decided to check it after class.

In this example the teacher provides 'trauma' as an approximate synonym (in the metaphorical sense but not the medical sense) for the word 'scar'. The student appears to have been unable to make the link between the two synonymous *metaphorical* meanings in English. The Chinese equivalent of trauma (创伤), which has both metaphorical and medical equivalence in Chinese, enlightens him to some extent. However, he is unable to visualize its orthographic representation and claims to have checked it after the lesson. One speculation here might be whether Student 9 would have been able to retain the phonological shape of 'trauma' long enough in order to be able to look it up in the dictionary if he had not had the Chinese equivalent to hook it on.

In summary, we have abundant evidence here that the teacher taking on the role of 'dictionary and dictionary designer' is having a considerable effect on the students' cognitive strategic responses. We have fascinating insights into how these teachers are pre-emptively realizing problems that students might face with form–meaning mapping relationships in the two languages and the two cultures and they are using a panoply of linguistic devices to assist with deep rather than approximate or vague comprehension. Students' reactions are on the whole positive because they appear to realize that these episodes of focus on form–meaning–culture relationships

are triggering deeper cognitive strategies that might help them understand text at a more fundamental level and potentially remember the lexical items better. The NNST's ability to code-switch seems to be a very powerful component in that panoply of linguistic devices. We have evidence here (to our knowledge for the first time) for the often vaunted claim that NNSTs have greater understanding of the learners' problems (Medgyes, 1999).

Whilst analysing the Tao Guo data, however, we became increasingly aware of the differences inherent in the two writing systems and the impact that these differences have on our understandings of words and concepts. As proposed above (Bassetti, 2005), the English writing system foregrounds the word to the detriment of the morpheme; the Chinese writing system foregrounds the morpheme to the detriment of the word. In Bassetti's study (English L1) participants were invited to segment Chinese text into words and she found that they did so quite differently (i.e. separated them at different points) from Chinese L1 participants. We therefore wondered if the different conceptualization of language might influence the way that Chinese L1 speakers attempt to learn low-frequency English words. This was investigated in the study by Huili Chen outlined below (Chen, 2007).

In Chen's study, the investigation into optimal teacher code-switching for lexical items was pursued in terms of the type of L1 information that a Chinese teacher of English might provide his/her students. Given our earlier explanation of the Chinese writing system and its effect on the conceptualization of English words, Chen dismissed the simple notion of providing L1 equivalents, opting for what we might call 'information type in response to word type'. She therefore explored whether English words would be learned differently if they were:

(a) the type of word whereby L1 information could be given via only two characters – what we might consider as providing 'near equivalence' (more prototypically a word) or
(b) the type of English word whereby the L1 information could only be deemed sufficient via a phrase or circumlocution, as represented by a minimum of five characters (less prototypically a word).

The idea being pursued here was as follows. If different word types are learnt differently by Chinese learners of English, then teachers should take this into consideration in their code-switching behaviour. She therefore devised two sets of English words (controlling for word frequency and length) to be learnt by advanced users of English. These were students in the UK studying a subject other than the English language and who might indeed be faced with learning a number of low-frequency words in their respective academic subjects.

Examples of English words where 'near equivalence' was conveyed by two characters included:

canvass – 游说 v.
ferment – 发酵 v.
debacle – 惨败 n.

Examples of English words where no equivalent in Chinese could be found and which therefore had to be rendered by a phrase or circumolocution included:

suborn – 用行贿或其他方法唆使某人做非法的事 v.
regale – 以食物和饮料供人享用 v.
halter – 袒肩露背的连衣裙 n.

For the sake of convenience we will now refer to these two word types as 'equivalents' and 'circumlocutions'. Having controlled for a number of variables (e.g. English proficiency, vocabulary knowledge, length of residence in the UK), and having ensured that they had no prior knowledge of the words, she arrived at a sample of 20 mainland Chinese adult students and set them the task of learning 40 English words (20 'equivalents type' words and 20 'circumlocution type' words), each of which she had provided with appropriate L1 information. The learning task lasted 15 minutes. After a 5 minute rest period, each participant was asked to recall the Chinese for the target words using Superlab (a computerized stimulus presentation program, http//www.superlab.com). Two weeks later 15 of the 20 participants undertook a delayed test using the same procedure. The remaining five participants took part in a stimulated recall of their learning.

In a paired-samples *t*-test, Chen found significant differences ($p < .05$) of recall accuracy between the two groups of words at post-test (but none at delayed test, where many words had been forgotten). The equivalent words were more correctly recalled than the circumlocution words. However, when examining the accuracy of recall for the circumlocution words, Chen noticed that there was in fact more *partial* recall for these words than for the equivalent words. Participants were able to recall some of the information that had been provided for many of the circumlocution words whereas, in the case of the equivalent words, recall was all or nothing. This finding was corroborated by the students' verbal report data from the stimulated recall. Participants suggested that the information provided in the circumlocution had impressed them more deeply and that the short meanings were harder to retain.

The stimulated recall data also suggested that participants tackled the learning very differently. Some appeared to focus on the form of the English word, trying to relate it to some other English word or breaking the

word up into constituent parts (a possible influence of the Chinese syllabic conceptualization); others focused on its visual 'shape' and the affective response that this generated. There was a strong suggestion, therefore, that participants were transferring L1 learning strategies to L2 word learning. There was little or no mention of meaning-related strategies in the sense of creating sentences or stories with the words whether in L1 or L2. Whether this strategy transfer was something that should be encouraged by teachers or discouraged is hard to say given the small size of the qualitative sample and given that these were all, apparently, successful learners of English.

Some tentative conclusions

In this chapter we have attempted to provide an explanation as to why the overarching question of teacher code-switching is being investigated through the dimension of vocabulary acquisition. We have argued that switching for semantic reasons is prevalent in broadly communicative classrooms and the switching reflects an underlying complexity of form–meaning relationships in the two languages available to bilingual teachers and learners alike. These relationships are particularly complex in the case of Chinese and English because of their different writing systems, a difference which may mask different ways of conceptualizing the world.

Although our research agenda has so far not led us to be able to say what optimal teacher code-switching is, it has presented us with a number of important findings and further avenues to go down. Firstly we have the 'no effect' finding. So far we have no firm quantitative evidence that code-switching either enhances vocabulary learning or detracts from it. We are following up this early result with a number of partial replication studies as this finding runs counter to theories of vocabulary acquisition which suggest that stronger links are established in L1. Early results of one ongoing study by Lili Tian in fact suggests counter-evidence to 'no effect', in favour of the code-switching condition.

Whilst awaiting the full results of these follow-up studies, one conclusion we can draw from the 'no effect' finding (should it persist) is that brief switches do not appear to disrupt the flow of communication and may, indeed, speed up or 'lubricate' that communication such that the pedagogical objective retains the purpose of discourse or text access rather than where there is prolonged focus on form–meaning relationships.

Our verbal report data seem to suggest that learners mentally translate L2 information into L1 information regardless of whether the teacher provides it in L1. However, we do not know whether they translate it because they are encouraged to do so by a teacher's pedagogical approach which includes code-switching. A future study might usefully explore, through stimulated recall, whether learners convert L2-only information provided by the teacher (as would be the case with an NST) into L1 information.

Whether the NNST switches or the NST sticks to the L2, it remains to be seen whether the information given to the students is triggering the best possible combination of vocabulary-processing strategies in the minds of learners. Once again, this vocabulary processing may well be specific to the relationship between the first and second language, as in our investigation of Chinese learners of English. In this the teacher may have an extremely important part to play. For example, should teachers be providing Chinese circumlocutions of some English words, Chinese translations of English circumlocutions, or simply definitions in L2?

We are planning to continue with our code-switching and vocabulary research agenda as we believe that it is in the way that learners process information about new lexical items that we may find a key to unlock the door of what seems to be an intractable debate regarding L1 use in the L2 classroom.

9

Lexical Development in Instructed L2 Learners of French: Is there a Relationship with Morphosyntactic Development?[1]

Annabelle David, Florence Myles,* Vivienne Rogers* and Sarah Rule***
**Newcastle University, UK*
***University of Southampton, UK*

Introduction

Although a number of studies have documented French L2 development (e.g. Ayoun, 2007; Dewaele, 2005; Myles and Towell, 2004; Prévost and Paradis, 2004), most focus on one or more aspects of morphosyntax, and there is comparatively little research on lexical development in French as an L2 (but see David, 2008; Graham, Richards and Malvern, 2008; Milton, 2006; Richards, Malvern and Graham, 2008). Even fewer studies investigate the relationship between the development of morphosyntax and vocabulary and the role that the lexicon might play in morphosyntactic development. This relationship is of interest as it might have implications not only for language acquisition theory, but also for better understanding the relationship between the lexicon and syntax in natural languages. For instance, some recent theories of language suggest that the brain is modular and that the different modules (e.g. phonology, syntax and lexis) are independent from one another and do not interact with each other except at the level of interfaces (which are external to them) (Jackendoff, 2002). One of the implications for language acquisition is that the processes by which these modules are acquired might also be separate. Hence (second) language learners might develop their lexicon and morphosyntax independently from one another at different rates and at different times.

Some acquisitionists believe, on the other hand, that this view of language is fundamentally flawed and have proposed that the knowledge structures that comprise fundamental language domains are constructed within a unified developing system using a common set of learning mechanisms and resources (e.g. MacWhinney, 2001). For example, grammatical abilities (e.g. the production of inflectional morphemes like *daddy goed*) emerge over the course of building a lexical system (Bates and Goodman, 1999).

In our current state of knowledge, the relationship between the lexicon and morphosyntax remains unclear, however, and the study of second language development can shed some light on the modularity of the different language subsystems by investigating whether they develop independently of one another or whether their development goes hand in hand.

Background

L1 acquisition researchers have investigated the issue of the relationship between lexicon and grammar to a larger extent than L2 researchers. Richards and Malvern (2004) find a significant positive correlation between lexical diversity (measured with *D*) and mean length of utterance (MLU) in 38 children aged on average 30 months old. Bates and Goodman (1997) find a correlation between lexicon and grammatical complexity (as measured by parental questionnaires based on typical child productions). The reasons behind this apparent link are open to debate. Locke (1997) argues that in order for L1 learners to activate their Grammatical Analysis Module (GAM), which allows them to move from a stage during which they are producing lexical items in isolation to the next stage which is characterized by the production of two-word utterances, they need a critical mass of lexical items to have been acquired and stored during the previous stages. Locke (1997) mentions (based on Benedict's work, 1979) that the vocabulary burst, which takes place around the 50-word stage, is the trigger for the activation of the GAM. Therefore, one could conclude that a child needs at least 50 words in his/her productive vocabulary before being able to produce multi-word utterances (other than frozen or formulaic phrases). Some researchers have interpreted the co-occurrence of the lexical spurt and of the onset of word combinations as evidence that grammar learning is well under way before the vocabulary spurt. In other words, grammar facilitates lexical learning (e.g. Gleitman and Gleitman, 1992). Dixon and Marchman (2007) propose that the idea of a critical mass (of lexical items) triggering grammatical development does not mean that lexical and grammatical development are not simultaneously ongoing during the earlier periods of language acquisition. According to these researchers, the concurrence of the vocabulary spurt and early word combinations might be due to other factors such as the particular measures used or the form of knowledge being reported. It remains the case, however, that in the context of L1 acquisition, it is difficult to investigate the direction of cause and effect between the development of vocabulary, grammar, processing and cognition, as they co-occur in time. The L2 context, on the other hand, should make it easier to separate them, given the cognitive maturity of the learners.

These issues have not been studied as systematically in the context of L2 acquisition. Researchers have shown that different measures of vocabulary

knowledge do correlate positively with other aspects of language learning (writing ability, reading comprehension and grammatical knowledge). For example, Meara and Buxton (1987) claim that results of a yes/no receptive vocabulary test do correlate significantly with results on the Cambridge First Certificate. Milton (2006) analyses the learning of French L2 receptive vocabulary in secondary schools in Britain and concludes that in their last year of learning French (after seven years of French), learners know about 2000 words. He also shows a significant correlation between their grade at the final secondary schooling exam (A level) and the number of words they know. But L2 researchers have not, to our knowledge, examined the question of the relationship between these different types of knowledge nor its implications for current theoretical models of language acquisition.

One way in which second language research has been interested in the relationship between different aspects of language is within the larger aim of researching means of measuring and assessing students' language proficiency. Some measures of vocabulary proficiency have been shown to relate to proficiency in other areas of the language system. Unsworth (2005), for example, has shown positive correlations between 'lexical complexity' (as measured by Guiraud's index) and morphosyntactic complexity (as measured by verbal density) in children and adult L2 learners of Dutch. But Malvern, Richards, Chipere and Durán (2004), on the other hand, show that lexical diversity does not correlate with more general linguistic measures such as MLU in their study of L2 learners of French taking an oral exam (General Certificate of Secondary Education) at age 16. Similarly, David (in press) does not find any correlation between A level exam grades and receptive vocabulary scores in undergraduate students of French. Hence, if this relationship does indeed exist, it is not well documented and certainly not fully understood. In addition, the relationship between the development of these two aspects of the language system is rarely tracked in learners at different stages of development.

In all the studies mentioned above, the measures of grammatical development used are rather general and indiscriminate, assuming the acquisition of grammar is a unified phenomenon. However, SLA research over the last two decades or so has clearly shown that not all aspects of the morphosyntactic system are as amenable to acquisition as others (see, for example, Hawkins 2001a, b, 2004; Mitchell and Myles, 2004, White, 2003 for overviews). For example, Lardiere (1998) argues that the acquisition of syntax and the acquisition of morphology are separate phenomena.

In this chapter, we therefore investigate the relationship between lexical and grammatical development in second language learners at three distinct stages of development by looking at distinct areas of morphosyntax. The next section outlines briefly the theoretical frameworks we are using.

Theoretical framework

Vocabulary

Productive lexical development has only very recently been the focus of L2 studies (e.g. David, 2008; Graham et al., 2008; Marsden and David, in press; Richards et al., 2008; Tidball and Treffers-Daller, 2007), even though researchers agree that the lexicon is one of the key aspects of language. Recent studies have looked at documenting the number of words learners know at different stages (e.g. Milton, 2006) or what measures to use to quantify the developing lexicon (e.g. Malvern et al., 2004). In this chapter, we use a measure of lexical diversity (index of Guiraud) to provide a profile of the developing productive L2 lexicon.

The index of Guiraud (Guiraud, 1954) is the ratio of types to the square root of tokens (types/$\sqrt{}$tokens). It is one of the alternatives to TTR put forward to minimize the impact of text length. This index has been found to be the most stable amongst various other possible transformations of TTR for language learner data (Van Hout and Vermeer, 1988 cited in Daller, van Hout and Treffers-Daller, 2003). Tidball and Treffers-Daller (2005) found that the Guiraud index was one of the best measures to explain variation in their data (university-level learners of French). This measure was also chosen here because it is widely used in language proficiency measurement and, in preliminary work, it had been shown to be the most reliable in correlations with other aspects of linguistic complexity (David, 2007). The unit of count used here is the lemma as this has been shown to be the most valid unit of counting (Vermeer, 2004). The lemma is defined here in the morphological sense of the word, as the canonical form of a word or lexeme. Lemmas are especially significant in highly inflected languages such as French. For example, if the student produced the utterances in examples 1 and 2, the learner is credited with seven lemma types as proper nouns are excluded and *vais* (present first person singular) and *va* (present third person singular) are the same lemma (*aller*). The possessives *mes* (plural) and *ma* (feminine) also count as a single lemma.

1. Je vais à Paris dans mes vacances.
   I go to Paris in my holidays.
2. Ma sœur va dans Ecosse.
   My sister goes in Scotland.

Morphosyntax

In order to analyse the relationship between lexical and grammatical development, we will use, in the first instance, MLU as a general measure of complexity (which we define here as a multidimensional measure that relates to the 'internal diversity and degree of elaboration of the inter-language

system' – van Daele, Housen and Pierrard, 2007). MLU has been widely used in the L1 literature to explore this relationship in the early stages of development. We will then analyse in more detail this relationship by investigating the development of more specific areas of French morphosyntax which have theoretical significance.

Measuring grammatical or morphosyntactic complexity takes many forms (Thomas, 1994), and much effort has been put into finding the best measure(s). MLU remains one of the most commonly used measures as it is a good index of language development, especially in the early stages of child language acquisition, with relatively reliable results. It is generally assumed that MLU can be considered reliable and valid with L1 data (e.g. Rondal, Ghiotto, Bredart and Bachelet, 1987). However, it is considered by some as unsuitable for advanced adult productions (especially as very short utterances tend to lower the scores, see Dewaele, 2000). The maximum cut-off point for reliability is somewhere between three and four morphemes depending on which study one reads (see Klee and Fitzgerald, 1985 or Scarborough, Rescorla, Tager-Flusberg, Fowler and Sudhalter, 1991 for L1 examples).

Evidence based on a general measure of grammatical development such as MLU remains rather vague, however, and does not allow us to draw conclusions about which aspects of morphosyntactic development lexical development might be related to. According to current linguistic theory, Universal Grammar (Chomsky, 1995) consists of invariant computational devices (a syntactic component, a semantic component and a morphological component) and a universal set of different types of features (semantic, phonological and syntactic). Syntactic features can be divided into interpretable (e.g. the phi-features of nouns: person, number) and uninterpretable (e.g. grammatical gender, tense). In this chapter we investigate an uninterpretable feature, grammatical gender, as well as the way in which learners develop over time the functional projections necessary for the acquisition of syntax and morphology (see below).

Gender

French marks a grammatical distinction between masculine and feminine nouns [+/− masculine]:

3. **Le/un** bois [+masc] vert
   the/a green wood
4. **La/une** forêt [−masc] vert**e**
   the/a green forest

The grammatical feature controls agreement between the article, the noun and any adjectives present (Hawkins, 2001a). The presence or absence of a gender feature is a parameter of variation allowed by Universal Grammar.

The feature is present in French but not in English. It has been documented that child L1 learners acquire the gender feature early in development (Clark, 1985). Karmiloff Smith (1979) found that children as young as three years and two months are consistently able to choose an appropriate determiner with nonsense words, like *bicron* [+masc] or *plichette* [−masc], suggesting that children use phonological cues to determine the gender of the noun. Additionally, child bilingual learners exhibit no problems with gender concord (Granfeldt, 2005). In contrast, the acquisition of grammatical gender appears to be an area of persistent difficulty for English learners of L2 French (Hawkins and Franceschina, 2004).

Syntactic projections (lexical and functional)

In this section, we assume a modulated structure-building approach (Hawkins, 2001a) to L2 language acquisition in which learners originally project only lexical categories which transfer from their L1. Simplifying somewhat, lexical categories are roughly equivalent to 'content' words, such as nouns, verbs or adjectives. Functional categories contain more abstract grammatical features such as inflections, or complementizers). According to the modulated structure-building approach, learners do not originally project functional categories. These categories are necessary for the production of inflections such as tense markings and for various syntactic operations such as the formation of interrogatives or embedded clauses. These develop later in succession as learners build the syntactic tree in a gradual/stepwise fashion. They will first project lexical phrases such as noun phrases (XP where X is a lexical category such as noun or verb) then inflectional phrases (IP) and finally complementizer phrases (CP) as illustrated in Figure 9.1.

In this study, we will examine whether there is a link between vocabulary development and the production of these syntactic categories in our learners (XP, IP and CP). We would expect learners to, first, be able to produce lexical phrases such as noun phrases before being able to produce verb

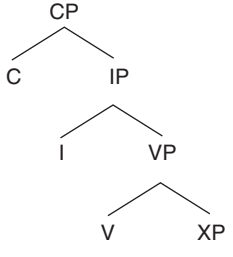


Figure 9.1 Simplified underlying sentence structure

phrases which require projecting a verb and all its arguments (i.e. complements, subject). This stage gives rise to verbless utterances, well documented in the literature (Housen, 2002; Lakshmanan, 1998; Myles, 2005). In a second stage, learners would project VP which would lead to sentences with untensed verbs. In a third stage, learner will be able to project IP, and inflected verbs will appear. Finally, with the projection of CP, the complementizer phrase, embedded sentences will be produced.

In child L1 acquisition of French, morphological development and syntactic development appear to run in parallel (Herschensohn, 2001, p. 275). Initially children use untensed verbs with strong pronouns (e.g. *moi*) or nouns as subjects. Children then go through a stage which has become known as the optional infinitive (OI) or root infinitive (RI) stage. At this time, children will use either tensed verbs with verb raising and subject clitics, or untensed forms without verb raising or subject clitics (as in examples 5 and 6):

5. *Moi, je tousse encore.* (adult-like) Philippe 2;2.2 (Pierce, 1992, p. 96)
   Me, I cough again
6. *Lancer la balle* Philippe 2;1.3 (Pierce, 1992, p. 109)
   to throw the ball
   [someone] throws the ball

Radford (1996) argues that in children the use of verbal morphology is linked to the acquisition of functional features. This means that when a child uses a tensed verb, it will have raised the verb to IP (Pollock, 1989). Rizzi (1994) argues that children do not consistently project the full tree while they are acquiring it (i.e. a child may project just the VP, just the IP or the whole CP during this OI stage). This hypothesis is known as the Truncation Hypothesis. Both these arguments explain why in child L1 we do not see the use of subject clitics with untensed verbs, nor the use of strong pronouns alone with tensed verbs.

In L2 acquisition, it has been argued by some researchers that, like in L1 acquisition, L2 learners do not initially project functional features and that these are acquired gradually (Hawkins, 2001a; Myles, 2005; Vainikka and Young-Scholten, 1996). This view is not uncontroversial (see Lardiere, 1998, 2000; Prévost and White, 2000; Schwartz and Sprouse, 1996 for an alternative analysis which suggests that learners transfer functional categories from their L1 from the outset, but have problems with the realization of surface morphological endings). Some researchers have argued that uninterpretable features (e.g. gender), which are not available in the learner's L1, will not be available via Universal Grammar in the learner's L2 (Hawkins and Chan, 1997).

This chapter aims to answer the following research question by analysing the acquisition of French after one, three and five years of learning French in the classroom: is there a link between the development of the lexicon

and different aspects of the development of morphosyntax in second language learners of French, or do they develop separately because they belong to different types of knowledge and involve different types of learning? Three possible links will be examined:

1. Link between the development of the lexicon and a general measure of morphosyntactic development (MLU);
2. Link between the development of the lexicon and the development of one uninterpretable feature (gender);
3. Link between the development of the lexicon and the development of syntactic structure (modulated structure building: XP, then VP, then IP, then CP).

These relationships will be investigated at three different stages of interlanguage development, as it is possible that such links exist at a particular stage but not others (e.g. some minimum level of lexical development might be a prerequisite for triggering syntactic development, but the link might weaken thereafter).

Method

Learners

This cross-sectional study describes, compares and analyses the linguistic development of 60 instructed learners of French in Years 8 (aged 12/13), 10 (aged 14/15) and 12 (aged 16/17) in the British school system. There are 20 learners in each year group. By the end of Year 8 (their second year of classroom learning) students will have received around 100–120 hours of instruction, about 240 hours by the end of Year 10 and by the end of Year 12, they will have had a maximum of 525 hours.

Oral task

All 60 participants carried out the same oral task which involved a conversation about a set of photos. This task takes the form of a one-to-one semi-structured interview in French between individual learners and a member of the research team. The task is in two parts. In the first part, the learners are shown stimulus photographs representing young people doing various age-appropriate activities (family life, pets, gap year, holidays, etc.). The learner is instructed to find out as much information as they can about the young people shown in the pictures, the location, and events by asking questions. In the second part, the researcher asks the learner a range of questions about topics such as their family life, interests and hobbies or holidays. For this second part, the photos only serve as a starting point to the conversation but the discussion is not solely based on the photos.

Data analysis

All of the oral conversations were transcribed by native or near-native speakers of French using the CHILDES guidelines into CHAT (MacWhinney, 2000). All of the transcripts were then tagged for part-of-speech using the MOR facility in CLAN. All of the data (soundfiles, transcripts and tagged files) are available through the FLLOC (French Learner Language Oral Corpora) website at: http://www.flloc.soton.ac.uk/.

In all our analyses, we have excluded formulaic language which is extremely common in early learners, and can substantially distort the investigation of L2 development (for a discussion of identification criteria used as well as the role played by formulaic sequences in early L2 development, see Myles, 2004; Myles, Hooper and Mitchell, 1998; Myles, Mitchell and Hooper, 1999).

Results

Vocabulary development

The index of Guiraud was calculated for all three year groups, and Table 9.1 presents the means, maximum, minimum and standard deviation. The difference (as measured by the Mann–Whitney test) between the Year 10 and 12 groups is statistically significant ($z = 4.085$, $p = <.001$) but not between Years 8 and 10 ($z = 1.677$, ns), suggesting that students have a significantly more diverse productive lexicon in Year 12 than in Year 10 but that the difference between Years 8 and 10 is minimal. This mirrors Milton's (2006) finding that relatively little progress is made in earlier years of L2 acquisition in receptive vocabulary in UK secondary schools.

The notion of a plateau in linguistic development has been observed by other researchers (e.g. Milton, 2006). Laufer (1989) suggests that a plateau in productive lexical learning can exist because of the nature of memory and the learning process. Milton (2006) proposes that this plateau may be a feature of French learning in general in British schools as, according to him, there are no apparent reasons for it. This plateau is also observed in L1 acquisition in which vocabulary growth is rather slow initially and then takes off very quickly at the onset of the two-word stage.

Table 9.1 Descriptive statistics for the index of Guiraud

Year	*Mean*	*Number of learners*	*Standard deviation*	*Minimum*	*Maximum*	*Median*
Year 8	4.00	20	.62	2.86	5.29	3.86
Year 10	4.33	20	.61	3.17	5.30	4.35
Year 12	5.40	20	.73	4.25	7.30	5.31
Total	4.58	60	.88	2.86	7.30	4.56

Table 9.2 Descriptive statistics for MLU

Year	*Mean*	*Number of learners*	*Standard deviation*	*Minimum*	*Maximum*	*Median*
Year 8	2.71	20	.75	1.61	4.37	2.52
Year 10	4.54	20	1.35	2.80	7.54	4.74
Year 12	5.99	20	1.25	4.05	8.50	5.98
Total	4.41	60	1.76	1.61	8.50	4.30

There is very substantial vocabulary development, however, between Years 10 and 12. The fact that this development is so extensive might be in part due to the fact that these learners have opted to study French in Year 12 as one of three or four specialist subjects and are thus self-selected successful learners. Additionally, they receive more intensive tuition (up to six hours a week).

Relationship between lexical development and MLU

Here, mean length of utterance was calculated using the MLU command of the CLAN program. This command was run on the MOR tier of the tagged files. Running the command on the tagged files provides an MLU score in morphemes rather than words.

Table 9.2 presents the MLU scores for the three groups of learners. There is a statistically significant difference (based on the Mann–Whitney test) between the MLU of Years 8 and 10 ($z = 4.382$, $p < .001$) and between the MLU of Years 10 and 12 ($z = 3.300$, $p = .001$). The three groups of learners significantly increase their mean length of utterance during the period studied. Figure 9.2 represents the relationship between MLU and Guiraud index for all three groups of learners.

There appears to be a link between lexical diversity and MLU: unsurprisingly, the more diverse the learners' vocabulary the longer MLU they appear to have. This is confirmed by correlations. There is a significant positive correlation between MLU and Guiraud when all the learners are grouped together (Pearson correlation, $r = .619$, $N = 60$, $p < 0.01$). However, the correlation is not significant within individual year groups: in Year 8 ($r = .188$, $n = 20$, ns), in Year 10 ($r = .435$, $n = 20$, ns) and in Year 12 ($r = .198$, $n = 20$, ns).

To conclude, it appears that vocabulary development and MLU do correlate, if somewhat tenuously. Let us now turn to the development of specific morphosyntactic properties.

Relationship between lexical development and morphosyntactic measures

As outlined above, in this section, we focus on two different aspects of morphosyntactic development: the development of uninterpretable features

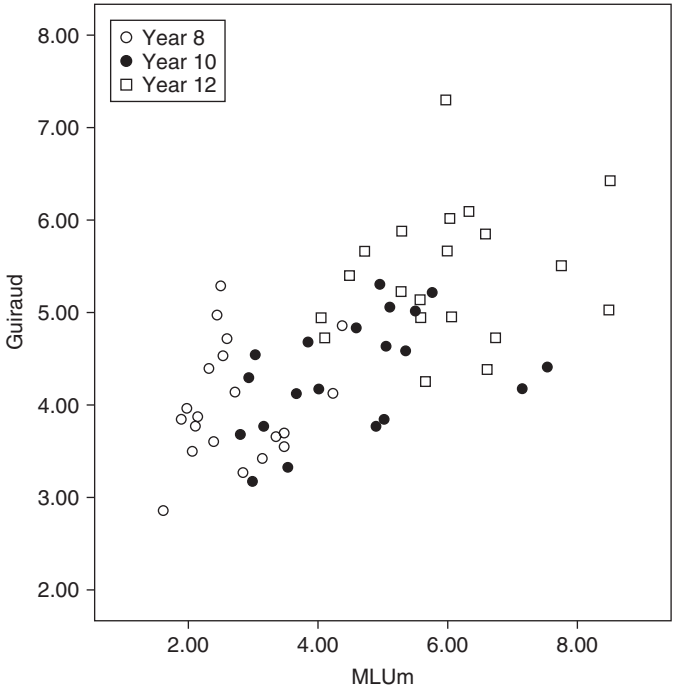


Figure 9.2 Scatterplot of the relationship between MLU (morphemes) and Guiraud

(gender), and modulated structure building as evidenced by the production of lexical phrases initially, then finiteness and embedded clauses.

Gender

The tagged data were searched using the CLAN analysis program COMBO to analyse the learners' development of grammatical gender, e.g. *le père*, *la table*, *une maison*. The program searched for tokens of masculine determiners (both definite and indefinite) followed by masculine and feminine nouns, e.g. *le père*, **le table*, **un photo* and then feminine determiners (both definite and indefinite) followed by feminine and masculine nouns, e.g. *la table*, **la garçon*, **une chien*. The definite determiners included examples of possessives (e.g. *mon*, *ma*) and demonstratives (e.g. *ce*, *cette*) but excluded *quel/quelle* as the contrast between the two forms is only orthographic and is not audible. We counted types rather than tokens: each new example of a noun (either feminine or masculine), and whether the gender concord was correct or incorrect. For example, if a learner produced **la garçon* ten different times, this was counted as one (inaccurate) gender concord. We counted both *le garçon* and *un garçon* as two (accurate) types. If a learner

Table 9.3 Descriptive statistics for correct gender concord (%)

Year	*Mean*	*Number of learners*	*Standard deviation*	*Minimum*	*Maximum*	*Median*
Year 8	71.64	19	18.30	37.50	100	72.22
Year 10	70.32	20	14.96	42.11	100	70.36
Year 12	78.28	20	13.11	50.00	100	80.91
Total	73.44	59	15.69	37.50	100	73.33

produced the same noun with two different gender concords (e.g. **le fille* then *la fille*) these were counted as two types. Cases where learners produced an indeterminate form sounding somewhere in between *le* and *la* or *un* and *une* were counted as incorrect gender (some learners used this avoidance strategy systematically). Uses of *l'* were also excluded.

Table 9.3 shows the results for correct grammatical gender assignment for the three groups of learners. Year 8 results are out of 19 rather than 20 as one learner did not produce any determiners.

Surprisingly, differences between the year groups are rather small. Mann–Whitney tests reveal that there is no significant difference between scores for Years 8 and 10 ($z = .07$, ns), nor between Years 10 and 12 ($z = 1.881$, ns). In addition, all the scores are fairly high (above 70 per cent) when compared to the literature (e.g. Hawkins, 2001b). The relative homogeneity of results could be explained by the very small set of familiar nouns that the Year 8 and possibly Year 10 students produced (e.g. repeatedly using *le garçon*, *la mère*). One could argue that these nouns are learned with their correct definite determiner as frozen phrases, which would explain why the learners in each year group who were 100 per cent accurate were also the learners who did not produce more than two or three different types of determiner phrases. The more diverse a learner's lexicon is the more chances s/he has of producing wrong genders as they are taking more risks.

No developmental pattern is visible for gender and this is confirmed by the lack of significant correlation with lexical diversity (Pearson correlation, $r = .148$, $N = 59$, $p = .262$) for the three year groups. Theoretical explanations for the lack of correlation between these two aspects will be explored in the discussion section.

Projection of XP (where X is a lexical category)

To investigate the relationship between vocabulary development and structure building, we first examined the emergence of lexical categories. We counted all verbless utterances, that is, utterances comprising primarily noun phrases with the occasional prepositional phrase, very common in early learners (e.g. *la mère à la ville:* the mother in the town). These utterances are a good indication that learners are at the lexical stage and

Table 9.4 Descriptive statistics for verbless utterances (%)

Year	*Mean*	*Number of learners*	*Standard deviation*	*Minimum*	*Maximum*	*Median*
Year 8	9.52	20	12.67	0	40	4.06
Year 10	14.72	20	18.20	0	60	7.92
Year 12	.97	20	1.81	0	6	.00
Total	8.40	60	13.86	0	60	2.82

not yet projecting higher-level categories such as IP or CP. Utterances with verbs are more difficult to analyse, as they often contain default forms which are used in both finite and non-finite contexts, making it difficult to know if IP has been projected or not.

Once again, using the COMBO command of CLAN, we searched for all utterances which did not have a verb and counted all utterances lacking a verb where one was clearly necessary (e.g. **elle quel âge?:* she what age?). Proportions were then computed of verbless utterances out of the total number of utterances produced by the learners. The total number of utterances excludes single word utterances including replies to questions (e.g. *Où habites-tu? Londres:* where do you live? London).

Table 9.4 illustrates the results (in percentages) of the verbless utterances produced by learners. Mann–Whitney test results show that the difference between Years 8 and 10 is not significant ($z = 1.107$, ns). However, the difference in percentages between Years 10 and 12 is significant ($z = 3.858$, $p < .001$). These findings suggest that our learners are well on the way out of the early stage when they project lexical phrases and lack functional projections, as verbless utterances have almost completely disappeared by Year 12.

A Pearson correlation between the index of Guiraud and the percentage of verbless utterances is significant at 0.05 level ($r = -.258$, $N = 60$, $p = 0.46$). This is illustrated by Figure 9.3. It indicates that those learners who produce more VPs are also those who have the more diverse vocabulary. Although there is a large amount of individual variation in how proficient learners are within each year group (with some learners in Year 12 still performing at Year 8 level and vice versa), learners seem to be consistent in their proficiency across the various areas of development we investigate. In other words, learners who are poor at vocabulary also tend to be poor at syntactic projections and learners who have a large vocabulary also tend to do well on syntactic projections.

Projection of IP

In order to determine if learners project IP, we counted the number of finite verbs with subject clitics. The reason for excluding finite verbs with lexical subjects is that, as mentioned in the previous section, many learners use

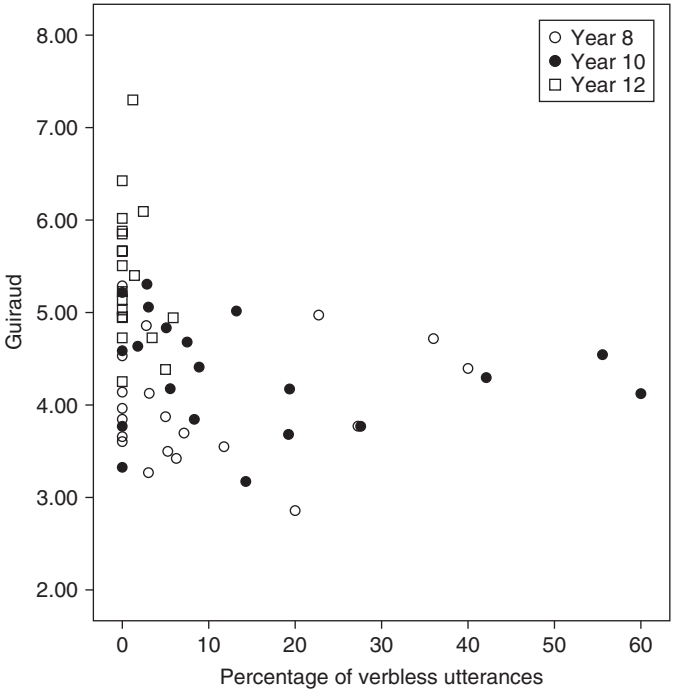


Figure 9.3 Scatterplot of the relationship between verbless utterances and Guiraud

default verb forms and it can be difficult to determine whether they are really finite or not. Additionally a lexical subject does not a priori involve IP. For example, in a learner production such as *ma famille aller en vacances* (my family go-INF on holidays), consisting of an NP followed by a VP, there is no evidence of an IP projection, as tense is not marked and the subject does not have to raise to IP. When learners produce subject clitics, they have to raise to the IP position in French, because of a strong uninterpretable feature in IP (the verb-raising parameter) which forces movement to this position. This feature is responsible for a number of syntactic differences between French and English such as adverb placement, subject–verb inversion in interrogatives, and post-verbal negation. Clitics are therefore good evidence that IP is being projected and that learners have moved beyond the lexical projection stage.

We calculated the proportion of clitics with finite verbs out of the total number of all clitics used (finite and non-finite contexts), excluding chunks such as *j'aime* (I love) as in *j'aime le football* (I love football). Table 9.5 presents the percentage of finite verbs produced with clitic subjects out of all clitics used (finite + non-finite contexts).

Table 9.5 Descriptive statistics for clitics produced with finite verbs (%)

Year	*Mean*	*Number of learners*	*Standard deviation*	*Minimum*	*Maximum*	*Median*
Year 8	86.62	20	16.54	33.33	100	90.91
Year 10	83.73	20	18.14	33.33	100	87.92
Year 12	94.76	20	3.73	85.71	100	95.20
Total	88.37	60	14.86	33.33	100	91.99

Table 9.6 Descriptive statistics for the proportion of embedded sentences

Year	*Mean*	*Number of learners*	*Standard deviation*	*Minimum*	*Maximum*	*Median*
Year 8	.07	20	.31	.00	1.39	.00
Year 10	.35	20	.86	.00	2.86	.00
Year 12	2.08	20	1.65	.00	5.06	2.44
Total	.83	60	1.40	.00	5.06	.00

The difference between Years 8 and 10 is not statistically significant ($z = .978$, $p = .328$) whereas the difference between Years 10 and 12 is significant ($z = 3.317$, $p = .001$). The marginally higher percentage for Year 8 (when compared to Year 10) is due to two Year 10 learners who produce a very large proportion (around 66 per cent) of non-finite verbs with clitics (which show that it is somewhat dangerous to overgeneralize on the basis of one or two individuals, as has sometimes been done elsewhere in the literature).

The correlation between lexical diversity and clitics is not statistically significant (Pearson correlation, $r = .111$, $N = 60$, ns). We will suggest why this might be the case in the discussion section below.

Projection of CP

Finally, embedded clauses were counted to provide evidence for the CP stage. Here, embedded clauses are all dependent clauses (including subordinate clauses), for example, *je voudrais un chien parce que j'aime le chien* (I would like a dog because I like the dog). The analysis of embedded clauses is based on the number of different types of dependent clauses, that is to say, if a student produces ten utterances all starting with *je pense que …* (I think that …), then this is only counted once. Proportions are calculated out of the total number of utterances (once again excluding single word utterances).

Table 9.6 gives the proportion of embedded sentences out of the total number of utterances, for each year group. Our data reveal that learners produce significantly more embedded clauses in Year 12 than 10 (Mann–Whitney test: $z = 3.547$, $p < .001$) but that the difference is not statistically significant between Years 8 and 10 ($z = 1.117$, ns). This is not surprising as

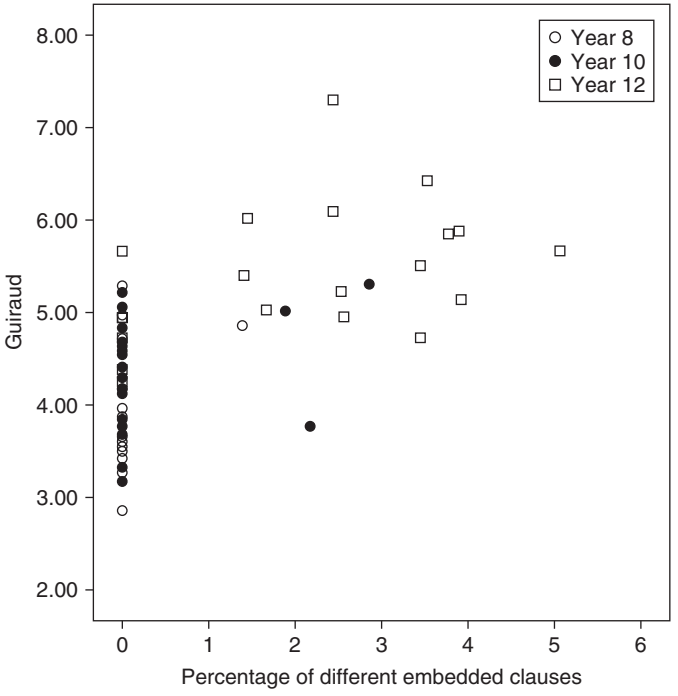


Figure 9.4 Scatterplot of embedded clauses and Guiraud index

the median for both groups is zero. The overall figures remain very small as in Year 12, only about 2 per cent of learners' productions contain embedded clauses of different types. Despite those small numbers, the correlation between the index of Guiraud and the percentage of different embedded clauses is significant at 0.01 level (Pearson correlation, $r = .633$, $N = 60$, $p < .001$), with learners with the most diverse lexicon producing the largest number of embedded sentences, as illustrated by Figure 9.4.

Discussion and conclusion

The picture that emerges from the results of this study can be summarized as follows. Firstly, development in our learners is, generally speaking, rather minimal between Years 8 and 10 (it only reaches significance in terms of MLU), but becomes statistically significant for all measures except gender (i.e. for lexicon, MLU, projection of XP, IP and CP) between Years 10 and 12. Secondly, significant correlations have been found between lexical development and MLU, projection of XP and projection of CP. No correlations were found between lexical development and gender and projection of IP.

We would like to suggest, somewhat tentatively at this stage, that the development of the L2 lexicon is related to the development of syntactic complexity in general terms, but not to the acquisition of uninterpretable features. Two different tasks seem to be confronting learners in their acquisition of morphosyntax: on the one hand, they have to develop the ability to produce functional projections increasingly higher up the syntactic tree, starting with lexical projections and graduating to functional projections such as CP over time. On the other hand, they have to acquire the uninterpretable features which are part of functional categories and differ from one language to another. In the case of French, grammatical gender is one such feature, and its acquisition clearly seems to be unrelated to the development of the lexicon, with no evidence of learners with a more diverse vocabulary having mastered gender any better than learners with a smaller vocabulary. Another uninterpretable feature our learners have to acquire is the verb-raising parameter attached to IP. Again here, the development of this property seems unrelated to lexical diversity. This is not to say that learners have not acquired those properties; our results clearly indicate that those learners who produce clitics are projecting IP and the associated verb-raising parameter. Similarly, some learners are fairly accurate in assigning gender. But what our results show is that the acquisition of uninterpretable features does not seem to be linked with more general measures of development such as lexical diversity or more general syntactic measures. More general measures, however, such as MLU or the ability to project functional categories correlate with vocabulary development, suggesting that those two go hand in hand.

These results are tentative at present, and we would need to test our hypotheses on larger numbers of learners, or better still, in a longitudinal design. But what we think they show conclusively is that we have to be wary of assuming that the acquisition of morphosyntax is a unified phenomenon; different properties clearly seem to develop in different ways. What does this all mean in terms of theoretical implications? It would seem to support a modular view of learning, with 'pure' syntax (as realized in uninterpretable features) developing separately from the lexis, or from the computational mechanisms which enable the construction of increasingly complex syntactic structures.

Note

1. The research reported here is based on data collected during the FLLOC projects (directed by Florence Myles and Ros Mitchell) funded by the UK Economic and Social Research Council (ESRC) award numbers R000223421, RES000220070, the Arts and Humanities Research Council (AHRC) RE-AN9057/APN-15456, AR112118 and the British Academy SG 41141 since 2001, at the University of Southampton and Newcastle University.

10

A New Method of Measuring Rare Word Diversity: the Example of L2 Learners of French

David Malvern and Brian Richards
The University of Reading, UK

Introduction

In his summary of measures of effective vocabulary use, Read (2000) includes among his four dimensions of lexical richness 'A selection of low frequency words … rather than just general, everyday vocabulary [including] technical terms and jargon as well as the kind of uncommon words that allow writers to express their meanings in a precise and sophisticated manner' (p. 200). He refers to this characteristic as lexical sophistication.

The importance of this concept can be gleaned from a perusal of the chapters in this volume. For first language Henrichs and Schoonen (Ch. 1) consider the relationship between parents' use of an academic register and their children's language over time. They note the importance of features of academic language for school success. Dickinson, Flushman and Freiberg (Ch. 2) note how children acquire less common and more complex vocabulary once they have moved beyond an initial core lexicon and identify exposure to sophisticated vocabulary as an important factor in children's language and literacy development (see also chapters in Dickinson and Tabors, 2001). Sealey (Ch. 3) draws attention to how values about the sophistication of vocabulary are enshrined in the English National Literacy Strategy and passed on to pupils in school. Alternative words are encouraged that are 'interesting' and 'accurate' rather than 'common choices'. In relation to second language learners' performance on different tasks, Skehan (Ch. 7) measures lexical sophistication as one of two dimensions of vocabulary richness and discusses the factors that influence it. Finally, Milton and Alexiou (Ch. 12) note the strong relationship between the frequency of words and the likelihood of their being learned, a phenomenon that systematically underpins the design of recent language tests (see also Milton, 2007b).

Second language researchers concerned with the measurement of lexical richness have argued strongly that indices that are entirely quantitative, and treat all vocabulary items as of equal value (e.g. measures of overall

lexical diversity), give an incomplete picture and need to be supplemented by a qualitative dimension (Daller, Van Hout and Treffers-Daller, 2003; Meara and Bell, 2001; Richards and Malvern, 2007). Thus, measures of lexical sophistication are now frequently incorporated into studies of language development across different languages (e.g. Daller et al., 2003; Daller and Xue, 2007; Tidball and Treffers-Daller, 2007, 2008; Richards, Malvern and Graham, 2008).

As Richards et al. (2008) have pointed out, lower-frequency vocabulary, especially in English, is often associated with technical terminology, greater levels of precision, abstraction and semantic and morphological complexity. Words are longer, more difficult to spell, later acquired, perceived as being more literary and are more difficult for learners. For English and some other European languages many such words have their origin in Greek and Latin and are central to educational achievement. It is not only in first and second language and bilingualism, however, that lexical sophistication has been an issue. In fact, there is a long tradition of measuring the deployment of rare words, and rare word diversity in a number of language disciplines and linguistic fields. These include genre studies and stylistics such as Ménard's (1983) seminal work *Mesure de la richesse lexicale*, and Biber's (1988) research on lexical specificity and variation across speech and writing. More recent research includes technical English compared with other genres (Milton and Hales, 1997), and attempts to date literary works from measures of lexical richness (Smith and Kelly, 2002). Clinically oriented studies, such as the conversational competence of aphasic patients (Holmes and Singh, 1996) and of those with Alzheimer-type dementia (Bucks, Singh, Cuerden and Wilcock, 2000) also measure lexical sophistication (see Malvern, Richards, Chipere and Durán, 2004, for further examples across the fields of language and applied linguistics).

Measuring rare word diversity

What nearly all the above studies have in common is that they set out to measure not the *frequency* of use of less common vocabulary, but its *diversity*. The former is relatively unproblematic from a measurement perspective – the proportion of tokens in a text that are rare can be calculated irrespective of the length of the text. The latter, on the other hand, is more useful because it reflects the range of sophisticated vocabulary that the speaker or writer brings to the task, but brings with it a serious problem of measurement. Most of the methods used to measure rare word diversity have been either simple counts of the number of rare words in a sample, or extensions of, and variations on, traditional lexical diversity measures, for instance the type token ratio (TTR) or its derivatives such as Guiraud's index (Guiraud, 1960), applied to the rare, or 'advanced' words in the sample (see Malvern et al., 2004, for an overview). Unfortunately, these traditional methods

suffer from their being dependent on the size of the language sample used to make the measurement. TTR is well known to fall with increasing text length. 'Advanced' TTR has the same flaw and it too will inevitably fall with increasing sample size. The 'Advanced' Guiraud (advanced types/$\sqrt{}$tokens) (Daller et al., 2003) is also a function of sample size and can be shown mathematically to rise to a maximum value as the size of the language sample increases before falling, slowly at first, but eventually more rapidly, as the token count becomes very large, in the same way as Guiraud's index itself does. Although both Ménard's (1983) original rarity measure (the proportion of low-frequency word types) and the similarly defined 'rare word density' of Snow, Tabors and Dickinson (2001) used to assess the linguistic environment of the home, are type/type (rather than type/token) measures, nonetheless, as is clear from Ménard's work, both also depend on the size of the language sample from which the measure is derived.

In order to overcome this difficulty, we have developed a measurement for lexical diversity which is not a function of sample size in the way these traditional measures are. This measure, *D*, is based on mathematically modelling how the TTR of any given language sample falls with increasing tokens. If a graph is drawn of the TTR against token count (*N*), the value of the TTR will fall in a curve which starts off dropping steeply, then falls more and more gently while continuing to tend towards zero for very large *N*. All language samples do this, but the more diverse the deployment of vocabulary in a sample the less steep will be the fall. A set of texts of differing lexical diversity will be represented by a set of similarly shaped curves, with the graphs for more lexically diverse ones falling less steeply and lying above those of the less diverse texts which will fall faster. Our mathematical model for all these curves is an equation containing a parameter, *D*, which relates to how the graph falls and can, therefore, be used as an indication of the lexical diversity of a language sample – the bigger *D* the more diverse the text (see Figure 10.1). A full description of *D* and the program to make the measurement (VOCD) can be found in Malvern et al. (2004). For now it is sufficient to note that it provides a valid and robust measure of lexical diversity or how well vocabulary is deployed.

At first this seems to overcome the problem – we could simply calculate a *D* for rare words only, by running VOCD on a transcript filtered so that only the rare words are entered into the calculations. Unfortunately, to discern how words are deployed requires that there are a sufficient number of them to investigate, and the VOCD program does not compute a value of *D* on samples with fewer than 50 tokens. This would mean that only samples with 50 or more rare word tokens in them could be used to calculate D_{rare} directly, and such samples are few and far between. Put simply, the real problem with measuring the diversity of rare word usage is that rare words are rare, and a way has to be found round the problem that language

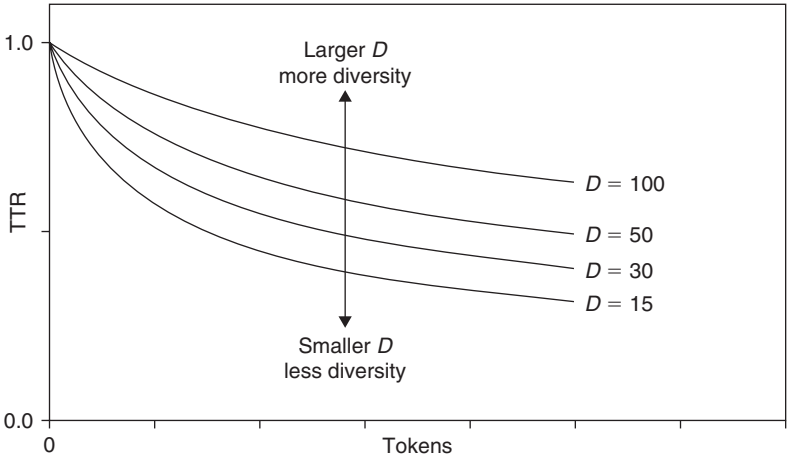


Figure 10.1 TTR versus token size (*N*) for language samples of different diversity

samples of typical sizes found in educational or similar research simply do not contain enough rare words for direct calculation of D_{rare}.

A way round this is to recognize that a particular language sample usually consists of a lot of words from the basic (non-rare) vocabulary and a few words which are rare, and to use VOCD to calculate both the diversity for the complete sample (D_{all}) and the diversity for just the subset of the whole sample consisting of words drawn from the basic vocabulary (D_{basic}), provided that there are at least 50 tokens from the basic vocabulary. The difference between these two, $D_{all} - D_{basic}$, represents the additional diversity above basic vocabulary use contributed by the rare words in the sample. We call this difference the rare word diversity (RWD) of the sample (i.e. RWD = $D_{all} - D_{basic}$).

If in the sample being tested rare words are deployed with more variation than is found within the basic vocabulary, then the overall diversity would be higher when the rare words are included (i.e. $D_{all} > D_{basic}$) and RWD would be large and positive. If no rare words were used at all, then $D_{all} = D_{basic}$ and obviously RWD = 0, whereas if there are only a few rare words or the ones that are there are used in much the same way as the basic vocabulary, then $D_{all} \approx D_{basic}$ and RWD would be small. More interestingly, perhaps, is that if RWD is large and negative, it shows that there may well be a noticeable number of rare word tokens in the text, but there is little variability within them and these tokens consist largely of repetition of a few rare word types. Consequently the overall diversity is driven down by the inclusion of these rare tokens and is lower than the diversity

found in deploying just the basic vocabulary ($D_{all} < D_{basic}$). RWD, then, can be used as an indicator of how well the speaker or writer is deploying rare words. It should be noted that RWD is not the same as D_{rare}, but the two are clearly mathematically related and should correlate perfectly if D_{rare} could be calculated directly.

Selecting the unit of analysis

In order to calculate an RWD, the words in the sample need to be partitioned between rare and basic vocabulary, which requires a definition of either rare words or of basic vocabulary – either will do, as the other category becomes all other words not included in the one defined category. Basically there are three ways of achieving this. The first makes use of data *extrinsic* to the data set under study, by appealing to information about word usage in the language in general. This is usually in the form of frequency information from large-scale data sets such as national language corpora. For example in a study of L2 Swedish learners, Hyltenstam (1988) used a million word corpus from Swedish newspapers to define (very) rare words as those not in the most frequent 7000 words. Similarly, in a study of EFL learners, Afitskaya (2002) used the British National Corpus 1000 most frequent words to mark the cut-off point between basic and advanced vocabulary.

The second method uses official curricula word lists, for instance Arnaud's (1984) identification of basic vocabulary as the French Ministry of Education official list of 1522 words that pupils were expected to master by the time they enter the *lycée*, or Linnarud's (1983) use of school L2 vocabulary lists. The third method is to use researched expert judgements, for instance the 3000 words judged by teachers to be known to fourth graders (Chall and Dale, 1995), used as a definition of basic vocabulary by Dickinson and Tabors (2001) in the Home–school study of language and literacy development; and the judgements of seven teachers of Turkish as a foreign language used by Daller et al. (2003) in a study of language dominance in bilinguals. For a systematic exploration of the effect on results of different operationalizations of such extrinsic criteria, see Tidball and Treffers-Daller (2008).

When the criterion for rareness is external to the subjects under study and derived by appeal to such wider usage of the language, RWD is consistent with Read's (2000) definition of lexical sophistication (see above) and RWD is essentially measuring the diversity in the deployment of words used relatively rarely in the particular language in general. A second form of rare word usage can be derived from information *intrinsic* to the data set under study. For instance, all the types used by the subjects under investigation could be listed along with their frequency of occurrence within the pooled data, and then a boundary could be drawn between basic and rare

vocabulary for this group of subjects by imposing a condition appropriate to the study. Examples could include:

- A word will be considered rare if its average frequency of use per subject falls below a half (say). This would be suggesting that if the internal data for the group indicate that it is likely that on average there are more people who do not use the word than there are who do, we can consider the word to be rare; otherwise it would be basic vocabulary.
- Or deciding that the basic vocabulary consists of those types which everybody uses at least once (and all other types are considered rare).

Of course, another specific condition could be chosen; the important point is that this way of looking at things uses the intrinsic information in the data set to determine what is a rare and what is a basic type by stipulating a definition of one or the other suited to the problem in hand. When this kind of procedure forms the basis for determining which words are rare and which basic, RWD would be measuring what Read (2000) calls lexical individuality and Linnarud (1983) refers to as lexical originality, that is diversity in the deployment of words used only rarely among one's peers in a particular context. It goes without saying that the list of rare words derived from this approach is ungeneralizable, and would not normally be applied to any subject outside the sample from which the list was obtained.

In order to use RWD as a measure of either lexical sophistication or lexical originality/individuality, D_{all} can be calculated straightforwardly from running VOCD on the full text, and D_{basic} from applying VOCD to the text with the rare words removed. To do the latter, first the chosen criteria for determining rare and basic words are applied to the pooled list of types found in the language samples under study to arrive at two sub-files – that of basic types and that of rare types. In practice, of course, only one subfile is needed, as D_{basic} can be calculated by running VOCD either with an include file of basic types or an exclude file of rare types. For each individual transcript, then, subtracting D_{basic} from D_{all} yields its RWD. In order to test these procedures and explore the validity of RWD, we applied its measurement to a set of language samples as follows in the next sections.

Validation trial of RWD

The data set consisted of written texts produced by 32 seventeen-year-old school-based learners of French as a foreign language. The students attended four 11–18 comprehensive schools in the Reading and Oxford areas and were in Year 12 – the first year of post-compulsory education in England. They had studied French for approximately 2½ hours per week over the previous five years (Years 7–11), and had shown sufficient achievement at the national GCSE 16+ examination and aptitude for the language to have

opted for more advanced study, taking an Advanced Subsidiary (AS) course in the year when the data were collected. They could best be described as lower intermediate learners of French. The data analysed here were collected between 7½ and 10 months into their AS course.

The texts were produced as handwritten narratives in response to a sequence of six pictures about a family visit to a stately home and were completed in 30 minutes. The handwritten narratives were typed in Word and converted to the CHAT format of the CHILDES system (MacWhinney, 2000) for analysis by the CLAN programs which include VOCD. During the transcribing, spelling was corrected and tags inserted to permit the exclusion of the following: illegible or otherwise unrecognizable words, non-French words, including proper names that are the same in English, and numbers written as digits. Each file was then edited by hand so that all words would be processed as their root (uninflected) forms and compound words or expressions and reflexive verbs analysed as single lexical items. Finally complete word lists pooled from all students were obtained from CLAN and checked by two modern foreign language experts – Brian Richards and a French language teacher with extensive experience of the relevant age group.

Two extrinsic sources were used to define rare words. First the word list *Le français fondamental 1er degré* (FF1) (Gougenheim, Rivenc, Michéa and

Table 10.1 Descriptive statistics for RWD for 32 students

N	*Min*	*Max*	*Range*	*Median*	*Mean*	*SD*
32	2.12	16.14	14.02	5.4	6.23	3.22

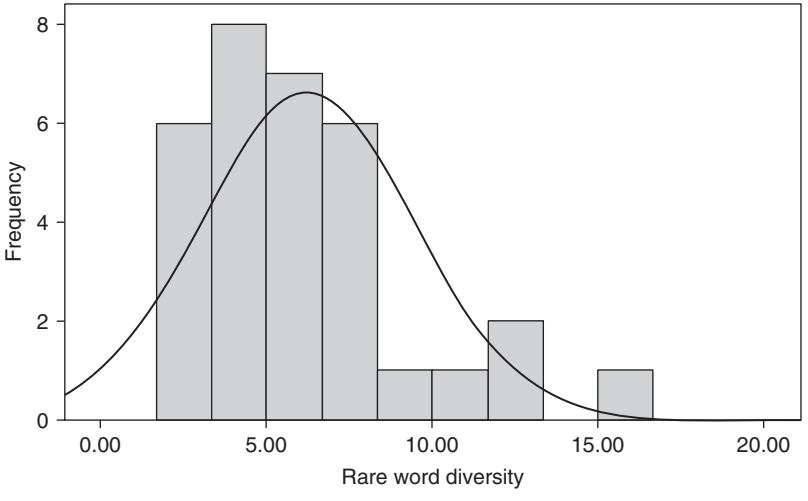


Figure 10.2 Distribution of RWD in trial texts for 32 students

Sauvageot, 1964) was used as an approximation to basic vocabulary. All files were run through CLAN using FF1 as an exclude file,[1] and the words remaining were then scrutinized as candidates for basic words by the two experts. Some words not in FF1 but in common use among the students (e.g. autobus, super) were added to the FF1 file. The final enhanced FF1 file consisted of 1370 types, which defined basic vocabulary for the calculations of RWD for each text.

All texts were run through VOCD in their complete state to obtain a set of D_{all} for each, and then again but using the FF1 file as an include only file to filter out rare words and obtain a D_{basic} for each student. Subtraction of D_{basic} from D_{all} for each text individually provided its RWD. The descriptive statistics of the RWDs obtained for the 32 texts are given in Table 10.1 and the distribution shown in Figure 10.2.

In addition, two students' texts (henceforth Text 1 and Text 2) were selected as the bases to explore the properties of RWD artificially by systematically manipulating the number of rare types and rare tokens in each (henceforth simulation data).

Construct validity using simulation data

As a measure, then, RWD meets the basic requirement of producing a reasonable distribution which seems capable of differentiating among the texts being measured. One noticeable outcome, however, is that while in principle at least negative values are possible, all the values of RWD are positive for this group. In order to check that the procedure for calculating RWD using VOCD in this way *would* produce a negative result if there were high enough repetition and low variation in the rare words used in a text, two transcripts were chosen for exploration – one with low overall diversity and one with high. The first contained 7 rare word tokens all different from each other, while the second contained 28 rare word tokens with altogether 26 rare word types. In both texts the different rare words were replaced by a single, substitute 'rare' word type, thereby removing all variation in the deployment of rare words. The effect of this is shown in Table 10.2 which presents the data for the two texts before and after removal of the variation in rare word usage.

As can be seen in each case, obviously the overall number of tokens is unaltered and the number of basic types, basic tokens and hence *D* for the basic vocabulary all remain the same. While the number of rare word tokens also stays the same, the number of types drops to one which affects first D_{all} and hence RWD.

Text 1, which is modest in length and in overall diversity to start with, has only seven rare tokens but at first they are all different types. Reducing these to one type repeated seven times produces a negative RWD as expected. Text 2 is much longer and more diverse overall and in both the basic and rare

Table 10.2 Two texts, before (original, actual data) and after (simulated data) reducing the number of rare word types to one

	Text 1		*Text 2*	
	Before	*After*	*Before*	*After*
Tokens	72	72	247	247
Types	37	31	110	85
Rare tokens	7	7	28	28
Rare types	7	1	26	1
D_{all}	19.32	12.56	42.14	24.44
D_{basic}	13.54	13.54	30.01	30.01
RWD	**5.78**	**−0.98**	**12.13**	**−5.57**

vocabulary used. It has 28 rare tokens consisting of 26 rare word types to begin with, so replacing these with a single type produces a greater reduction in overall diversity, D_{all}, and hence also results in a negative value for RWD. Further, Text 2 after the change has noticeably more repetition (1 type repeated 28 times) than Text 1 (1 type repeated 7 times), hence the RWD for Text 2 afterwards is a larger negative value (−5.57) than that for Text 1 (−0.98).

We can further demonstrate that RWD responds in the appropriate way when the number and distribution of rare words in a text are changed by concentrating on Text 1 and altering the number of rare word types in smaller steps. Text 1, which in its unadulterated form had seven rare word tokens made up from a total of seven rare word types, was altered systematically to produce seven versions, each containing the seven rare word tokens in precisely the same locations as the original seven, but made up of progressively fewer types with one repeated more often (see Table 10.3).

These different versions of Text 1, then, represent progressively less variation in the rare words used (fewer rare word types) coupled with an increase in the amount of repetition of rare words in the text (one rare word type occurs with increasing frequency) and, if RWD is measuring rare word diversity, its value would be expected to show a continual fall over the seven versions. Table 10.3 shows that this prediction is borne out. Moreover the fall is regular along a straight line, forming an interval scale for changing diversity within this text. All texts do this, and Figure 10.3 gives the values for RWD obtained from applying the same procedure to Text 2, plotted from the lowest diversity (just one type repeated 28 times) to the highest possible (28 different types) to illustrate its scale in ascending order of types and diversity.

This simulation demonstrates that RWD behaves as expected with an engineered decrease in diversity, and does indeed go negative when the deployment of rare words is less diverse than that found in the basic vocabulary of a language sample. The reductions in diversity shown in Table 10.3 and Figure 10.3 were achieved by systematically reducing the

Table 10.3 Original and simulation data: RWD for seven versions of Text 1

	Version						
	Original 7 rare types	*6 rare types*	*5 rare types*	*4 rare types*	*3 rare types*	*2 rare types*	*1 rare type*
Rare tokens	7	7	7	7	7	7	7
Rare types	7	6	5	4	3	2	1
(number × frequency)	(7×1)	(1×2 + 5×1)	(1×3 + 4×1)	(1×4 + 3×1)	(1×5 + 2×1)	(1×6 + 1×1)	(1×7)
D_{all}	19.32	18.43	17.15	15.91	14.75	13.62	12.56
D_{basic}	13.54	13.54	13.54	13.54	13.54	13.54	13.54
RWD	**5.78**	**4.89**	**3.61**	**2.37**	**1.21**	**0.08**	**−0.98**

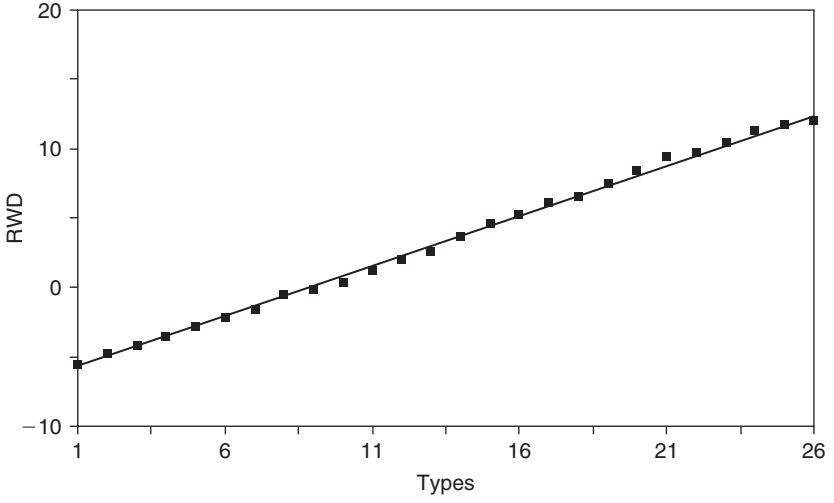


Figure 10.3 Simulation data: RWD for different versions of Text 2 with increasing number of rare word types

number of rare word types so that in each version one type was repeated more and more frequently. There are, of course, other patterns of repetition in which the reduction in the number of types could be achieved. For example, in the version of Text 1 with four types, the seven tokens were made up of one type repeated four times and three other types appearing only once each. An alternative version with seven tokens and four types would be three types each repeated twice and one type appearing once. We would expect the RWD for this alternative four-type version to be near but not the same as the first four-type version, and to show a slightly higher

Table 10.4 Simulation data: RWD for increasing numbers, and different combinations in repetition, of types

Types	*Tokens*	*Number of combinations*	*Combinations (Number of types × frequency)*	*RWD*
1	7	1	1×7	−0.98
2	7	1	$1 \times 6 + 1 \times 1$	0.08
3	7	3	$1 \times 5 + 2 \times 1$	1.21
			$1 \times 4 + 1 \times 2 + 1 \times 1$	**1.61**
			$2 \times 3 + 1 \times 1$	**1.78**
4	7	3	$1 \times 4 + 3 \times 1$	2.37
			$1 \times 3 + 1 \times 2 + 2 \times 1$	**2.84**
			$3 \times 2 + 1 \times 1$	**3.08**
5	7	2	$1 \times 3 + 4 \times 1$	3.61
			$2 \times 2 + 3 \times 1$	**3.99**
6	7	1	$1 \times 2 + 5 \times 1$	4.89
7	7	1	7×1	5.78

diversity than the first which has one type with considerably more repetition than any in the new version. This is precisely what happens, and RWD for the second four-type version is 3.08 – higher than the first four-type version (2.37).

Diversity, then, can increase in more than one way. Along the basic straight line scale formed by the points where, as the number of rare word types is increased, all the repetition continues to be loaded onto one particular type (Table 10.3/Figure 10.3), or by keeping the number of types the same but altering the pattern of repetition in other ways. Table 10.4 shows all the possible ways of doing this for Text 1 and Figure 10.4 illustrates how when both kinds of pattern are included for this text there are intermediate values of RWD (NB Figure 10.4 is plotted to show increasing RWD with increasing diversity). It should be stressed that for points which have different combinations in the way types are repeated, the TTR for each combination is the same but the diversity is clearly different, which illustrates another flaw in TTR – it is insensitive to how types are repeated as opposed to simply how many there are for a given number of tokens. RWD, then, produces the sort of values, range and distribution we would expect of a measure of rare word diversity and we can now proceed to explore its validity further through investigating appropriate correlations.

Face validity and intrinsic comparisons across all 32 students using actual data

The first set of variables which are of interest are intrinsic to the texts under study and a description of the desirable correlations a valid measure of rare

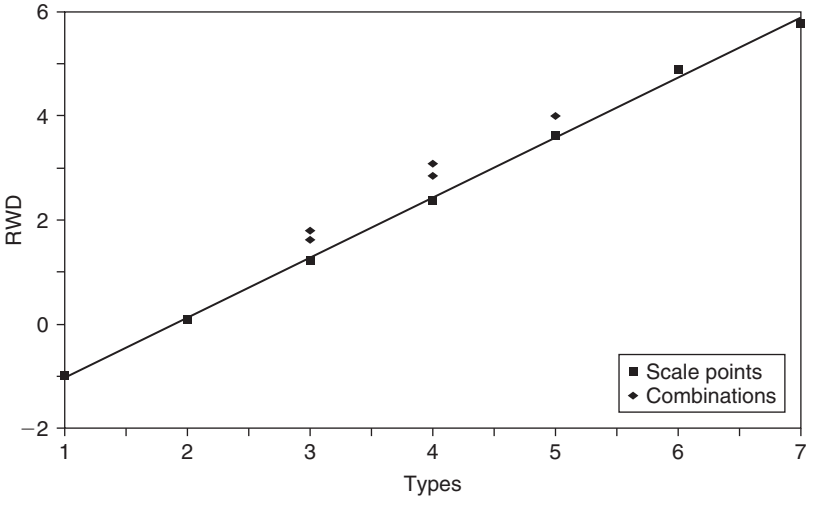


Figure 10.4 Simulation data: RWD for different versions of Text 1 with increasing rare word types, showing scale points and intermediate combinations

word usage ought to satisfy can be made in the form of the following predictions. First, we would expect there to be a relationship between RWD and the number of rare types and rare tokens in a text. As we have seen, simply the numbers of rare types or tokens are flawed measures in themselves as they will depend on the size of the language samples tested. We would expect, then, that the numbers of rare word types and tokens would be very strongly correlated with each other, and significantly correlated with overall text length. On the other hand, while RWD might well be correlated with rare tokens, it should be more highly correlated with rare types but it should not be significantly correlated with overall length of text. These predictions are borne out by the analysis as Table 10.5 shows.

The very form of the expression for RWD predicts that there will be a relationship between it and both D_{all} and D_{basic}. Both mathematically (RWD is proportional to $+D_{\text{all}}$) and in terms of face validity we would expect RWD to be significantly positively correlated with D_{all} – more rare word variation should emerge from more varied texts overall. More significant, however, is that on the one hand in the mathematical expression for RWD, we have RWD proportional to $-D_{\text{basic}}$, which suggests a negative correlation between RWD and D_{basic}, but on the other hand face validity would indicate that the greater one's capacity to vary the basic vocabulary the more one should be able to vary the use of rare words too. Obtaining a significant positive correlation for RWD with D_{basic}, then, would be strong evidence for its validity. The results are given in Table 10.6, and, with both correlations highly significant, provide strong indication of the validity of RWD.

Table 10.5 Spearman intercorrelations between RWD, text length, and the numbers of rare word tokens and types ($N = 32$)

	Text length in tokens	*Number of rare word tokens*	*Number of rare word types*	*RWD*
Text length in tokens	–	0.64 ($p < .001$)	0.649 ($p < .001$)	0.191 (ns)
Number of rare word tokens		–	0.933 ($p < .001$)	0.622 ($p < .001$)
Number of rare word types			–	0.752 ($p < .001$)
RWD				–

Table 10.6 Spearman correlations between RWD and D_{all} and D_{basic} ($N = 32$)

	D_{all}	D_{basic}
RWD	.677 ($p < .001$)	.502 ($p = .003$)

Convergent validity

For students in this study, we have three measures of their general language competence. First there is the actual grade they achieved in the national 16+ examination taken at the end of their previous year of study. Second and third, both their teachers and the students themselves were asked to provide a predicted grade for their current course of study – Advanced Subsidiary level French. For good evidence of validity a positive significant correlation with all three is to be expected, but as the achieved and predicted grades also depend on other language abilities the actual value need not be particularly high. Table 10.7 presents the results, which once again are entirely as predicted for RWD to be valid.

Conclusions

When applied to rare words, traditional measures of diversity, for example the number of word tokens or types or simple proportions like the TTR and derivatives such as Guiraud's index, have the same flaws they have normally and in particular are functions of the length of the language sample analysed. In this chapter we have proposed a new measure, rare word diversity (RWD), which exhibits the appropriate properties for construct validity. Through simulated versions of two real texts, it can be seen how RWD can vary from large and positive to large and negative and in doing so have

Table 10.7 Spearman correlations between RWD and three measures of language ability

	GCSE French grade	*AS teacher predicted grade*	*AS student predicted grade*
RWD	.411 (p = .022; N = 31)	.443 (p = .016; N = 29)	.425 (p = .015; N = 32)

Table 10.8 Interpretation of positive, negative and zero values of RWD

$D_{all} > D_{basic}$ *RWD positive*	$D_{all} \approx D_{basic}$ *RWD ≈ 0*	$D_{all} < D_{basic}$ *RWD negative*
The diversity of rare word deployment is higher than that for basic words and the use of rare words adds to the overall diversity	Either there are no rare words deployed, or the diversity of the rare words that are used is the same as that for the basic vocabulary and their use does not add to or subtract from the overall diversity	There is less diversity in the rare words deployed than in the basic vocabulary, i.e. the speaker/writer is using a few rare word types with high repetition
High rare word diversity	Low rare word diversity	Negative rare word diversity

meaning with respect to changes in the diversity with which rare words are used in a language sample, as summarized in Table 10.8.

RWD, then, has construct validity, and through descriptive and correlation studies of a trial with 32 texts drawn from 17-year-old lower intermediate learners of French, we provide evidence as to its face and convergent validity. The evidence is all as predicted for validity and generally strong. In particular is the striking evidence that RWD correlates positively with D_{basic}, as face validity would indicate, in spite of its mathematical dependence on *minus* D_{basic}, and that RWD correlates highly significantly with all three available measures of language competence at about the level of absolute values which would correspond to two closely related but different variables. These results are given added weight when it is known that in the above we have taken the more cautious stance of not assuming a parametric distribution, and have reported Spearman's rho. We can further report that, in the event, the same results but with even stronger absolute values were obtained from Pearson correlations – hence the evidence presented here is the more conservative of the two.

RWD can be used to measure both lexical sophistication and lexical originality/individuality and is theoretically consistent with the measure of overall lexical diversity, *D*, and related measures such as inflectional

diversity (ID) for morphological development and limiting relative diversity (LRD) for style previously proposed and summarized in Malvern et al. (2004). Unlike Advanced TTR, Advanced Guiraud and similar measures, RWD is not a function of the token size of the sample. That is not to say that there is no connection with text length. In this example, for instance, because the texts were all produced in a fixed time limit, we would expect in general terms the more competent linguists to write more fluently and produce longer texts. We would also expect more competent linguists to use more rare words and exhibit more diversity in their deployment. Hence there may well be something of a small positive correlation because of this, and that is precisely what we observe – the correlation between RWD and text length is .19 (ns). The same argument applies to traditional measures, and we might expect Advanced TTR, for example, also to have a small positive correlation with text length; but this positive link is more than balanced out by TTR's being a decreasing function of the size of the sample, and the two actually correlate *negatively* at −.18.

RWD is superior to Advanced TTR, Advanced Guiraud and the like in another respect in that unlike them, and as we have shown, it is capable of distinguishing between two language samples with precisely the same number of rare word types and tokens (and hence of the same value for Advanced TTR and Advance Guiraud, etc.) but different patterns of repetition, which represent differing diversities. That is to say, RWD will measure not just how many and how many different rare words are included, but it will also take into account *how often* each rare word type is used.

Lastly, although for clarity and because of its widespread interest to linguistic scholars, we have chosen to introduce this measure with respect to rare word deployment, the arguments for it would apply to other classes of words which occur with relative infrequency. The method could be applied to other kinds of specialized words such as a specified technical vocabulary or borrowings from a second language and the like.

Note

1. We would like to express our thanks to Jeanine Treffers-Daller for her assistance and advice on this aspect of the analysis.

11

Vocabulary Knowledge and Academic Success: a Study of Chinese Students in UK Higher Education[1]

Michael H. Daller and Huijuan Xue
University of the West of England, Bristol

Introduction

The number of Chinese students undertaking international education has been increasing steadily over the past decade, and this upward trend is still ongoing (Goh, 2007; TEIU, 2008). Study failure is obviously a major concern for both international students themselves and universities in host countries. Previous attempts to relate the study success of overseas students to their English language proficiency have used scores from standardized tests such as the International English Language Test System (IELTS) or the Test of English as a Foreign Language (TOEFL). Whilst such tests are valid tools as entry tests (Rosenfeld, Oltman and Sheppard, 2004; Taylor and Falvey, 2007), they do not seem to be good predictors of academic success on their own. We therefore tried to find other measures to complement the information provided by IELTS by using data from 23 overseas students from China in the present study. Apart from their IELTS scores, we used two measures of lexical diversity (*D* and Guiraud) and two measures of lexical sophistication (Lexical Frequency Profile and Guiraud Advanced) based on their written essays. In addition, we employed C-tests which focus on vocabulary but also measure other aspects of foreign language proficiency (Eckes and Grotjahn, 2006). The correlations we found clearly showed that the C-tests focused on lexical sophistication rather than diversity. The present study revealed the C-test to be a powerful tool in that it allowed us to predict over one-third of the variance in the modules failed by the students during their first year in UK higher education (HE). We conclude that lexical sophistication is most closely related to overseas students' academic success.

Second language proficiency and study success

The role of English language proficiency and its relation to the study success of international students have been the focus of many studies. Graham

(1987) reviewed several earlier studies, one of which was carried out with 2075 foreign students from 1964 to 1965 in the USA, showing that 'English language proficiency, as measured by the institution's own test, was one of the least predictive of the variables in the study ($r = -.046$)' (Graham, 1987, p. 509). One could argue that this raises questions about the validity of the test, for which no detailed descriptions were available. However, a number of other studies Graham reviewed similarly revealed the weak correlations between language proficiency test scores, for instance TOEFL results at entry to HE, and subsequent academic success. She drew the conclusion that there was no clear-cut relationship between English proficiency and academic success. Many other factors seemed to be involved in the study success of international students, but it was likely that a minimum threshold level of English proficiency was required before other factors became important (Graham, 1987, p. 517). Still, exactly what constitutes this minimal level remains unclear.

Patkowski (1991) carried out a study with 271 EFL students. He used a university entrance test which included reading, writing and maths subtests to predict grade point average (GPA), but only 13.7 per cent of the variance of GPA scores could be explained even when all three subtests were combined in a multiple regression. He came to the conclusion that these proficiency tests were poor predictors of academic success. Comparable findings were reported in a study of 89 students at an Australian university where Dooey (1999) investigated whether students' IELTS scores at entry were related to the average marks they obtained in the first two semesters. The study showed that IELTS scores were generally not related to study success. Only the reading subtest revealed a moderate correlation (.396) with the marks.

On the other hand, Yule and Hofman (1990) approached the correlation between EFL proficiency and academic success from a different angle. They carried out a study with 233 international graduate teaching assistants. The dependent variable was whether or not they received positive or negative recommendations for the assignment of teaching duties after a two-year study period. The authors found that there was a difference in the average TOEFL scores at the beginning of the course between the group that were successful in getting a teaching contract at the end of the course and those who were eventually unsuccessful. The mean score of the former was 607 ($SD = 39$) and the mean for the unsuccessful students was 560 ($SD = 25$). At first sight, this difference of approximately 10 per cent did not seem to be very large, but statistically it was highly significant ($p < .001$) (Yule and Hoffman, 1990, p. 231). Although the authors did not report the effect size, this could be easily computed based on the reported t-value and the degrees of freedom ($t(231) = 9.34$) (Field, 2005). In this case, it amounted to an effect size of $R^2 = 0.275$.

Feast (2002) made a further study of IELTS scores and academic success. The subjects were 101 international students at an Australian university who were monitored over a maximum period of five semesters. The results

showed that the IELTS scores had a statistically significant relationship with study success. However, other factors, including the country of birth, were significant as well. Feast concluded that 'Chinese born students [were] likely to be relatively more successful than non-Chinese born students with the same IELTS score, all other variables being equal' (Feast, 2002, p. 79). Overall, this study convincingly confirmed that there was a relationship between IELTS scores and study success but that many other factors also played a role. It is important to note here that the variability in the IELTS scores was higher than in other studies with a range of 4.5–8.5 (Feast, 2002, p. 75). This meant that students were included in this study who would not have been admitted in other institutions as the minimum entry requirement is normally an IELTS score of 6.0. It was therefore not clear whether the findings could be gener–alized to a situation where all students meet this minimum requirement.

Brooks and Adams (2002) investigated the correlation between study success and reported foreign language use in an Australian context. They made a comparative study of 32 international students with 112 local students and found that the marks of the former were lower than those of the latter. Likewise, the reported use of English outside teaching hours was also lower for the international students. Despite the absence of information about overseas students' first language, their lower reported use of English indicated that their first language (L1) was not English. According to the researchers, the use of the second language (L2) had some effect on students' academic success. Factors like educational culture or lower English language proficiency might also explain the lower marks of international students, although Brooks and Adams did not attempt to include such factors in the research design.

Bayliss and Raymond (2004) reported two studies where the scores in a language proficiency test were linked to study success. The participants in the first study were 34 Chinese overseas students enrolled on a Master's programme in Business Administration. The students were tested twice, first at the start of the programme in April 2000 and then in November 2000. The test used was the CanTEST from the University of Ottawa which included the subtests listening, reading, 'scim-scan ability', writing and a Cloze test. Study success was defined as the average mark (GPA) from the modules that the students took in this first part of their MA programme. Significant correlations were found between the GPAs and the April listening test ($r = .49$), the November Cloze test ($r = .34$) and the November reading comprehension test ($r = .62$). Interpretation of the data was difficult but one possibility could be that listening skills were important before and at the beginning of a programme whereas reading skills became more important later in their studies. Bayliss and Raymond reported that the majority of students had great problems with the volume of reading required.

In the second study discussed by Bayliss and Raymond (2004), the subjects were 136 students enrolled on a law course taught in French.

The students were classified as either being Francophone (n = 100) or Anglophone (n = 36). Both groups completed a French test which included subtests for listening, dictations, reading, writing and error correction, an item requiring students to identify errors in a written text. Weak to moderate statistically significant correlations were found between the academic success of the Francophone students and their scores in all subtests. This meant that all aspects of language proficiency played a role in academic success for such students who were taught in their first language. By contrast, the only significant correlation for the Anglophone students was between reading comprehension and average marks (r = .37). One should bear in mind that the two groups differed substantially in size and that other correlations for the Anglophone students might not be statistically significant because of the small size. Nevertheless, such findings once again indicated that reading comprehension was an important part of study success for international students and, as revealed in other studies in a monolingual and bilingual context, for instance studies by Qian (1999) and Ransdell (2002), reading comprehension is closely related to vocabulary knowledge.

Most of the studies reviewed so far operated with scores from tests such as IELTS or TOEFL which, as reported earlier, might be a problem because the variance in test scores of students entering HE is normally low. This might explain the inconsistency in findings, with either no significant correlations or only weak ones between study success and language proficiency in most studies. Tests such as IELTS and TOEFL include vocabulary knowledge as part of band descriptors (IELTS, 2008). However, in recent years there has been a shift in applied linguistics towards a clearer focus on vocabulary as a central part of proficiency in L1 and L2. This shift has been documented by a number of books and special issues of journals (Daller, Milton and Treffers-Daller, 2007a; Malvern, Richards, Chipere and Durán, 2004; Nation, 1990; Read, 2000; Treffers-Daller, Daller, Malvern, Richards, Meara and Milton, 2008). Various studies (Daller, 1999; Daller, Van Hout and Treffers-Daller, 2003) have clearly revealed vocabulary knowledge as one of the major aspects of foreign language proficiency.

One study with a focus on the correlation between vocabulary and academic success in the context of Teaching English as a Second Language (TESL) was carried out by Morris and Cobb (2003). They examined the language proficiency of 122 TESL trainees and analysed the vocabulary in 300-word samples of their writing with software based on frequency lists. The frequency bands were the most common 1000 words (K1), the next most common 1000 words (K2), the Academic Word List (Coxhead, 2000) and off-list words. In addition, they took the percentage of function words into account. Quite contrary to earlier findings by Morris and Tremblay (2002) that more proficient EFL students used more function words, which seemed at first sight counter-intuitive, Morris and Cobb (2003) found a negative correlation (r = −.34) between the use of function words and the

marks that students obtained for the two obligatory courses they took. The highest correlation ($r = .37$) was found between the students' marks and the score that they received on the Academic Word List. The authors concluded that although this was clear evidence that vocabulary profiles could predict study success to a certain extent, they were unsuitable as a single, stand-alone measure for decisions about admission to TESL training programmes.

In summary, we conclude that there is a correlation between academic success and foreign language proficiency but that such relationships may be difficult to identify for methodological reasons. The low variance of some variables used makes it hard to obtain high correlations. In addition, many other variables apart from language proficiency seem to influence students' academic success. The existing literature does not allow us to specify the extra-linguistic variables in detail. However, good candidates for successful linguistic predictors seem to be measures that focus on all skills, such as the average overall IELTS scores, measures of vocabulary knowledge and reading proficiency. The present study will therefore try to relate study success to the vocabulary knowledge and IELTS scores of Chinese students undertaking HE in the UK.

Measures of vocabulary knowledge

The measurement of lexical knowledge is a complex issue since there are various distinct aspects. Four aspects of lexical knowledge or lexical richness have been identified by Read (2000): lexical variation or diversity, lexical sophistication, lexical density and number of errors. We will focus on the first two aspects because they have been used most widely in recent research. Lexical diversity is the range of vocabulary and the avoidance of repetition and lexical sophistication is the use of sophisticated vocabulary. We include measures of lexical diversity and lexical sophistication in this study to find out which aspect of vocabulary knowledge is more important for study success.

D as a measure of lexical diversity

A number of previous studies have shown that lexical diversity measured by the relationship between the total number of words (tokens) and the number of different words (types) in a text, can be a good indicator of vocabulary knowledge. Johnson (1944) introduced the type-token ratio (TTR) as measure of lexical diversity and it has been used widely. TTR is, however, problematic as the probability of the occurrence of new words in a text decreases with text length. Therefore TTR is a function of text length and not suitable for comparing texts of different lengths. The index *D* (Malvern et al., 2004) was developed to overcome this problem. *D* is the single parameter in the equation for this falling TTR against token curve and allows comparison of speakers or writers irrespective of the length of the text produced. The higher the *D*, the greater the lexical diversity of the text. It is an indication

of a combination of the vocabulary resources of the author of a text and his or her skill in deploying them in a way that minimizes repetition.

This measure has been used in research on L1 development but also to distinguish between L2 learners at different levels. Tidball and Treffers-Daller (2007) showed that it discriminated clearly between L2 learners of French at different levels and native speakers and the effect sizes (η^2) for differences between groups in this study have a range from .61 to .67. Similarly, Daller and Xue (2007) showed that *D* was the vocabulary measure that discriminated most clearly out of six measures of vocabulary richness between two groups of Chinese EFL learners. Therefore, we conclude that *D* is a good candidate when relating vocabulary knowledge to study success of international students. In addition to *D*, we computed the index of Guiraud (types/square root of tokens) which has been used widely as a measure of lexical diversity (see also Treffers-Daller, this volume).

Lexical sophistication

A further aspect of lexical knowledge is lexical sophistication, the use of low-frequency or 'difficult' vocabulary. This aspect has been investigated in various educational contexts, including bilingual children (Daller et al., 2003; Vermeer, 2001). To measure lexical sophistication it is necessary to find a criterion to define the difficulty and/or frequency of words. One operationalization is based on frequency lists. In the present study, we used the program 'Range' (Heatley, Nation and Coxhead, 2002; Nation, URL) which made use of three bands: 1K, 2K and 3K. Words not within the bands were classified as 'Not on List' (NoL). The first two bands (K1 and K2) were based on West's (1953) word list, and the third, on the Academic Word List compiled by Coxhead (2000). The output of the program gave the percentage of words from each list and allowed us to create a lexical frequency profile (LFP) for each text. The output also allowed us to compute Guiraud Advanced (advanced types/square root of tokens). We classified as advanced all types that were not in the K1 and K2 list. This measure has been used successfully in the description of the academic profile of bilinguals (Daller et al., 2003). The measure Guiraud Advanced has also recently been used in a study based on 55 essays written by young adults in their first language (Wray, A., Mollet, Fitzpatrick, Wright and Wray, N., in preparation). The exact details are not yet published but a first result is that there are strong correlations between a verbal IQ test and Guiraud Advanced (personal communication from Eugene Mollet). The test used by Wray et al. was the verbal IQ subtest of the Multidimensional Aptitude Battery MAB-II (Sigma Assessment Systems, URL).

The C-test as a measure of vocabulary knowledge

From the literature review above we conclude that academic success can be predicted to a certain extent with standardized tests, such as IELTS, and that reading proficiency and vocabulary knowledge seem to be important factors

for study success in international education. A test format closely associated with reading proficiency and vocabulary is the C-test. This test format is a further development of the Cloze test but instead of deleting whole words only the second half of every second word is deleted. It has been used in more than 200 L1 and L2 studies (Grotjahn, 2007) and in recent research on L1 attrition (Opitz, 2008).

C-tests consistently yield significant correlations with all other aspects of language proficiency, including oral proficiency, in various studies (Eckes and Grotjahn, 2006; Sigott, 2004, 2006). For example, Arras, Eckes and Grotjahn (2002) reported a Spearman correlation of .64 between a C-test and a 'simulated oral proficiency interview' in a study of 145 learners of German. These significant correlations with all aspects of language proficiency led many researchers to the conclusion that the C-test was a test of general language proficiency (Cohen, Segal and Weiss, 1985; Grotjahn, 1992, 1995; Klein-Braley, 1985a, b; Raatz, 1985). However, the validity of the C-test and what it measures has been widely discussed. Alderson viewed its validity as a 'worrisome question' (2002, p. 28) and argued that 'claiming that there was a unitary competence, or a general language proficiency … [was] now generally discredited' (Alderson, 2002, p. 21).

The concept of a unitary competence underlying different skills in a foreign language was developed in the 1970s (Daller, 1999). The main argument for such a competence was the high correlations that could be found between tests of different aspects of foreign language proficiency. Despite Alderson's claim that this hypothesis was now discredited, these high correlations were still found in later studies. As Singleton and Singleton (2002) noted:

> High correlations have been found between sets of scores from tests purporting to measure grammatical knowledge and sets of scores from tests purporting to measure lexical knowledge, and there has been little success in attempts to demonstrate that 'grammar tests' and 'vocabulary tests' tap fundamentally distinct aspects of linguistic knowledge. (p. 154)

C-test scores are found to be correlated with scores in the four classical skills (reading, writing, listening and speaking) in many studies rather than with tests of vocabulary. However, a closer look at published research findings seems to confirm that the C-test operates largely at the lexical level. Vocabulary is relevant to all sub-skills; therefore the concept of general language proficiency can perhaps be reinterpreted, at least to some extent, as vocabulary knowledge.

Little and Singleton (1992) administered C-tests in French and German to university students (L2 learners), and analysed the items that posed particular difficulties. They viewed the test as 'an instrument with a clear lexical focus' (p. 175) and concluded that in filling C-test slots the subjects gave 'priority to a ready lexical solution over morphosyntactic and more

general semantic issues' (p. 188). Stemmer (1992) also carried out a study on the C-test with reading aloud protocols. She drew the conclusion that the informants operated predominantly within one meaning unit, whilst higher processing strategies at a macro level were less involved. This strongly suggests that the C-test does indeed have a lexical focus. Sigott (2004) administered decontextualized C-test items (containing truncated words only) and a C-test in its canonical form to 60 university students of English in Austria. He confirmed that students with a higher English proficiency operated at the lexical level when filling in C-test gaps. Students with a lower proficiency depended more on contextual information than those with a higher proficiency. In Sigott's study, English proficiency was measured with the Oxford Placement Test (OPT) which showed a high correlation with the C-test results (r = .83). The OPT was meant to measure the foreign language proficiency of the students in relation to their potential academic success. The high correlation between the OPT and C-test scores indicated that the latter may also be a good candidate to measure academic success (see also below).

The view that the C-test is to a large extent a vocabulary test is also supported by other correlational studies. Grotjahn and Stemmer (1985) carried out a study with 115 students of French, adopting a C-test and the Bochum Diagnostic Test for French (BDTF). The highest correlations (r = .63) were found with the two subtests 'pronouns' and 'verbs' of the BDTF. Klein-Braley (1985b) used the Duisburg Diagnostic Test for English (DELTA) and a C-test over a period of four years with students (N = 202) in the English department. She found high correlations between all subtests of DELTA and the C-test. In a factor analysis the highest loading on the first factor was found for the vocabulary subtest (loadings between .83 and .94). Furthermore, Cohen et al. (1985) reported C-test studies carried out in Hebrew. They concluded that the test encouraged micro-level processing and that 'students who did not understand the macro-context could still mobilize their vocabulary skills' (p. 125).

There are also clear indications that C-test scores are related to academic success. In a study with 358 pupils at secondary schools in Germany, Klein-Braley (1985a) showed that C-test scores in the first language were directly linked to grade and school type. Pupils who attended school types that lead to HE (*Gymnasium*) consistently had higher C-test scores than those that prepared for vocational training (*Realschule*, *Hauptschule*). This finding is supported by a study carried out with 75 pupils (mean age 10.8) by Raatz (1985) who found a correlation of .51 between the C-test scores and a test of nonverbal intelligence (figure completion and figure sequence completion tasks). Furthermore, Coleman (1994) employed a C-test to investigate the language proficiency of students enrolled at foreign language degrees in the UK. His study showed that the highest correlation between a C-test in French and five subsets of A level exams was with the subset 'reading and writing' (r = .78).

Overall, we conclude that the C-test has a specific focus on vocabulary and is potentially a good predictor of academic success. The specific focus on vocabulary might be precisely the underlying cause for the high correlations between C-test scores and test scores in all other aspects of language proficiency, including the important aspect of reading in a foreign language. Since they have to fill in gaps that are created randomly (every second word is truncated), test-takers also encounter different words and, depending on the text, infrequent words. The C-test format therefore taps into lexical diversity as well as sophistication. In addition, we conclude that there is a relationship between C-test scores and IQ scores which will also contribute towards the usefulness of this test format for predicting academic success.

The present study tested the following hypotheses:

1. There is a relationship between academic success and language proficiency.
2. Vocabulary knowledge is an important aspect of study success in a foreign language and can be used to predict academic success.
3. C-test scores are an indication of vocabulary knowledge and are therefore related to academic success.

An additional research question was whether lexical diversity or lexical sophistication was more important for academic success.

Subjects and data collection

The subjects in the present study were 23 Chinese students attending a British university and following a one-year taught course for postgraduates in a business school. The first data collection took place in China in February 2004, half a year before the students came to Britain. A C-test was administered under controlled conditions as part of the selection process. The students also obtained IELTS scores in China. Those with scores between 5.5 and 6.0 attended an eight-week pre-sessional English language programme in the British university before starting their postgraduate studies in September 2004. These students took the C-test at the beginning and end of the programme. Some with a score above IELTS 6.5 attended a three-week pre-sessional course and others did not. Before the start of their Master's programmes, the same C-test was administered to these students again and they were also asked to write an essay on 'The Internet'. They were informed this time that the test was to investigate their English proficiency and would not affect starting their studies. The essays were then transcribed into the CHAT format which allowed the computation of *D* and other measures with the CLAN software (MacWhinney, 2000). Spelling mistakes were corrected but the texts were not lemmatized. The number of types, tokens and *D* were computed. Students' IELTS scores and previous

learning history in China were obtained as additional background information (e.g. whether they had already been awarded academic degrees or not before taking UK HE, and whether they switched subjects in Britain).

Predictor and dependent variables

The following predictor variables were used in the present study:

- The IELTS scores prior to entry into UK HE;
- Whether or not the student had been awarded a Bachelor's or similar degree in China before coming to the UK;
- The values for *D*, Guiraud, LFP and Guiraud Advanced obtained from the essays written by the students;
- C-test scores from February 2004 and September 2007.

We operationalized study success with two variables:

- Whether a student failed at least one module in the first year or passed all modules at the first attempt;
- The number of modules failed in the first year.

Results

Predictor variables

The students' IELTS scores are listed in Table 11.1. It should be noted that these scores were obtained in China before their pre-sessional course. According to the teachers on this course, the IELTS scores typically rose by half a band in an eight-week course.

Students' educational background is shown in Table 11.2. The variables were whether the students had studied the same subject at home

Table 11.1 Distribution of IELTS scores

IELTS score	*Number of students*
5.5	1
6.0	10
6.5	9
7.0	3

Table 11.2 Educational background of the students

	Obtained degree in China	*Subject switched*
Yes	9	20
No	14	3

or whether they switched subjects in Britain, and whether they had been awarded a Bachelor's degree in China.

It was surprising that most students had chosen to take a different subject (a Master's in Business) instead of pursuing what they had studied previously. This could, however, be explained as a clear change in career perspective by the students.

Figure 11.1 shows the distribution of *D* scores. Apart from one outlier all lay between 48 and 110 with a median of 80. The standard deviation of 20.56 indicated a good spread of the scores, revealing the potential for this measure to be a good predictor if the range of vocabulary usage was indeed related to individual differences in academic achievement in the students from China.

The results for both C-tests illustrated in Table 11.3 and Figure 11.2 reveal the spread of two test scores transformed into percentages. Both C-tests turned out to be highly reliable: Cronbach's alpha was .873 for February 2004 and .876 for September 2004. The correlation between the two tests was significant and moderately strong ($r = .604$, $p = .008$, $N = 18$).

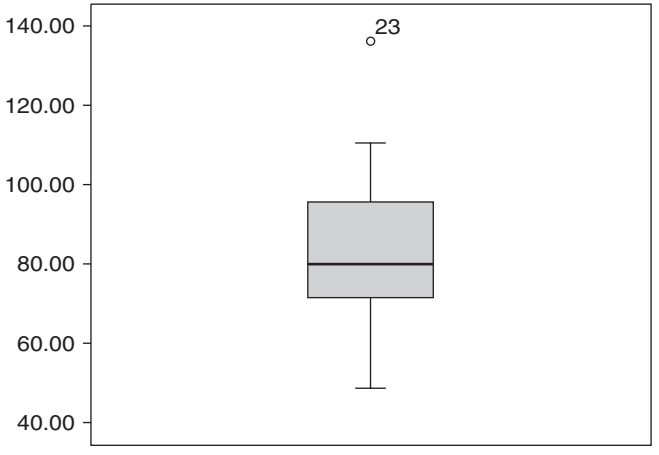


Figure 11.1 Spread of *D* scores

Table 11.3 C-test results

C-test	*Number*	*Min.*	*Max.*	*Mean*	*SD*
Feb. 2004	21	30.83	71.67	55.27	13.26
Sept. 2004	20	43.13	83.75	58.81	10.81

Table 11.3 shows a slightly higher mean for September 2004. However, it can be seen from Figure 11.2 that there is little, if any, difference in the median scores and overall the difference was not significant (paired *t*-test, $t = 1.42$, d.f. = 17, $p = .173$). Note that there were only 18 paired test scores because not all students did both tests.

Modules failed

As mentioned earlier, we operationalized study success in two ways. Firstly, we divided the group into those who passed all modules at the first attempt and those who had at least one failed module. Only 9 out of 23 students passed all modules at their first attempt and 14 students failed one or more modules (see Table 11.4). Failure seemed to be a serious issue for the cohort. More than half failed one module and over a quarter of the students failed four or more modules while taking taught courses in their first year.

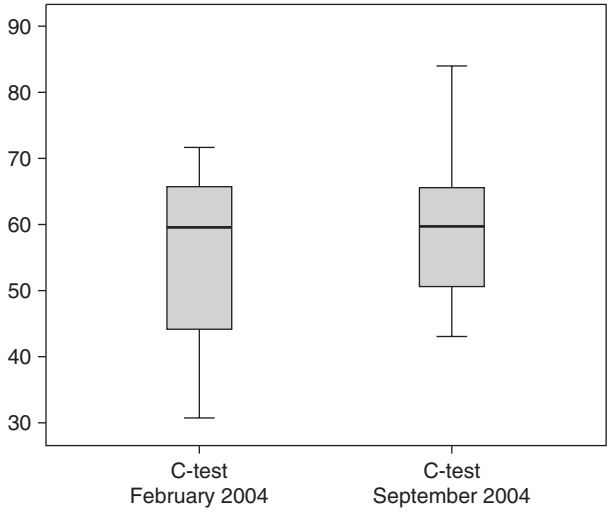


Figure 11.2 The spread of C-test scores

Table 11.4 Failed modules for the Chinese students

	Number of failed modules								
	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Number of students	9	5	1	2	4	1	0	0	1

Predicting academic success

In order to investigate the contribution that variables make to predicting study success, we computed the correlations between variables and the number of failed modules. Two variables (IELTS and C-test February 2004) yielded significant results with a one-tailed test. The Spearman correlation between IELTS scores and the number of failed modules was −.382 (p = .036, N = 23). An even stronger negative correlation was found between the C-test scores from February 2004 and the number of failed modules (−.565, p = .004, N = 21).

Figure 11.3 illustrates the relation between C-test and the number of failed modules, with a linear line of best fit included. The data point for the student who failed eight modules was not included in the graph since no C-test data were available.

No other variable correlated significantly with the number of failed modules. However, some intercorrelations were significant and gave some insight into the specific focus of such variables (see Table 11.5). The only variable from the Lexical Profile Analysis to enter into significant correlations was the number of types not on the list (types NoL) and the number of tokens not on the list (tokens NoL). The C-test scores are also included in Table 11.5 to investigate the lexical focus of the test.

As can be seen from the table, *D* and Guiraud were significantly intercorrelated which is in line with expectations, as both measures focus on lexical diversity. Guiraud Advanced and types NoL were also significantly

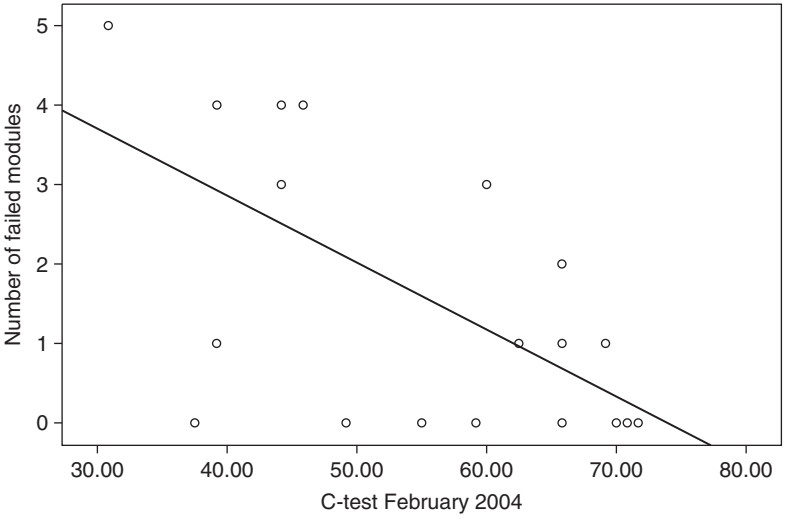


Figure 11.3 C-test scores and number of failed modules

Table 11.5 Intercorrelations between lexical measures (Pearson)[a]

	C-test	*D*	*Tokens NoL*	*Types NoL*	*Guiraud*	*Guiraud Advanced*
C-test	–	.135	.393	.522**	.167	.613 ***
D	–	–	.236	.475**	.564***	.275
Tokens NoL	–	–	–	.737***	.422	.616***
Types NoL	–	–	–	–	.369	.696***
Guiraud	–	–	–	–	–	.518**
Guiraud Advanced	–	–	–	–	–	–

[a] $N = 19$ for all intercorrelations with the C-test, otherwise $N = 21$.

** Significant at the .05 level.

*** Significant at the .01 level.

correlated which is an indication that they both measure the same aspect of vocabulary knowledge, in this case lexical sophistication. The C-test was significantly correlated with these two measures of lexical sophistication but not with the two measures of diversity. This is a clear indication that the C-test is sensitive to lexical sophistication rather than just the range of one's vocabulary.

We also wanted to predict the variable 'Fail' which put the students into two categories: those who passed all modules and those who had at least one fail at the first attempt. We used a logistic regression with 'IELTS scores' initially as the only predictor variable. This model was significant ($\chi^2 = 8.416$; d.f. = 3; $p = .038$; d.f. was 3 because there were four IELTS levels in the study: 5.5, 6.0, 6.5 and 7.0). This statistical procedure produced beta values for each 'IELTS score' separately, and the highest negative beta value was obtained for IELTS score 5.5 ($\beta = -38.56$), followed by score 6.0 ($\beta = -20.7$), 6.5 ($\beta = -19.5$) and score 7.0 (β set to zero). The lower the IELTS score, the more likely it was, therefore, that the student failed a module. The same computation was carried out with the C-test scores but no significant model could be obtained.

Overall, the present study showed that failure of at least one module could be predicted by IELTS scores. The lower the IELTS scores, the higher risk of failing at least one module. IELTS scores also explained about 11 per cent of the variance of the number of failed modules.

Interestingly, none of the measures taken in September 2004 in the UK before the students started their studies, including the C-test, predicted their academic success. One possible explanation could be that the C-test taken in February 2004 in China had been administered under strictly controlled conditions, with several British and Chinese staff monitoring, and the test was taken as part of the admissions procedure. By contrast, the students knew in September 2004 that they had secured a place at a British university and that tests and essays would not affect their further study.

This might have impacted on their motivation and reduced the validity of the measures (for an in-depth discussion on test-taker motivation and test validity see Nation, 2007). The relatively modest correlation between the two C-tests could be a further indication of decreased motivation, although firm conclusions were not possible with the existing data set.

Conclusion

The present study shows that it is possible to predict study success and failure before overseas students come to the UK. English proficiency tests such as IELTS and a C-test are useful predictor variables. The most powerful predictors for the number of failed modules are the C-test scores obtained more than half a year before the students came to the UK. Almost 40 per cent of the variance in the number of failed modules during the first-year taught course of Master's programmes can be predicted from their C-test scores. In addition, we used measures of lexical diversity (*D* and Guiraud) and measures of lexical sophistication (Guiraud Advanced and LFP) derived from students' essays. The present study reveals that the measures of lexical sophistication correlate significantly with each other and so do the measures of lexical diversity, which supports the validity of the measures used. However, the C-test correlates significantly only with measures of lexical sophistication (and not with measures of lexical diversity) which is an indication that knowledge of infrequent words may be related to academic success in a foreign language. The highest correlation between the C-test and any measure of vocabulary richness was found with Guiraud Advanced. As a recent study has shown, Guiraud Advanced appears to be related to verbal intelligence. This and earlier findings on the relationship between the C-test format and intelligence scores lead us to the conclusion that lexical sophistication and (verbal) intelligence both play a role in C-test completion. Therefore this test format is an excellent tool to predict study success.

The present study provides implications for admissions procedures. Language tests which tap into lexical sophistication and verbal intelligence can be developed to measure overseas students' ability to study in a foreign language before they embark on international education.

Note

1. We are very grateful for the comments on an earlier draft of this chapter made by Brian Richards and an anonymous reviewer. We would like to thank Paul White for his advice on statistics.

12

Vocabulary Size and the Common European Framework of Reference for Languages

James Milton and Thomaï Alexiou***
**Swansea University, UK*
***Aristotle University, Greece*

Introduction

In its earliest stages of development the Common European Framework of Reference for Languages (CEFR) included vocabulary lists in its materials and these gave some indication of the scale of the vocabulary knowledge that the creators were envisaging at the various levels of the framework. More recently these have been removed and learners, textbooks and course syllabuses are placed into the framework levels according to skills-based rather than knowledge-based criteria (Council of Europe, 2003). The purpose of this chapter is to see what happens when vocabulary size measures are placed back into the framework and there are two reasons for wanting to do this. One is academic interest in seeing what vocabulary sizes emerge at the CEFR levels and considering how these compare across levels and across languages. The second reason is a practical one and is to help to make the framework more robust. The skills-based criteria have the virtue of making the framework flexible and highly inclusive, and almost any course, textbook or learner should be able to find a place in the system. However, the penalty for such flexibility is that the levels become imprecise; it is often possible to place learners or textbooks at several of the CEFR levels. This potentially devalues the framework and diminishes its usefulness. The British foreign language exam system in schools, for example, has been criticized for being misplaced within the system and, as a consequence, for misleading those who try to use it (Milton, 2007a). The presence of a more objectively assessed, knowledge-based measure, such as vocabulary size, ought to help avoid this kind of ambiguity and the problems associated with it.

In this chapter, therefore, we intend to review the evidence we have from a variety of learners in different countries and learning different languages where we are able to tie vocabulary size scores to different levels in the CEFR hierarchy. We intend to draw on results we have from learners

in Spain, Hungary, Greece and Britain, learning English, French and Greek as foreign languages. By looking at modern foreign language learners at different levels we should be able to see whether the vocabulary knowledge changes systematically as the language level of learners increases. It would be expected that learners at the lower levels of the hierarchy, A1 or A2, would know fewer words and expressions in the foreign language than learners who are very advanced and who are taking courses or studying for exams at C1 or C2 level. This in turn raises the possibility of comparing vocabulary knowledge levels across languages and across different language systems; is the knowledge of French learners in Britain, say, comparable in some meaningful way to the knowledge of Greek learners of French or even Greek learners of English? It is not always obvious how to compare knowledge across different languages but one method for comparing vocabulary sizes will be proposed and examined. This should begin to tell us whether the CEFR hierarchies are as robust as we would like them to be, and whether vocabulary size measures can help to add a useful degree of precision to the difficult art of placing learners at the correct CEFR level.

Background to the CEFR and the place of vocabulary knowledge within it

The CEFR was created to provide a framework of comparison in the study and testing of languages. There were many issues involved in creating such a framework which has taken over 25 years to accomplish. It requires, for example, the development of a common set of terms and references so that professionals across Europe can speak to each other on aspects of language learning and language level, and be confident that what they intend to convey will be understood in the same way. For most users, that is learners, parents, teachers and employers, the most obvious intention of the framework is to bring order to the plethora of courses, exams and awards which learners can take. Even within a single language it was frequently unclear how one exam related to another in its demands and in its difficulty. Students we have dealt with in Greece often confidently assert that the Michigan Proficiency exam is easier than the Cambridge Proficiency in English exam, although there is very little evidence to suggest whether this is the case or not. How should this kind of opinion be interpreted? Would it be appropriate to value a pass in the Cambridge exam more highly than the Michigan in determining, say, whether a candidate has the qualifications for entry to a university course requiring a language qualification? The presence of the CEFR, even if it is no more than a common vocabulary to describe the hierarchy of levels, ought to allow questions like these to be answered rather better. It should allow exams, for example, to be placed within a framework so that users can see which exams are intended to be at different levels and which are intended to be similar.

The CEFR is not intended to be specific to one country; it is designed to be a common framework which can be applied to language courses and exams across Europe. It ought to tell us, for example, whether learners from Germany, Spain and Italy with school leaving certificates in a foreign language are of the same standard so we can know whether they could enter a course of study requiring a set level of ability. Or it might tell us whether these learners have the foreign language ability required for a job. In terms of textbooks and teaching materials, the CEFR has been taken up by the EFL world in particular and by EFL publishers. It is now common for textbooks and for language courses to be described in terms of the framework. Therefore, in principle at least, a course designed for learners at, say B1 level ought to be able to select materials from a range of textbooks designed for students at this level and all of them should be appropriate in some meaningful sense. Additionally, the CEFR should, in principle, allow direct comparison between learners, courses and course books in different languages. Because the framework is not language specific, by implication it should allow intelligent comparisons between exams or learners of Italian, German, Greek or any other language. In Britain, for example, the age 18 Advanced level foreign language exams are pitched at the CEFR B2 level and so Advanced level students of, say German, should have the same kind of knowledge and skills as learners of Italian also taking Advanced level. Both of these should be comparable with learners in other countries following courses at B2 level in German and Italian.

At the outset of the project which created the CEFR the descriptors which were created included word lists. The *Threshold* level materials (for example, Coste, Courtillon, Ferenczi, Martins-Baltar and Papo, 1987; van Ek and Trim, 1990) and some of the *Waystage* materials (for example, van Ek, 1990) contain such lists. The level descriptors are generated from the notional-functional categories which underlie the framework. While this mode of analysis now looks rather old-fashioned, the word lists they contain are, nonetheless, both useful and usable. The word lists at *Threshold* level (CEFR B1) contain about 2000 words and the *Waystage* level (CEFR A2) materials contain word lists with about 1000 words. However, the overall framework document (Council of Europe, 2003) has concentrated on skills and can-do lists, and language-specific items, such as the word lists, are absent. No one is saying, of course, that the skills which define the framework are divorced from language knowledge such as vocabulary knowledge. The word lists have not been disowned by the framework. Nonetheless, they appear to have receded into the background and the scale of vocabulary knowledge which might reasonably be associated with the CEFR levels is now an unknown quantity.

There is a case for arguing that a measure such as vocabulary size ought to fit well into a hierarchy of level such as the CEFR. There is growing evidence that vocabulary size measures correlate well with overall measures of

language ability such as scores on the Cambridge IELTS test (Milton, Wade and Hopkins, forthcoming). They also correlate well with all four skills, and with reading and writing in particular. Staehr (2009), for example, is able to gain correlations of .83 between scores gained by his 88 testees on a test of receptive vocabulary size (Schmitt, Schmitt and Clapham, 2001) and on a multiple choice test of reading comprehension. A correlation of .73 was found with the same group between vocabulary scores and assessments on an academic writing task. Both correlations are statistically significant. While lower correlations are found by Staehr with listening and speaking skill scores, his results still explain between 35 and 40 per cent of variance in the scores for these skills. Using a combination of both phonological and orthographic tests of vocabulary size, Milton et al. (forthcoming) are able to explain over 40 per cent of variance on scores in IELTS speaking and listening sub-skill scores. This suggests, perhaps not surprisingly, that the skills of listening and speaking access different lexical resources from reading and writing. Listening and speaking rely on aural word knowledge; reading and writing on knowledge of the written form of words. Nonetheless, they emphasize how important vocabulary knowledge is to all language skills. Other studies suggest that vocabulary size scores correlate well with hours of instruction and teacher assessments (Orosz, 2007) and with the size and frequency distribution of the vocabulary content available to learners in course books (Vassiliu, 2001).

There is good reason for thinking, therefore, that if the CEFR has validity as a hierarchy of language level and ability, then each succeeding increase in level in the CEFR should be matched by an increasing demand in the vocabulary knowledge of the learners who take exams at that level. With each progressively higher CEFR band, there should be higher mean scores on vocabulary size measures with groups of learners. If this were not seen then the validity of the CEFR would be called into question. The word lists in the early CEFR materials appear to reflect this pattern. The *Threshold* (B1) level word lists are indeed larger, implying greater knowledge by learners at this level, than the *Waystage* (A2) material lists. The information included with Meara and Milton's (2003, p. 5) Swansea Levels Test (*XLex*), explicitly links the EFL vocabulary size scores to attainment in Cambridge EFL exams and these exams, of course, have a place in the hierarchy of CEFR levels. The range of scores they suggest for each level is shown in Table 12.1.

Vocabulary size measures also have a distinct benefit in language measurement terms, of being, or appearing to be, more countable and therefore objective, than the kind of subjective evaluations of level which abound in other aspects of language. Modern methodology, for example Meara and Jones's *Eurocentre's Vocabulary Size Test* (EVST) (1990), allows a numerical estimate of a learner's vocabulary to be made, and a learner with, say, 2000 words out of the 10,000 in this test, can be argued to have double the knowledge of another learner with only 1000 words. It is impossible in the current state of knowledge to characterize knowledge of grammatical

Table 12.1 EFL vocabulary size, formal EFL exams and the CEFR (from Meara and Milton, 2003, p. 5)

CEFR level	*Cambridge exam*	*XLex score (max. 5000)*
A1	Starters, Movers and Flyers	<1500
A2	Kernel English Test (KET)	1500–2500
B1	Preliminary English Test (PET)	2750–3250
B2	First Certificate in English (FCE)	3250–3750
C1	Cambridge Advanced English (CAE)	3750–4500
C2	Cambridge Proficiency in English (CPE)	4500–5000

structure, or ability in a language skill such as reading, in this way, and subjective judgements have to be made on these matters. No matter how carefully these judgements are criterion referenced, it is very difficult for assessors to apply them consistently across the millions of foreign language learners we have in Europe. The presence of a vocabulary size measure, if this can be linked to CEFR levels, ought to make any hierarchy of levels more robust. It would introduce an element of objective assessment and knowledge-based assessment into the process of placement which, as it currently stands, is entirely subjective.

Measuring vocabulary size

Recent years have seen the development of rather more systematic and principled methods for estimating the vocabulary knowledge in foreign language learners. There is considerable evidence that there is a strong frequency effect in the learning of foreign language vocabulary (for example, Milton, 2007b). In effect this means that the more frequent a word is then the more likely it is to be learned. This is not a perfect rule, of course. Word learning will also be dependent on what thematic material the learner has been exposed to in textbooks and on word difficulty factors such as whether the words encountered are cognate or not. But frequency still has a very powerful effect, probably more powerful than the other factors and, as a consequence, recent vocabulary tests have drawn on frequency information and focused their test items in the most frequent bands. Nation's *Vocabulary Levels Test* (Schmitt et al., 2001), Meara and Jones's *EVST* (1990) and Meara and Milton's *Swansea Levels Test XLex* (2003) all do this, for example. What emerges from these tests appears to be good characterizations of learners' vocabulary knowledge. In the case of the latter two they provide believable estimates of vocabulary size within the frequency bands they test.

In this chapter we have used vocabulary size estimates arrived at using the *XLex* test which has the virtue of having equivalent versions available in English, French and Greek. All three make estimates of knowledge of the

most frequent 5000 lemmatized words in these languages. The frequency information in English is drawn from Nation (1984) and Hindmarsh (1980), in French from Baudot (1992) and in Greek from the Hellenic National Corpus (Hatzigeorgiu, Mikros and Carayannis, 2001). *XLex* asks learners to respond to 120 test items presented in a yes/no format. The words are presented in turn and learners must respond either 'yes' they know the word, or 'no' they do not know the word; 100 real words are included, 20 drawn from each of the first five most frequent 1000 word bands. In addition the test contains 20 pseudo-words, words constructed to look and sound like real words but which do not exist, and therefore cannot be recognized. The responses to these words allow the responses to real words to be adjusted for guesswork and overestimation. While the yes/no task appears simple it can be quite difficult where a word is only vaguely recognized or is partially known. It presents a challenge to even the most scrupulously honest learners as to how best to answer some of the items which are only vaguely recognized. The pseudo-words allow some kind of recognition of this difficulty, and compensation to be made for the differing strategies which learners may employ. A score of 50 is given for each 'yes' response to a real word and a deduction of 250 is made for each 'yes' response to a pseudo-word. The scores that emerge are estimates of the number of words that each learner has identified out of 5000. It is common to eliminate data which demonstrate an unacceptably high level of pseudo-word recognition, and are arguably unreliable as a result. However, there is no set level at which a set of answers moves from being reliable to unreliable and in analysing data for this chapter we have not eliminated such sets of responses. While we now have a lot of experience in pseudo-word construction in English, we know much less about the way these things perform in French or Greek.

Subjects and method

In EFL, the vocabulary size scores have been recorded in a state secondary school in Hungary (Orosz, 2007) and in a private language school in Greece (data from Milton, 2007b). The learners have been grouped according to the CEFR level of the class they are in and, where appropriate, the CEFR level of the exam they are taking. The learners in Greece routinely take the Cambridge Preliminary English Test (PET) at level B1, Cambridge First Certificate in English (FCE) at level B2 and Cambridge Certificate of Proficiency in English (CPE) at level C2. Data were collected from 88 Greek learners at all CEFR levels. The learners in Hungary take the state maturity exams at levels B1 and B2. Data were collected from 144 Hungarian learners at these levels.

In French as a foreign language the vocabulary size scores have been recorded in a state secondary school and university in Britain (Milton, 2006, 2008), two private language schools in Greece, and from two schools in the Spanish state education system. The learners have been grouped

according to the CEFR level of the class they are in and, where appropriate, the CEFR level of the exam they are taking. The learners in Britain take GCSE exams at B1 level, Advanced levels at B2 level, and we have results for British university graduates in French which we have assumed will be at C2 level although there is no formal statement on the part of the university to confirm that this is the case. Data were collected from 155 learners at these levels. The French learners in Greece do not appear to be taking any formal exam but are grouped for teaching into CEFR levels. Data were collected from 65 Greek learners of French at all CEFR levels. Like the Greek learners, the learners of French in Spain are grouped for teaching according to CEFR levels and are not, to the best of our knowledge, taking formal exams. Data from 50 Spanish learners of French were collected at all CEFR levels.

In Greek as a foreign language, data have been collected from learners at the Centre of Modern Languages in Thessaloniki. Data were collected from 64 learners, from a variety of first language backgrounds, at CEFR levels A1, A2, B1 and B2.

Vocabulary size and CEFR levels in English

The mean vocabulary size scores at each CEFR level from the 88 EFL learners in Greece are presented in Table 12.2. The mean vocabulary size scores for the 144 learners in Hungary at CEFR levels B1, B2 and C1 are presented in Table 12.3. The *XLex* scores suggested in Meara and Milton (2003) are

Table 12.2 EFL vocabulary size and the CEFR among learners in Greece

CEFR level	*XLex*	*Mean*	*Max.*	*Min.*	*SD*	*n*
A1	<1500	1477.27	2100	150	580.37	22
A2	1500–2500	2156.81	3250	700	664.45	22
B1	2750–3250	3263.63	4000	2750	434.79	11
B2	3250–3750	3304.54	4350	2550	666.50	11
C1	3750–4500	3690.90	4300	2650	471.07	11
C2	4500–5000	4068.18	4500	3700	261.02	11

Table 12.3 EFL vocabulary size and the CEFR among learners in Hungary

CEFR level	*XLex*	*Mean*	*Max.*	*Min.*	*SD*	*n*
A1	<1500					
A2	1500–2500					
B1	2750–3250	3135.90	4700	1130	434.79	66
B2	3250–3750	3668.42	4950	1880	666.50	72
C1	3750–4500	4340.00	4650	4000	471.07	6
C2	4500–5000					

included for reference although it should be noted that the learners in this case are in classes preparing to take EFL exams at the various CEFR levels, while the *XLex* scores in Meara and Milton are for those actually taking the exams.

The results, superficially at least, look rather persuasive. There is a hierarchy of CEFR levels in each case, and in each case also there is a hierarchy of mean vocabulary size scores. Learners at A1 appear, on average, to know fewer words than learners in level A2 who, in turn, know on average fewer words than those in level B1, and so on up the levels. Even with relatively small numbers it is possible to argue that this tendency is statistically significant. An ANOVA on the Greek data confirms that there are significant differences between the means at different levels, $F(5, 82) = 50.197$, $p < .01$, and the same is true of the Hungarian data, $F(2, 141) = 14.896$, $p < .01$. The two systems also appear, from this limited sample, to be similar and conform quite closely to the levels of vocabulary knowledge suggested by Meara and Milton (2003), especially at the lower levels. Both systems suggest considerable vocabulary knowledge is required, approximately 3000 words, before learners progress from the elementary stages of performance at A1 and A2 level to intermediate B1 level, and a score in region of 3500 words is associated with B2 level. Learners at Advanced levels know even more than this. These encouraging similarities in mean scores disguise considerable individual variation, however, as the maximum and minimum scores and standard deviations reveal. While the mean scores for groups suggest an encouraging general tendency, it seems likely that there are no clear thresholds where a certain minimum score is a requirement of passing from one level of skill or ability to another. The reasons why this might be so are discussed later in the chapter.

Vocabulary size and CEFR levels in French

The mean vocabulary size scores at each CEFR level from the 155 French as a foreign language learners in Britain at CEFR levels B1, B2 and C2 are presented in Table 12.4 The mean vocabulary size scores for the 65 learners

Table 12.4 French as a foreign language vocabulary size and the CEFR among learners in Britain

CEFR level	*Mean*	*Max.*	*Min.*	*SD*	*n*
A1					
A2					
B1	952.04	1900	0	440.28	49
B2	1882.58	3650	650	562.21	89
C1					
C2	3326.47	4150	2050	711.75	17

Table 12.5 French as a foreign language vocabulary size and the CEFR among learners in Greece

CEFR level	*Mean*	*Max.*	*Min.*	*SD*	*n*
A1	1125.71	2550	0	620.40	35
A2	1756.25	2500	1500	398.60	8
B1	2422.72	3400	1800	517.37	11
B2	2630.00	2850	2250	251.49	5
C1	3212.50	3750	2600	473.24	4
C2	3525.00	4150	2900	883.88	2

Table 12.6 French as a foreign language vocabulary size and the CEFR among learners in Spain

CEFR level	*Mean*	*Max.*	*Min.*	*SD*	*n*
A1	894.44	2850	350	604.61	18
A2	1700.00	2750	500	841.50	9
B1	2194.44	3100	1100	717.39	9
B2	2450.00				1
C1	2675.00	3600	1900	643.23	6
C2	3721.42	4200	3200	416.19	7

of French as a foreign language in Greece at all CEFR levels are presented in Table 12.5. The mean vocabulary size scores for the 50 learners of French as a foreign language in Spain at all CEFR levels are presented in Table 12.6. There are no guideline scores for what vocabulary sizes might be associated with each level.

There is rather more variation apparent in the French learners than in the EFL data but in one respect, at least, the results are consistent and encouraging. As with the EFL data there is a hierarchy of mean vocabulary size in each set which rises in line with the CEFR levels. An ANOVA using the British data confirms that there are significant differences between the mean scores at each CEFR level, $F(2, 152) = 126.055$, $p < .01$. A Tukey analysis further confirms that the difference between the mean score at each CEFR level in these data is also statistically significant. The Greek and Spanish data also confirm the relationship. ANOVAs give the results $F(5,59) = 23.713$, $p < .01$ for the Greek data and $F(5, 44) = 21.401$, $p < .01$ for the Spanish data.

It is less easy than with the EFL data to suggest that there is much consistency across the CEFR levels in different countries. While in Spain and Greece learners seem to need to know, on average, over 2000 French words to progress beyond the elementary A1 and A2 levels, the British data suggest that this can be achieved with less than half this number, fewer than

Table 12.7 Greek as a foreign language vocabulary size and the CEFR among learners in Greece

CEFR level	*Mean*	*Max.*	*Min.*	*SD*	*n*
A1	1492.10	2400	500	705.58	19
A2	2237.50	3150	1500	538.58	12
B1	3338.23	4150	1950	701.13	17
B2	4012.50	4750	3450	415.33	16
C1					
C2					

1000 words. In all cases the figures suggest that many fewer words in French are required to achieve this level of proficiency than in EFL. As with the EFL data, the mean scores, and general tendency of groups, disguise the wide variety of individual scores which the standard deviation figures illustrate.

Vocabulary size and CEFR levels in Greek

The mean vocabulary size scores at each CEFR level from the 64 Greek as a foreign language learners at CEFR levels A1 A2, B1 and B2 are presented in Table 12.7. As with the French figures, there are no guideline scores for what vocabulary sizes might be associated with each level.

As with all the other sets of data, the Greek figures reveal a hierarchy of vocabulary size scores for each successive CEFR level where we have results. Again, an ANOVA confirms the relationship between CEFR levels and differences in vocabulary size, $F(3, 60) = 57.150$, $p < .01$, and the differences between the mean scores at each level are significant. The mean scores at each level in these data are larger than the scores in the EFL and French as a foreign language data. Again, there is considerable variation of individual scores within each level and overlap in vocabulary scores between the levels.

Vocabulary size and CEFR levels

The data from users of the CEFR system, collected from four countries and three different foreign languages, show what one would expect. As learners get better in their foreign languages, and move upwards through the CEFR levels, they tend to know progressively more vocabulary. Regression analysis allows the relationship between a learner's vocabulary size and the CEFR level he or she has attained to be modelled and suggests just how strong the relationship between the two variables can be. A series of these analyses have been carried out on the data collected for this chapter and give the results shown in Table 12.8.

It appears that in Spain and Greece the CEFR level a learner achieves is particularly sensitive to their vocabulary knowledge; 60–70 per cent of variance

Table 12.8 Linear regression modelling the relationship between vocabulary size and CEFR level [a]

Learners	*R*	R^2	*Adjusted* R^2	*Standard error of estimate*
EFL learners in Greece	.842	.708	.705	0.9465
EFL learners in Hungary	.417	.174	.168	0.5229
French FL learners in Britain	.664	.441	.437	0.7065
French FL learners in Greece	.809	.654	.648	0.8562
French FL learners in Spain	.825	.681	.675	1.0519
Greek FL learners in Greece	.844	.713	.708	0.8480

[a]All regressions are statistically significant.

in CEFR levels can be explained by vocabulary size. In Britain there is still a strong relationship and over 40 per cent of variance can be explained in this way. This observation fits well with other observations (Milton, 2006; Richards, Malvern and Graham, 2008) that exam success in foreign languages in Britain is related to vocabulary size. Only in Hungary does the strength of this relationship diminish. It is not immediately obvious why these data should be so very different from the others.

In the EFL data it appears that there is some agreement on actually what levels of vocabulary might be associated with each CEFR level, at least at the lower levels. Thus, the British vocabulary size test writers and schools in Hungary and Greece appear to agree that learners at A1 and A2 level probably know less than 3000 of the most frequent words in English. Learners at B1 level will know about 3000, and learners at B2 level will know about 3500. Statistics can be misleading and we are dealing with small samples here. Nonetheless, the differences in the mean EFL vocabulary scores in Greece and Hungary at levels B1 and B2 were not statistically significant. At advanced level C1, on the other hand, the Greek and Hungarian vocabulary scores are significantly different ($t(15) = 3.092$, $p < .01$). Thus, while the Hungarian mean falls within the range suggested in Meara and Milton (2003), the Greek mean is well below it. The reason for this is considered later in the chapter.

At first sight the French results are more varied, but this is due to the influence of the British data. The Spanish and Greek data coincide closely at almost every CEFR level. The numbers are small but the differences between the means at every level in the Spanish and Greek data are not statistically significant. It is the British data which differ markedly from the other two and the mean vocabulary scores at every level of the CEFR are lower in the UK than elsewhere. It was pointed out at the outset of the chapter that foreign language exams, and the CEFR levels they have been placed at, have been criticized within the UK. Given the relationship between vocabulary size and overall language knowledge and skill, it would seem that the CEFR

levels in French have been interpreted very differently in Britain than elsewhere in Europe.

A characteristic of all the French data is that mean vocabulary scores for the CEFR levels are lower than for EFL. We have only one set of data for learners of Greek as a foreign language, but these figures are higher than those in either EFL or French. This, of course, raises the question of how the CEFR levels in the different languages are to be compared. Does the lower vocabulary score associated with French CEFR levels, for example, mean that the French levels are lower and much easier to achieve than the EFL ones, or do these differences reflect some systematic difference between the languages whereby fewer lexical resources are needed in French to achieve the kind of communicative skill that the CEFR levels describe?

Comparing vocabulary sizes across languages

There is evidence that there may be systematic differences between the vocabulary sizes required for the CEFR levels in different languages. For example, it may be possible to achieve certain levels of competence in a foreign language, such as reading with full comprehension, with fewer words in French than is possible in English. It is commonly accepted that full comprehension in a skill such as reading will require the reader to recognize almost all the words he or she encounters. A figure of 95 per cent of the words in a text for general comprehension (Laufer, 1989), or 98 per cent for reading for pleasure (Hu and Nation, 2000), are the kind of figures which are often quoted. Nation further suggests (2001, p. 147) that there is a threshold at about 80 per cent coverage which is required for gist understanding. But it is possible to achieve this kind of coverage with fewer words in French than in English. And it seems that rather more words in Greek are required to achieve this figure than in either French or English.

The reason for this is that languages are different in structure and the ways words are created and used. Some of the most frequent words in English are prepositions like *of* and *up*, for example, but other languages inflect much more than English and these prepositions are likely to be absent from the frequency lists in, say, Hungarian or Finnish and that will affect the coverage of the most frequent words in these languages. The most frequent preposition in French, *de*, is much more frequent than any equivalent in English. More relevant to English and French, is that English is a language where speakers can reputedly use a particularly large vocabulary and often appear to have a variety of words available for just a single idea (e.g. Bryson, 1990, p. 61). Part of the reason may be historical. English differentiates, for example, between many farmyard animals and the meat which comes from them, between *pork* and *pig* and between *sheep* and *mutton*. English too appears to have two different sets of vocabulary available for formal occasions, such as writing an essay, and for less formal

occasions, such as telephoning friends. An *argument* among friends might be called a *controversy* in academic circles. The train which, to you and me, *stops* at Paddington, *terminates* there in the language of the train conductor who is being formal. Regardless of whether English speakers really do have very large vocabularies to work with, it appears that not all languages make this formal and informal language distinction, or at least not in the choice of vocabulary, in the way English does. It seems that in French the most frequent vocabulary does the service both of everyday language and the specialist academic vocabulary which English requires. Thus, Cobb and Horst (2004) point to the coverage provided of academic texts by the most frequent 2000 words in French. The figure they quote of nearly 89 per cent (p. 30) would be equivalent to the General Service Word List of 2000 words in English plus Coxhead's Academic Word List of some 600 words. Arguably, 2000 words in French will do the work of some 2600, carefully selected rather than purely frequency-based, words in English. To help illustrate this we have plotted the coverage provided by frequency lists in English and in French as shown in Figure 12.1. It is apparent from this that the figure of 80 per cent coverage required for gist understanding requires 2000 words in English but substantially less in French, maybe only about 1500 words or fewer.

It may be possible to use the differences in coverage which frequency lists in different languages provide, to understand how the vocabulary size requirements of the CEFR levels might vary between these languages. If achievement of B2 level in EFL requires about 3000 words, which would provide about 85 per cent coverage of normal texts, then the volume of

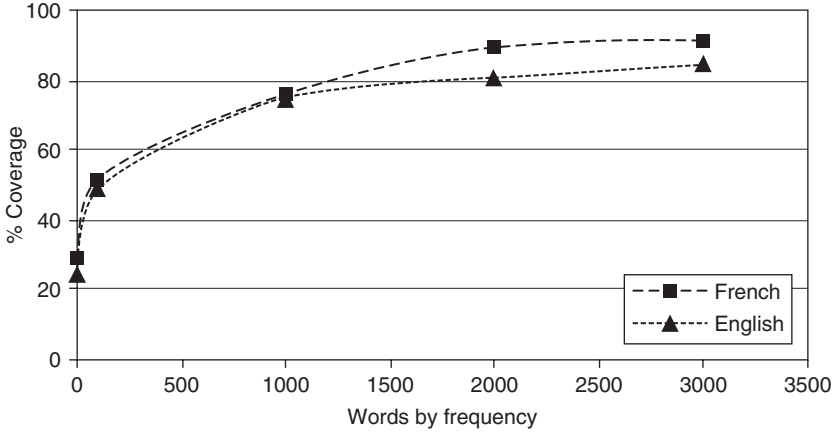


Figure 12.1 A comparison of coverage of text between Carroll et al.'s (1971) corpus of English and Baudot's (1992) French corpus

vocabulary in French producing the same coverage, perhaps 1800 words, might be expected for the same level. The information provided by the original CEFR word lists broadly supports this idea. While the original *Threshold* (B1) level word lists in English and French were both in the region of 2000 words, the figure for English is higher at about 2200 words (van Ek and Trim, 1990), and for French is lower at about 1800 words (Coste et al., 1987). It might be argued, therefore, that the vocabulary size figures for French are likely to be lower than their EFL equivalents at the CEFR levels above the most basic.

How might Greek as a foreign language compare? Does Greek provide figures which might also suggest a systematic difference? Figure 12.2 overlays the line for coverage from Carroll, Davies and Richman's (1971) corpus of English with the lemmatized Hellenic National Corpus's coverage (Hatzigeorgiu et al., 2001) and provides something like an equivalent list. At the outset the first few words are comparatively more frequent in Greek than in English; in Greek the definite article is very highly frequent even compared to English. Thereafter, Greek vocabulary provides proportionately less coverage and the two plot lines cross over (see Figure 12.2). The most frequent 5000 words in Greek provide about 83 per cent coverage which is substantially less than in English. A particular feature of Greek is the very high number of *hapax legomena* (words which occur only once in a corpus) which comprise 49.4 per cent of the corpus in Greek but is nearer to 30 per cent in English and French (Mikros, personal correspondence).

It appears from these data that rather more words are required in Greek than in English for any level of coverage beyond the smallest, and that more words would be needed in Greek to achieve the levels of communicative

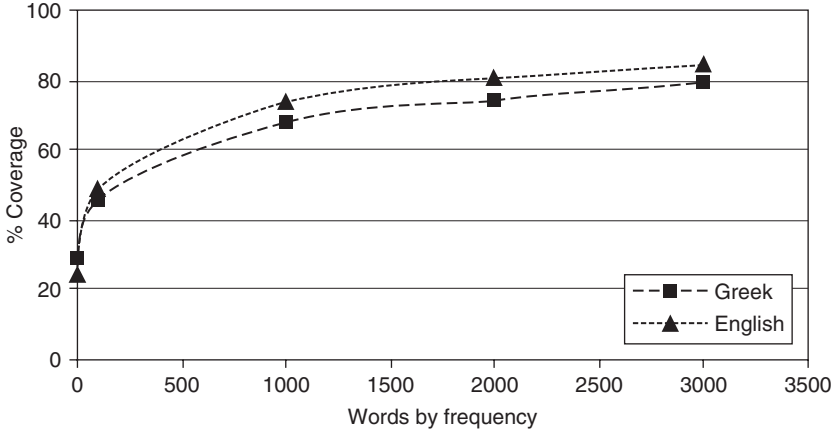


Figure 12.2 A comparison of text coverage between Carroll et al.'s (1971) corpus of English and the Hellenic National Corpus (Hatzigeorgiu et al., 2001)

Table 12.9 Summary of mean scores for each CEFR level in three foreign languages

CEFR level	*French in UK*	*French in Spain*	*French in Greece*	*EFL in Greece*	*EFL in Hungary*	*Greek in Greece*
A1		894.44	1125.71	1477.27		1492.10
A2		1700.00	1756.25	2156.81		2237.50
B1	952.04	2194.44	2422.72	3263.63	3135.90	3338.23
B2	1882.58	2450.00	2630.00	3304.54	3668.42	4012.50
C1		2675.00	3212.50	3690.90	4340.00	
C2	3326.47	3721.42	3525.00	4068.18		

ability that fewer words in English would allow. It might be expected, therefore, that levels of vocabulary knowledge required for the various CEFR grades would be higher in Greek than in either English or French. What happens when the vocabulary scores for the different languages at each CEFR level are compared with each other? Do the differences in vocabulary size which these coverage differences suggest, emerge in the CEFR framework? In Table 12.9 we have presented all the vocabulary size mean scores at each of the CEFR levels.

Encouragingly, these data support the differences which coverage figures suggest should occur. At every CEFR level the mean French vocabulary scores are smaller than the mean scores for EFL at the equivalent levels and at every level where we have data, the mean Greek vocabulary scores are higher than both the mean EFL and French scores.

Discussion and conclusions

At one level these results have produced exactly what was hoped for and expected. As learners get better in their foreign language, and become more skilled, able and communicative, they tend to know more words. The vocabulary size scores which emerge suggest that certain levels for vocabulary knowledge are associated with performance at each CEFR level. This supports the idea that the CEFR system can work in establishing equivalent levels in foreign languages across different countries and examinations systems. The EFL data in Greece and Hungary broadly conform well to the vocabulary levels suggested by the writers of the vocabulary testing software at each of the CEFR levels. The EFL system, at least in Greece, has the benefit of being tied strongly to the Cambridge testing system which itself has a vocabulary level attached to it in the form of Hindmarsh's (1980) list. This list of 4500 words and phrases should form the basis of test construction at the Cambridge FCE (B2) level and the use of this list has probably helped fix the standard of this exam over time. The mean vocabulary scores that learners produce at this level, approximately 3500 out of the 5000 most frequent words, fit well with the kind of vocabulary size implicit in Hindmarsh's list,

which includes not only the most frequent words required for communication but also words tied to the kind of thematic requirements of the FCE exam and which lie outside the most frequent 5000 words of English.

French does not have a fixed point of reference like the Hindmarsh list to help establish vocabulary norms at each CEFR level. Nonetheless, it is possible to argue that the French foreign language data from Greece and Spain suggest that the CEFR's skills-based criteria have allowed very similar levels of knowledge to be tied to the CEFR levels. This suggests that the system can be quite workable. The British data, however, reveal the weakness of the skills-based criteria when used in isolation from more objective evaluation methods. The British scores for learners of French are not just different from the Greek and Spanish scores, they are so different that the abilities of the learners in Britain cannot possibly be equivalent to learners at the equivalent CEFR levels in Greece or Spain. The presence of a vocabulary knowledge indicator will surely help the British system in evaluating where it stands in relation to the CEFR and in adjusting its level appropriately so that it will fit more convincingly within the framework. There appears, therefore, to be a real place for these vocabulary size measures.

The vocabulary score hierarchies which have emerged from this exercise appear to be different between languages. The EFL scores are higher than the French foreign language scores, and the Greek vocabulary scores are higher than both. There is no reason for thinking that the achievement of a level of competence in the CEFR system should require a single vocabulary size in all languages. Languages differ and it is quite likely that it is possible to be rather more communicative and fluent with fewer vocabulary resources in some languages than in others. It appears possible to argue, however, that this kind of variation is linked to coverage which also varies from language to language. The volume of data represented here is small but it suggests that the CEFR levels are associated with levels of coverage of text, and that these coverage figures will allow us to estimate vocabulary size equivalences between languages. To progress from elementary, A1 and A2 levels, for example, it seems that learners need to know a volume of vocabulary which will give more than 80 per cent coverage. In EFL that would require knowledge of over 2000 words and in French rather fewer. This in turn means that vocabulary size guidelines can be produced across the languages to which the CEFR is applied, tying it together in a way this is not possible at the moment. At the moment we assume that the French learners at B2 level in Greece and Spain, for example, are similar in performance and knowledge to EFL learners at B2 level in Greece and Hungary, but we have no real way of demonstrating this without reference to something like vocabulary levels. It is early days, but this method of rationalizing how vocabulary size scores in difference foreign languages might link to the CEFR looks promising and would merit more systematic investigation with larger numbers of learners, in more countries and learning more foreign languages.

It should not be thought, however, that a set level of vocabulary is a requirement of achieving CEFR levels; for example, that EFL learners must have 3500 words before they can achieve CEFR B2 level. The relationship between vocabulary size and level of attainment need not be fixed in this way. Vocabulary size scores are likely to be indicators of knowledge and attainment rather than an absolute determiner of these things. It has been commented on several times in this chapter that while the mean scores for each CEFR level appear to vary predictably and to tell a comprehensible story about how vocabulary knowledge and attainment develop, there is nevertheless considerable variation in vocabulary scores at every level of the CEFR. There are several reasons why the relationship between vocabulary size and skill in communication of language performance need not be fixed.

One reason for the slightly messy individual data which this investigation has produced is the imperfect way learners are assigned to their classes in foreign languages, and the idiosyncratic way they may progress. It has been assumed that learners have been assigned correctly to classes and that every individual in a B1 level class, for example, is really at B1 level. In reality there is no guarantee that this is the case. Learners can be assigned to a class for many reasons other than level of knowledge and performance. They can be grouped with other learners of the same age, for example, or to keep a group of friends together. Again, learners may have been assigned to the closest practical level even if it is not the correct one. Where a school contains bilingual learners or the children of native speakers of the foreign language alongside beginners, for example, it may not be practical or financially possible to arrange classes across the entire range of language ability and for every year in a school. And again, once the class has begun learners can progress at very different rates according to their interest and motivation. Even where a class begins a year's study at the same level of ability, some learners will always make better progress than others. It seems inevitable in this kind of research, therefore, that learners of different levels will be grouped in a way that makes the results less clear.

An additional factor which is likely to obscure the relationship between vocabulary size and CEFR level, is that language testing is not direct or precise. In language testing we are dependent for valid results not only on the creation of good tests to reveal aspects of language proficiency, but also on the ability of the learner to play along with the system and willingly and correctly show what they know. This is not always easy. Learners may not be interested in the test, or they may become bored, tired or ill and misrepresent their knowledge. Equally, they may choose, particularly in objective-style testing, to make educated guesses about their answers in order to gain the highest possible score rather than the score which most accurately reflects their knowledge. Some variation, it seems, is just an inevitable consequence of the language testing system.

These factors should not disguise a third reason why vocabulary size and ability level are not precisely tied and this reason is rather more important in this context. This is that the relationship between foreign language knowledge and the ability or skill in using that knowledge may vary from one individual to another. Communicative ability in reading and listening, for example, rests to a degree on anticipating what is likely to come next and in making intelligent guesses as to the meaning of the writer or speaker. Some learners can use limited data and achieve comprehension more easily than others who require more complete knowledge to draw the same conclusions. Likewise, some foreign language users manage to be much more creative and intuitive than others in their ability to use the limited language knowledge they have for communication. We have very little understanding of this type of variation and have no real way of characterizing it usefully. As a consequence we currently find it hard to explain away completely satisfactorily the range of vocabulary scores that learners in the same class, or at the same level, can produce, and this is an area that bears further investigation.

For these reasons the kind of vocabulary data which is likely to emerge and be most useful for the CEFR system will be ranges of vocabulary knowledge associated with the CEFR levels, and which will act as guidelines. Groups of learners might be expected to conform to these guidelines quite well since the progress of vocabulary knowledge among *groups* of learners is now becoming quite well understood. Individuals are likely to be less predictable, however, and while it is unlikely that learners will depart enormously from the guidelines, some individuals are likely to fall outside any vocabulary range that is set.

In conclusion, therefore, it seems that it is quite workable to put vocabulary knowledge measures back into the CEFR. While specifying lists of required vocabulary may no longer be appropriate, a vocabulary size metric can offer much to the framework. The vocabulary size scores which emerge among learners at different levels of the framework are relatively predictable and understandable, and it appears that vocabulary size estimates are already associated with each of the CEFR levels, even if users of the system are not aware of this. This chapter has been able to codify what some of these levels are. We have even suggested a way of handling and explaining the way vocabularies will vary between languages so the CEFR system can remain generalizable across all languages and countries. This process has already revealed the kind of discrepancy to which a system without an objective style of measurement is prone, and it has highlighted the way the British placements of foreign language qualifications appear very different from the kind of expectations which are common on the rest of the continent of Europe. By reintroducing a vocabulary size measure to the CEFR the system can, very likely, be made more robust so that misplacements of this kind can be recognized and corrected.

References

Adolphs, S., and Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, *24*, 425–38.

Afitskaya, N. (2002). Assessing vocabulary richness: LS* and LS**. *University of the West of England, Faculty of Languages and European Studies Research and Consultancy Bulletin*, *22*, 4–9.

Aitchison, J. (2003). *Words in the mind: an introduction to the mental lexicon*. Oxford: Blackwell.

Akhtar, P. (1985). *Amar's last wish* (Gujarati translation by Pragna Thakar). London: Magi.

Alderson, C. (2002). Testing proficiency and achievement: principles and practice. In J. A. Coleman, R. Grotjahn and U. Raatz (eds), *University language teaching and the C-test* (pp. 15–30). Bochum: AKS Verlag.

Alladina, S., and Edwards, V. (1991). *Multilingualism in the British Isles* (Vols 1 and 2). Harlow: Longman.

Anderson, R. C., and Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (ed.), *Comprehension and teaching: research reviews* (pp. 77–117). Newark, DE: International Reading Association.

Anderson, R. C., and Nagy, W. E. (1992). The vocabulary conundrum. *American Educator: the Professional Journal of the American Federation of Teachers*, *16*, 44–7.

Anglin, J. M. (1993). Vocabulary development: a morphological analysis. *Monographs of the Society for Research in Child Development, 165*(10, Serial No. 238).

Arnaud, P. J. L. (1984). The lexical richness of L2 written productions and the validity of vocabulary tests. In T. Culhane, C. Klein Bradley and D. K. Stevenson (eds), *Practice and problems in language testing: papers from the International Symposium on Language Testing* (pp. 14–28). Colchester: University of Essex.

Arnett, K. (2001). The accommodation of grade 9 students with learning disabilities in the applied core French classroom. Unpublished master's thesis, The Ontario Institute for Studies in Education, University of Toronto, Toronto, Canada.

Arnfast, J. S., and Jørgensen, J. N. (2003). Code-switching as a communication, learning, and social negotiation strategy in first-year learners of Danish. *International Journal of Applied Linguistics, 13*, 23–53.

Arras, U., Eckes, T., and Grotjahn, R. (2002). C-Tests im Rahmen des 'Tests Deutsch als Fremdsprache' (TestDaF). Erste Forschungsergebnisse. In R. Grotjahn (ed.), *Der C-Test. Theoretische Grundlangen und praktische Anwendungen* (Vol. 4) (pp. 175–210). Bochum: AKS Verlag.

Association for Language Awareness (ALA). (n.d.). Language awareness defined. Retrieved 18 October 2008, from: http://www.lexically.net/ala/la_defined.htm

Auerbach, E. (1993). Re-examining English only in the ESL classroom. *TESOL Quarterly, 27*, 9–32.

Ayoun, D. (ed.). (2007). *French applied linguistics*. Amsterdam: John Benjamins.

Barnes, S., Gutfreund, M., Satterly, D., and Wells, G. (1983). Characteristics of adult speech which predict children's language development. *Journal of Child Language, 10*, 65–84.

Bassetti, B. (2005). Effects of writing systems on second language awareness: word awareness in English learners of Chinese as a foreign language. In V. Cook and

B. Bassetti (eds), *Second language writing systems* (pp. 335–56). Clevedon: Multilingual Matters.
Bates, E., and Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes, 12*, 507–84.
Bates, E., and Goodman, J. C. (1999). On the emergence of grammar from the lexicon. In B. MacWhinney (ed.), *The emergence of language* (pp. 29–70). Mahwah, NJ: Erlbaum.
Baudot, J. (1992). *Fréquences d'utilisation des mots en français écrit contemporain*. Montreal: Les Presses de l'Université de Montréal.
Bayliss, D., and Raymond, P. (2004). The link between academic success and L2 proficiency in the context of two professional programs. *The Canadian Modern Language Review, 61*, 29–51.
Beck, I. L., McKeown, M. G., Hamilton, R., and Kucan, L. (1997). *Questioning the author*. Newark, DE: International Reading Association.
Beck, I. L., McKeown, M. G., and Kucan, L. (2002). *Bringing words to life*. New York: Guilford Press.
Beck, I. L., McKeown, M. G., Sandora, C., Kucan, L., and Worthy, J. (1996). Questioning the author: a yearlong classroom implementation to engage students with text. *The Elementary School Journal, 96*, 385–414.
Beck, I. L., Perfetti, C. A., and McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology, 74*, 506–21.
Becker, J. D. (1975). *The phrasal lexicon* (Advanced Research Projects Report No. 3081). Washington, DC: Agency of the Department of Defense.
Bell, H. (2003). Using frequency lists to assess L2 texts. Unpublished doctoral dissertation, University of Swansea, Wales.
Benedict, H. (1979). Early lexical development: comprehension and production. *Journal of Child Language, 6*, 183–200.
Bensoussan, M. (1983). Dictionaries and tests of EFL reading comprehension. *English Language Teaching Journal, 37*, 341–5.
Bereiter, C., Engelmann, S., Osborn, J., and Reidford, P. (1966). *Preschool education today*. New York: Doubleday.
Berman, R. A., and Slobin, D. I. (1994). *Relating events in narrative: a crosslinguistic developmental study*. Hillsdale, NJ: Erlbaum.
Bernstein, B. (1973). *Class, codes and control* (Vol. 1). London: Routledge & Kegan Paul.
Bialystok, E. (2001). *Bilingualism in development*: *language, literacy, and cognition*. Cambridge: CUP.
Bialystok, E., Craik, F. I. M., and Luk, G. (2008). Lexical access in bilinguals: effects of vocabulary size and executive control. *Journal of Neurolinguistics, 21*, 522–38.
Biber, D. (1988). *Variation across speech and writing*. Cambridge: CUP.
Biber, D. (1995). *Dimensions of register variation*. Cambridge: CUP.
Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (eds). (1999). *Longman grammar of written and spoken English*. London: Longman.
Biemiller, A. (1999). *Language and reading success*. Cambridge, MA: Brookline Books.
Biemiller, A. (2006a). An effective method for building vocabulary in primary grades. *Journal of Educational Psychology*, *98*, 44–62.
Biemiller, A. (2006b). Vocabulary development and instruction: a prerequisite for school learning. In D. K. Dickinson and S. B. Neuman (eds), *Handbook of early literacy research* (Vol. 2) (pp. 41–51). New York: Guilford Press.

Biemiller, A. (2007). The influence of vocabulary on reading acquisition. *Encyclopedia of language and literacy development*. London: Canadian Language and Literacy Research Network.

Biemiller, A., and Boote, C. (2003). Vocabulary: needed if more children are to read well. *Reading Psychology, 24*, 323.

Biemiller, A., and Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology, 93*, 498–520.

Bishop, D. V. M., and Edmundson, A. (1987). Language-impaired 4-year-olds: distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders, 52*, 156–73.

Blackledge, A. (2001). Complex positionings: women negotiating identity and power in a minority urban setting. In A. Pavlenko, A. Blackledge, I. Piller and M. Teutsch-Dwyer (eds), *Multilingualism, second language learning, and gender* (pp. 53–75). Berlin: Mouton de Gruyter.

Blank, M., and White, S. J. (1986). Questions: a powerful but misused form of classroom exchange. *Topics in Language Disorders*, *6*, 1–12.

Bornstein, M. H. (1998). Vocabulary competence in early childhood: measurement, latent construct, and predictive validity. *Child Development, 69*, 654–71.

Bornstein, M. H., Haynes, M. O., and Painter, K. M. (1998). Sources of child vocabulary competence: a multivariate model. *Journal of Child Language, 25*, 367–93.

Bourdieu, P. (1991). *Language and symbolic power*. Cambridge, MA: Harvard University Press.

Bourdieu, P. (1997). The forms of capital. In A. H. Halsey, H. Lauder, P. Brown and A. S. Wells (eds), *Education, culture, economy, and society* (pp. 46–58). Oxford: OUP.

Brooks, G., and Adams, M. (2002). *Spoken English proficiency and academic performance: is there a relationship and if so, how do we teach*? Centre for Professional Development, Macquarie University. Retrieved 5 August 2006, from http://www.cfl.mq.edu.au/celebrate/pdf/papers/brooks1.pdf

Bryson, B. (1990). *Mother tongue*. London: Penguin.

Bucks, R. S., Singh, S., Cuerden, J. M., and Wilcock, G. K. (2000). Analysis of spontaneous, conversational speech in dementia of Alzheimer type: evaluation of an objective technique for analysing lexical performance. *Aphasiology*, *14*, 71–91.

Bullock, A. (1975). *A language for life*. London: HMSO.

Burgess, S., McConnell, B., Propper C., and Wilson, D. (2003). Girls rock, boys roll: an analysis of the age 14–16 gender gap in English schools. CMPO Working Paper (Series No. 03/084). Bristol: University of Bristol, CMPO.

Butzkamm, W. (1998). Code-switching in a bilingual history lesson: the mother tongue as a conversational lubricant. *International Journal of Bilingual Education and Bilingualism*, *1*, 81–99.

Cain, K. (2007). Deriving word meanings from context: does explanation facilitate contextual analysis? *Journal of Research in Reading*, *30*, 347–59.

Cain, K., Oakhill, J., and Bryant, P. (2004). Children's reading comprehension ability: concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96*, 31–42.

Campbell, J. R., Hombo, C. M., and Mazzeo, J. (eds). (2000). *NAEP 1999: trends in academic progress: three decades of student performance 2000*. Washington, DC: US Department of Education, Office of Educational Research and Improvement.

Canagarajah, A. S. (1995). Functions of codeswitching in ESL classrooms: socializing bilingualism in Jaffna. *Journal of Multilingual and Multicultural Development*, *6*, 173–95.

Canale, M., and Swain M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*, 1–47.
Carroll, J. B., Davies, P., and Richman, B. (1971). *The American heritage word frequency book*. Boston: Houghton Mifflin.
Carter, R. (2003). Grammar, spoken English and literacy. *Literacy Today*, *34*. Retrieved 24 October 2008 from http://www.literacytrust.org.uk/Pubs/carter.html
Carter, R., and McCarthy, M. (1995). Grammar and the spoken language. *Applied Linguistics*, *16*, 141–58.
Cazden, C. B. (2001). *Classroom discourse: the language of teaching and learning* (2nd edn). Westport, CT: Heinemann.
Celik, M. (2003). Teaching vocabulary through code-mixing. *English Language Teaching Journal*, *57*, 361–9.
Chafe, W., and Danielewicz, J. (1987). Properties of spoken and written language. In R. Horowitz and S. J. Samuels (eds), *Comprehending oral and written language* (pp. 83–113). New York: Academic Press.
Chall, J., and Dale, E. (1995). *Readability revisited and the new Dale–Chall readability formula*. Cambridge, MA: Brookline Books.
Chambers, F. (1991). Promoting use of the target language in the classroom. *Language Learning Journal*, *4*, 27–31.
Chamot, A., and O'Malley, M. (1994). *The CALLA handbook: implementing the cognitive academic language learning approach*. Reading, MA: Addison-Wesley.
Chan, B. (2004). Beyond contextualization: codeswitching as a 'textualization cue'. *Journal of Language and Social Psychology, 23*, 7–27.
Chen, H. (2007). Investigating the effects of different types of Chinese translation on advanced learners' lexical processing of low frequency English words. Unpublished master's thesis, University of Oxford, UK.
Chesterfield, R. A., and Chesterfield, K. B. (1985). Natural order in children's use of second language learning strategies. *Applied Linguistics, 6*, 45–59.
Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
Clark, C., and Rumbold, K. (2006). *Reading for pleasure: a research overview*. London: National Literacy Trust. Retrieved 24 October 2008, from http://www.literacytrust.org.uk/Research/readpleasure.html
Clark, E. V. (1985). The acquisition of romance with special reference to French. In D. Slobin (ed.), *The cross-linguistic study of language acquisition* (pp. 688–782). Hillsdale, NJ: Erlbaum.
Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge: CUP.
Clark, E. V. (2003). *First language acquisition*. Cambridge: CUP.
Clear, J., Fox, G., Francis, G., Krishnamurthy, R., and Moon, R. (1996). COBUILD: the state of the art. *International Journal of Corpus Linguistics, 1*, 303–14.
Cobb, T., and Horst, M. (2004). Is there room for an AWL in French? In B. Laufer and P. Bogaards (eds), *Vocabulary in a second language: selection, acquisition, and testing* (pp. 15–38). Amsterdam: John Benjamins.
Cohen, A., Segal, M., and Weiss, R. (1985). The C-test in Hebrew. *Fremdsprachen und Hochschule*. *AKS Rundbrief*, *13/14*, 121–7.
Cole, A. (1986). *Soma goes to market*. New Malden, UK: Suhada Press.
Coleman, J. A. (1994). Profiling the advanced language learner: the C-test in British further and higher education. In R. Grotjahn (ed.), *Der C-test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 2) (pp. 217–37). Bochum: Brockmeyer.

Coley, J. D. (1993). From college to classroom: three teachers' accounts of their adaptations of reciprocal teaching. *Elementary School Journal, 94*, 255–66.
Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition*, *26*, 227–49.
Collins, J., and Michaels, S. (1986). Speaking and writing: discourse strategies and the acquisition of literacy. In J. Cook-Gumperz (ed.), *The social construction of literacy*. Cambridge: CUP.
Collins, M. F. (2005). ESL preschoolers' English vocabulary acquisition from storybook reading. *Reading Research Quarterly, 40*, 406–8.
Connor, C. M., Morrison, F. J., and Petrella, J. N. (2004). Effective reading comprehension instruction: examining child × instruction interactions. *Journal of Educational Psychology, 96*, 682–98.
Connor, C. M., Morrison, F. J., and Underwood, P. S. (2007). A second chance in second grade: the independent and cumulative impact of first- and second-grade reading instruction and students' letter-word reading skill growth. *Scientific Studies of Reading, 11*, 199–233.
Cook, V. J. (1991). The poverty-of-the-stimulus argument and multicompetence. *Second Language Research, 7*, 103–17.
Cook, V. J. (1997a). Monolingual bias in second language acquisition research. *Revista Canaria de Estudios Ingleses*, *34*, 35–50 (also available from: http://homepage.ntlworld.com/vivian.c/Writings/Papers/MonolingualBias.htm)
Cook, V. J. (1997b). *Inside language*. London: Arnold.
Cook, V. J. (2001). *Second language learning and teaching* (3rd edn). London: Arnold.
Corson, D. (1984). The lexical bar: lexical change from 12 to 15 years measured by social class, region and ethnicity. *British Educational Research Journal, 10*, 115–33.
Corson, D. (1985). *The lexical bar*. Oxford: Pergamon.
Corson, D. (1995). *Using English words*. Dordrecht: Kluwer.
Coste, D., Courtillon, J., Ferenczi, V., Martins-Baltar, M., and Papo, E. (1987). *Un niveau seuil*. Paris: Editions Didier.
Council of Europe. (2003). *Common Framework of Reference for Languages*. Cambridge: CUP.
Cowie, A.P. (1999). *English dictionaries for foreign learners: a history*. Oxford: Clarendon.
Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*, 213–38.
Craik, F. I. M., and Bialystok, E. (2006). Planning and task managements in older adults: cooking breakfast. *Memory and Cognition, 34*, 1236–49.
Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism, 19*, 121–9.
Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In California State Department of Education (ed.), *Schooling and language minority students: a theoretical framework*. Los Angeles, CA: California State University, Evaluation, Dissemination and Assessment Center.
Cummins, J. (1998). E-lective language learning: design of a computer-assisted, text-based, ESL/EFL learning system. *TESOL Journal*, *7*, 18–21. Retrieved 24 October 2008, from http://www.iteachilearn.com/cummins/elective.html
Cummins, J. (2003). *BICS and CALP*. Retrieved 24 October 2008, from http://www.iteachilearn.com/cummins/bicscalp.html
Cunningham, A. E., and Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*, 934–45.

Daller, H. (1999). *Migration und Mehrsprachigkeit*. Frankfurt: Peter Lang.
Daller, H., and Phelan, D. (2007). What is in a teacher's mind? The relation between teacher ratings of EFL essays and different aspects of lexical richness. In H. Daller, J. Milton and J. Treffers-Daller (eds), *Modelling and assessing vocabulary knowledge* (pp. 234–45). Cambridge: CUP.
Daller, H., and Xue, H. (2007). Lexical richness and the oral proficiency of Chinese EFL students. In H. Daller, J. Milton and J. Treffers-Daller (eds). *Modelling and assessing vocabulary knowledge* (pp. 150–64). Cambridge: CUP.
Daller, H., Milton J., and Treffers-Daller, J. (eds). (2007). *Testing and modelling lexical knowledge*. Cambridge: CUP.
Daller, H., Van Hout, R., and Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics, 24*, 197–222.
David, A. (2007). Vocabulary and morphosyntactic complexity in French second language learners. Paper presented at the Symposium on Complexity, Accuracy and Fluency in Second Language Use, Learning and Teaching, Brussels, March.
David, A. (2008). A developmental perspective on productive lexical knowledge in L2 oral interlanguage. *Journal of French Language Studies*, *18*, 315–31.
David, A. (in press). Vocabulary breadth in French L2 learners. *Language Learning Journal*.
Davis, J. N. (1989). Facilitating effects of marginal glosses on FL reading. *The Modern Language Journal*, *73*, 41–8.
De Bot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics*, *13*, 1–24.
DeKeyser, R. (1991). Foreign language development during a semester abroad. In B. F. Freed (ed.), *Foreign language acquisition research and the classroom* (pp. 104–19). Lexington, MA: D. C. Heath.
Department for Education and Employment, and Qualifications and Curriculum Authority. (1999). *English: the National Curriculum for England*. London: QCA.
Department for Education and Skills. (1998). *The National Literacy Strategy*. Retrieved 24 October 2008, from http://www.standards.dfes.gov.uk/primary/publications/literacy/nls_framework/
Department of Education and Science. (1990). *Modern foreign languages for ages 11 to 16*. London: Her Majesty's Stationery Office.
Dewaele, J.-M. (2000). Saisir l'insaisissable? Les mesures de longueur d'énoncés en linguistique appliquée. *International Review of Applied Linguistics, 38*, 17–33.
Dewaele, J.-M. (ed.). (2005). *Focus on French as a foreign language: multidisciplinary approach*. Clevedon: Multilingual Matters.
Dickinson, D. K. (1987). Oral language, literacy skills and response to literature. In J. Squire (ed.), *The dynamics of language learning: research in the language arts* (pp. 147–83). Urbana, IL: National Council of Teachers in English.
Dickinson, D. K. (1994). Features of early childhood classrooms that support development of language and literacy. In J. Duchan, L. Hewitt and R. Sonnenmeier (eds), *Pragmatics: from theory to practice* (pp. 185–201). Englewood Cliffs, NJ: Prentice Hall.
Dickinson, D. K. (2001a). Large-group and free-play times: conversational settings supporting language and literacy development. In D. K. Dickinson and P. O. Tabors (eds), *Beginning literacy with language: young children learning at home and school* (pp. 223–55). Baltimore, MD: Brookes Publishing.
Dickinson, D. K. (2001b). Putting the pieces together: the impact of preschool on children's language and literacy development in kindergarten. In D. K. Dickinson and P. O. Tabors (eds), *Beginning literacy with language: young children learning at home and school* (pp. 257–87). Baltimore, MD: Brookes Publishing.

Dickinson, D. K., and Beals, D. B. (1994). Not by print alone: oral language supports for early literacy development. In D. Lancy (ed.), *Children's emergent literacy: from research to practice* (pp. 29–40). Westport, CT: Praeger.

Dickinson, D. K., and Porche, M. (under review). The relationship between teacher–child conversations with low-income four-year olds and grade four language and literacy development.

Dickinson, D. K., and Sprague, K. E. (2002). The nature and impact of early childhood care environments on the language and early literacy development of children from low-income families. In S. B. Neuman and D. K. Dickinson (eds), *Handbook of early literacy research* (pp. 263–80). New York: Guilford Press.

Dickinson, D. K., and Tabors, P. O. (eds). (2001). *Beginning literacy with language: young children learning at home and school*. Baltimore, MD: Brookes Publishing.

Dickinson, D. K., Darrow, C. L., and Tinubu, T. A. (2008). Conversations in preschool: toward cognitively enriching and responsive discourse. *Early Education and Development, 19*, 396–429.

Dickinson, D. K., McCabe, A., Anastasopoulos, L., Peisner-Feinberg, E., and Poe, M. D. (2003). The comprehensive language approach to early literacy: the interrelationships among vocabulary, phonological sensitivity, and print knowledge among preschool-aged children. *Journal of Educational Psychology, 95*, 465–81.

Dickinson, D. K., St. Pierre, R., and Pettengill, J. (2004). High quality classrooms: a key ingredient to family literacy programs. In B. Wasik (ed.), *Handbook of family literacy* (pp. 137–54). Mahwah, NJ: Erlbaum.

Dixon, J. A., and Marchman, V. A. (2007). Grammar and the lexicon: developmental ordering in language acquisition. *Child Development, 78*, 190–212.

Dooey, P. (1999). An investigation into the predictive validity of the IELTS test as an indicator of future academic success. In K. Martin, N. Stanley and N. Davidson (eds), *Teaching in the disciplines/learning in context. Proceedings of the Annual Teaching Learning Forum* (pp. 114–18). Perth: University of Western Australia. Retrieved 8 December 2006, from http://lsn.curtin.edu.au/tlf/tlf1999/dooey.html

Dudley-Marling, C. (2007). Return of the deficit. *Journal of Educational Controversy, 2*. Retrieved 24 October 2008, from http://www.wce.wwu.edu/Resources/CEP/eJournal/v002n001/Index.shtml

Duff, P. A., and Polio, C. G. (1990). How much foreign language is there in the foreign language classroom? *Modern Language Journal*, *74*, 154–66.

Dunn, L. M., and Dunn, D. M. (1959/2006). *Peabody picture vocabulary scale*. Bloomington, MN: Pearson Assessments.

Dupuy, B., Tse, L., and Cook, T. (1996). Bringing books into the classroom: first steps in turning college-level ESL students into readers. *TESOL Journal, 5*, 10–15.

Durán, P., Malvern, D., Richards, B., and Chipere, N. (2004). Developmental trends in lexical diversity. *Applied Linguistics, 25*, 220–42.

Eckes, T., and Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing, 23*, 290–325.

Edwards, J. (1979). *Language and disadvantage*. London: Edward Arnold.

Edwards, V. (2004). *Multilingualism in the English-speaking world: pedigree of nations*. Oxford: Blackwell.

Edwards, V. (2008). *The culture of reading: an evaluation of a key programme of PRAESA*. Cape Town: Project for the Study of Alternative Education in South Africa.

Edwards, V., Monaghan, F., and Knight, J. (2000). Books, pictures and conversations: using bilingual multimedia storybooks to develop language awareness. *Language Awareness, 9*, 135–46.

Elder, B. N. (2005). Early predictors of emergent and conventional literacy skills: a longitudinal investigation. Unpublished doctoral dissertation, University of California, Davis, United States.

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly, 24*, 174–87.

Elley, W. B. (1991). Acquiring literacy in a second language: the effect of book-based programs. *Language Learning*, *41*, 375–411.

Elley, W. B., and Mangubhai, F. (1983). The impact of reading on second language learning. *Reading Research Quarterly, 19*, 53–67.

Ellis, N. (1996). Sequencing in SLA: phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, *18*, 91–126.

Ellis, R. (1993). Second language learning and the structural syllabus. *TESOL Quarterly, 27*, 91–113.

Ellis, R. (1994). A theory of instructed second language acquisition. In N. Ellis (ed.), *Implicit and explicit learning of languages* (pp. 79–114). London: Academic Press.

Ellis, R. (1997). *The study of second language acquisition*. Oxford: OUP.

Erman, B., and Warren, B. (2000). The idiom principle and the open choice principle. *Text*, *20*, 29–62.

Eversley, J., and Baker, P. (eds). (2000). *Multilingual capital*. London: Battlebridge.

Feast, V. (2002). The impact of IELTS scores on the performance at university. *International Educational Journal*, *3*, 70–85.

Feitelson, D., Goldstein, Z., Iraqi, J., and Share, D. L. (1993). Effects of listening to story reading on aspects of literacy acquisition in a diglossic situation. *Reading Research Quarterly, 28*, 71–9.

Feuverger, G. (1994). A multicultural literacy intervention for minority language students. *Language and Education*, *8*, 123–46.

Field, A. (2005). *Discovering statistics using SPSS*. London: Sage.

Fielding, L. G., and Pearson, P. D. (1994). Reading comprehension: what works. *Educational Leadership*, *51*, 62–8.

Fishman, J. (1971). *Advances in the sociology of language*. The Hague: Mouton.

Fitzgerald, J. (1995). English-as-a-second language learners' cognitive reading processes: a review of research in the United States. *Review of Educational Research, 65*, 145–90.

Fitzgerald, J., and Noblit, G. W. (1999). About hopes, aspirations, and uncertainty: first-grade English language learners' emergent reading. *Journal of Literacy Research, 31*, 133–82.

Foster, P. (2001a). Rules and routines: a consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan and M. Swain (eds), *Researching pedagogic tasks: second language learning, teaching, and testing* (pp. 75–93). Harlow: Longman.

Foster, P. (2001b). Lexical measures in task-based performance. Paper presented at the American Association for Applied Linguistics Conference, St. Louis, MO, February.

Foster, P., and Skehan, P. (1996). The influence of planning on performance in task-based learning. *Studies in Second Language Acquisition, 18*, 299–324.

Foster, P., and Skehan, P. (1999). The effect of source of planning and focus on planning on task-based performance. *Language Teaching Research*, *3*, 185–215.

Foster, P., and Skehan, P. (under review). The effects of post-task activities on the accuracy and complexity of language during task performance.

Foster, P., and Tavakoli, P. (in press). Native speakers and task performance: comparing effects on complexity, fluency and lexical diversity. *Language Learning*.

Foster, P., Tonkyn, A., and Wigglesworth, G. (2000). Measuring spoken language: a unit for all reasons. *Applied Linguistics*, *21*, 354–75.

Fotos, S., and Ellis, R. (1991). Communicating about grammar: a task-based approach. *TESOL Quarterly*, *25*, 605–28.

Franklin, C. E. M. (1990). Teaching in the target language: problems and prospects. *Language Learning Journal*, *2*, 20–4.

Fröhlich, M., Spada, N., and Allen, P. (1985). Differences in the communicative orientation of L2 classrooms. *TESOL Quarterly*, *19*, 27–58.

Gallas, K. (1992). When the children take the chair: a study of sharing time in a primary classroom. *Language Arts, 69*, 172.

Garcia, G. E. (1991). Factors influencing the English reading test performance of Spanish-speaking Hispanic children. *Reading Research Quarterly, 26*, 371–92.

Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: a critical survey. *Applied Linguistics*, *28*, 2412–65.

Gendner, V., and Adda–Decker, M. (2002). *Analyse comparative de corpus oraux et écrits français: mots, lemmes et classes morpho-syntaxiques*. XXIVièmes Journées d'Étude sur la Parole, Nancy, 24–27 June 2002. Retrieved 24 October 2008, from ftp://tlp.limsi.fr/public/jep02POS.pdf

Genishi, C., Stires, S. E., and Yung–Chan, D. (2001). Writing in an integrated curriculum: prekindergarten English language learners as symbol makers. *Elementary School Journal, 101*, 399–416.

Gleitman, L. R., and Gleitman, H. (1992). A picture is worth a thousand words, but that's the problem: the role of syntax in vocabulary acquisition. *Current Directions in Psychological Science, 1*, 31–5.

Goh, J. (2007). *China education market briefing*. Retrieved 24 October 2008 from http://cemis.britishcouncil.org/download/banner.asp

Goldenberg, C. (1992). Instructional conversations: promoting comprehension through discussion. *Reading Teacher, 46*, 316–26.

Gougenheim, G., Rivenc, P., Michéa, R., and Sauvageot, A. (1964). *L'élaboration du français fondamental (1er degré): étude sur l'établissement d'un vocabulaire et d'une grammaire de base* (2nd edn). Paris: Didier.

Graham, J. G. (1987). English language proficiency and the prediction of academic success. *TESOL Quarterly*, *21*, 505–21.

Graham, S., Richards, B. J., and Malvern, D. D. (2008). Progress in learning French vocabulary in a one-year advanced course at school. *Journal of French Language Studies*, *18*, 349–64.

Granfeldt, J. (2005). The development of gender attribution and gender agreement in French: a comparison of bilingual first and second language learners. In J.-M. Dewaele (ed.), *Focus on French as a foreign language: multidisciplinary approaches* (pp. 164–90). Clevedon: Multilingual Matters.

Grosjean, F. (1997). The bilingual individual. *Interpreting*, *2*, 163–87.

Grosjean, F. (1998). Studying bilinguals: methodological and conceptual issues. *Bilingualism: Language and Cognition*, *1*, 131–49.

Grotjahn, R. (1992). Der C-test. Einleitende Bemerkungen. In R. Grotjahn (ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen*. Bochum: Brockmeyer.

Grotjahn, R. (1995). Der C-Test: state of the art. *Zeitschrift für Fremdsprachen-forschung*, *6*, 37–60.

Grotjahn, R. (2007). *Die C-Test Bibliographie*. Retrieved 5 August 2008 from http://homepage.ruhr-uni-bochum.de/Ruediger.Grotjahn

Grotjahn, R., and Stemmer, B. (1985). On the development and evaluation of a C-test in French. *Fremdsprachen und Hochschule. AKS Rundbrief, 13/14*, 101–20.

Guillot, M.-N. (2005). Il y a des gens qui disent que … 'there are people who say that …' Beyond grammatical accuracy in FL learners' writing: issues of non-nativeness. *International Review of Applied Linguistics in Language Teaching, 43*, 109–28.

Guiraud, P. (1954). *Les caractères statistiques du vocabulaire*. Paris: Presses Universitaires de France.

Guiraud, P. (1960). *Problèmes et méthodes de la statistique linguistique*. Dordrecht: D. Reidel.

Guo, T. (2007). A case study of teachers' codeswitching behaviours in mainland China's university EFL classrooms and students' reactions to the codeswitching. Unpublished doctoral dissertation, University of Oxford, UK.

Häggsblom, C. (2006). *Young EFL pupils reading multicultural fiction: an ethnographic case study of a Swedish language primary school in Finland*. Turku: Åbo Akademi.

Hakuta, K. (1986). *Mirror of language. The debate on bilingualism*. New York: Basic Books.

Halliday, M. A. K. (1987). Spoken and written modes of meaning. In R. Horowitz and S. J. Samuels (eds), *Comprehending oral and written language* (pp. 55–82). New York: Academic Press.

Halliwell, S., and Jones, B. (1991). *On target*. London: Centre for Information on Language Teaching and Research.

Hansen, C. C. (2004). Teacher talk: promoting literacy development through response to story. *Journal of Research in Childhood Education, 19*, 115.

Hart, B., and Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes Publishing.

Hatzigeorgiu, N., Mikros, G., and Carayannis, G. (2001). Word length, word frequencies and Zipf's law in the Greek language. *Journal of Quantitative Linguistics, 8*, 175–85.

Hawkins, R. (1989). Do second language learners acquire restrictive relative clauses on the basis of relational or configurational information? The acquisition of French subject, direct object, and genitive restrictive clauses by language learners. *Second Language Research, 5*, 156–88.

Hawkins, R. (2001a). *Second language syntax: a generative introduction*. Oxford: Blackwell.

Hawkins, R. (2001b). The theoretical significance of Universal Grammar in second language acquisition. *Second Language Research, 17*, 345–67.

Hawkins, R. (2004). The contribution of the theory of Universal Grammar to our understanding of the acquisition of French as a second language. *Journal of French Language Studies, 14*, 233–55.

Hawkins, R., and Chan, C. Y.–H. (1997). The partial availability of Universal Grammar in second language acquisition: the 'failed functional features hypothesis'. *Second Language Research, 13*, 187–226.

Hawkins, R., and Franceschina, F. (2004). Explaining the acquisition and non-acquisition of determiner–noun gender concord in French and Spanish. In P. Prévost and J. Paradis (eds), *The acquisition of French in different contexts: focus on functional categories* (pp. 175–205). Amsterdam: John Benjamins.

Head Start Bureau. (2000). *Celebrating cultural and linguistic diversity in Head Start*. Washington, DC: Commissioner's Office of Research and Evaluation and the Head Start Bureau.

Heaslip, P. (1978). *School dinners* (Panjabi translation by J. Kalra). London: Methuen.

Heath, S. B. (1983). *Ways with words. Language, life, and work in communities and classrooms*. Cambridge: CUP.

Heatley, A., Nation, I. S. P., and Coxhead, A. (2002). *RANGE and FREQUENCY programs*. Retrieved 2 August 2008 from http://www.vuw.ac.nz/lals/staff/Paul_Nation

Heaton, J. B. (1966). *Composition through pictures*. Harlow: Longman.

Henrichs, L. F. (2007). Academic language use during book reading at home and at school: an exploration of early exposure to academic language in two contexts. Paper presented at the 12th Biennial Conference for Research on Learning and Instruction, Budapest, Hungary, August.

Herbstein, D. (1980). I'm in need of small talk. *The Sunday Times*, 27 January, p. 12.

Herschensohn, J. (2001). Missing inflection in L2 French: accidental infinitives and other verbal deficits. *Second Language Research, 17*, 273–305.

Hindmarsh, R. (1980). *Cambridge English lexicon*. Cambridge: CUP.

Hoey, M. (1983). *On the surface of discourse*. London: George Allen and Unwin.

Hoey, M. (2003). What's in a word? *English Teaching Professional*, *27*. Retrieved 24 October 2008, from http://www.macmillandictionaries.com/med-magazine/August2003/10-Feature-Whats-in-a-word.htm

Hoey, M. (2005). *Lexical priming*. London: Routledge.

Hoff, E. (2003). The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74*, 1368–78.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26*, 55–88.

Hoff, E., and Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development, 73*, 418–33.

Hoff-Ginsberg, E. (1991). Mother–child conversation in different social classes and communicative settings. *Child Development, 62*, 782–96.

Holmes, D. I., and Singh, S. (1996). A stylometric analysis of conversational speech of aphasic patients. *Literary and Linguistic Computing, 11*, 133–40.

Hoover, W. A., and Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*, 127–60.

Hopper, P. (1988). Emergent grammar and the a priori grammar postulate. In D. Tannen, (ed.), *Linguistics and context: connecting observation and understanding* (pp. 117–34). Norwood, NJ: Ablex.

Hopper, P. (1998). Emergent grammar. In M. Tomasello (ed.), *The new psychology of language* (pp. 155–75). Mahwah, NJ: Erlbaum.

Hounslow Bilingual Support Project. (1990). *The fox and the crane* (Gujarati translation by K. Gomez). Hounslow: Hounslow Community Language Service.

Housen, A. (2002). A corpus-based study of the L2 acquisition of the English verb system. In S. Granger, J. Hung and S. Petch-Tyson (eds), *Computer learner corpora, second language acquisition and foreign language learning* (pp. 77–116). Amsterdam: John Benjamins.

Howatt, A. P. R. (1984). *A history of English language teaching*. Oxford: Oxford University Press.

Hu, M., and Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*, 403–30.

Hulstijn, J. H. (2001). Intention and incidental second language vocabulary learning: a reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (ed.), *Cognition and second language instruction* (pp. 258–86). Cambridge: CUP.

Hunston, S., and Francis, G. (2000). *Pattern grammar: a corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., and Lyons, T. (1991). Early vocabulary growth: relation to language input and gender. *Developmental Psychology, 27*, 236–48.

Huttenlocher, J., Levine, S., and Vevea, J. (1998). Environmental input and cognitive growth: a study using time-period comparisons. *Child Development, 69*, 1012–29.

Huttenlocher, P. R. (2002). *Neural plasticity: the effects of environment on the development of the cerebral cortex*. Cambridge, MA: Harvard University Press.
Hyltenstam, K. (1988). Lexical characteristics of near-native second-language learners of Swedish. *Journal of Multilingual and Multicultural Development*, *9*, 67–84.
IELTS. (2008). IELTS Task 2 Writing band descriptors (public version). Retrieved 2 September 2008 from http://www.ielts.org/pdf/UOBDs_WritingT2.pdf
Ife, A., Vives Boix, G., and Meara, P. (2000). The impact of study abroad on the vocabulary development of different proficiency groups. *Spanish Applied Linguistics*, *4*, 55–84.
Jackendoff, R. (2002). *Foundations of language: brain, meaning, grammar, evolution*. Oxford: OUP.
Jackson, L. (1974). The myth of the elaborated and restricted code. *Higher Education Review, 6*, 65–81.
Jacobs, G. M., Dufon, P., and Hong, F. C. (1994). L1 and L2 vocabulary glosses in L2 reading passages: their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading, 17*, 19–28.
Jarvis, S. (2002). Short texts, best fitting curves, and new measures of lexical diversity. *Language Testing, 19*, 57–84.
Jensen, A. R. (1969). How much can we boost I.Q. and scholastic achievement? *Harvard Educational Review, 33*, 1–123.
Jisa, H., and Kern, S. (1998). Relative clauses in French children's narrative texts. *Journal of Child Language*, *25*, 623–52.
Johnson, W. (1944). Studies in language behavior: I. A program of research. *Psychological Monographs*, *56*, 1–15.
Jones, L. (1979). *Notions in English*. Cambridge: CUP.
Karmiloff Smith, A. (1979). *A functional approach to child language: a study of determiners and reference*. Cambridge: CUP.
Kharma, N. N., and Hajjaj, A. H. (1989). Use of the mother tongue in the ESL classroom. *International Review of Applied Linguistics*, *27*, 223–35.
Kim, S. H., and Elder, C. (2005). Language choices and pedagogical functions in the foreign language classroom: a cross-linguistic functional analysis of teacher talk. *Language Teaching Research*, *9*, 355–80.
Kindler, A. (2002). *Survey of the States' limited English proficient students and available educational programs and services, 2000–2001. Summary report*. Washington, DC: National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs.
Klee, T., and Fitzgerald, M. (1985). The relation between grammatical development and mean length of utterance in morphemes. *Journal of Child Language, 12*, 251–69.
Klein-Braley, C. (1985a). Advanced prediction of test difficulty. *Fremdsprachen und Hochschule. AKS Rundbrief*, *13/14*, 23–41.
Klein-Braley, C. (1985b). C-tests as placement tests for German university students of English. *Fremdsprachen und Hochschule. AKS Rundbrief*, *13/14*, 96–100.
Knight, S. (1994). Dictionary use while reading: the effects on comprehension and vocabulary acquisition for students of different verbal abilities. *The Modern Language Journal*, *78*, 285–99.
Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Erlbaum.
Krashen, S. (1985). *Inquiries and insights: essays on second language teaching, bilingual education and literacy*. Hayward, CA: Alemany Press.
Krashen, S. (1993). *The power of reading*. Englewood, CO: Libraries Unlimited.

Krashen, S. (2007). Literacy campaigns: access to books is the first step. *Literacy Network News*. Retrieved 24 October 2008, from http://www.sdkrashen.com/articles/literacy_campaigns/index.html

Krashen, S. D., and Terrell, T. D. (1988). *The natural approach*. London: Prentice Hall.

Kroll, J., Bobb, S. C., and Wodnieczka, Z. (2006). Language selectivity is the exception, not the rule: arguments against a fixed locus of language selection in bilingual speech. *Bilingualism, Language and Cognition*, *9*, 119–35.

Kuiper, K. (2004). Formulaic performance in conventionalised varieties of speech. In N. Schmitt (ed.), *Formulaic sequences: acquisition, processing, and use* (pp. 37–54). Philadelphia, PA: John Benjamins.

Labov, W. (1973). The logic of non-standard English. In N. Keddie (ed.), *Tinker, tailor: the myth of cultural deprivation* (pp. 21–67). Harmondsworth: Penguin.

Lakshmanan, U. (1998). Functional categories and related mechanisms in child second language acquisition. In S. Flynn, G. Martohardjono and W. O'Neil (eds), *The generative study of second language acquisition* (pp. 3–16). Mahwah, NJ: Erlbaum.

Lancaster Corpus Linguistics website (University Centre for Computer Corpus Research on Language). Accessed 24 October 2008 at http://www.ucrel.lancs.ac.uk

Lardiere, D. (1998). Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research, 14*, 359–75.

Lardiere, D. (2000). Mapping features to forms in second language acquisition. In J. Archibald (ed.), *Second language acquisition and linguistic theory* (pp. 102–29). Balden, MA: Blackwell.

Laudanna, A., and Voghera, M. (2002). Nouns and verbs as grammatical classes in the lexicon. *Italian Journal of Linguistics*, *14*, 9–26.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren and M. Nordemann (eds), *Special language: from humans to thinking machines* (pp. 316–23). Clevedon: Multilingual Matters.

Laufer, B. (2005). Focus on form in second language vocabulary acquisition. In S. H. Foster-Cohen, M. P. Garcia-Mayo and J. Cenoz (eds), *EUROSLA Yearbook* 5, (pp. 223–50). Amsterdam: John Benjamins.

Laufer, B., and Hadar L. (1997). Assessing the effectiveness of monolingual, bilingual, and 'bilingualized' dictionaries in the comprehension and production of new words. *Modern Language Journal*, *81*, 189–96.

Laufer, B., and Nation, P. (1995). Vocabulary size and use: lexical richness in L2 written production. *Applied Linguistics, 16*, 307–22.

Laufer, B., and Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing, 16*, 33–51.

Layzer, J. I., Goodson, B. D., and Moss, M. (1993). *Life in preschool: volume one of an observational study of early childhood programs for disadvantaged four year olds*. Cambridge, MA: Abt Associates.

Leech, G., Rayson P., and Wilson A. (2001). *Word frequencies in written and spoken English*. Harlow: Longman.

Leseman, P. P. M., and van Tuijl, C. V. (2006). Cultural diversity in early literacy: findings from Dutch studies. In D. K. Dickinson and S. B. Neuman (eds), *Handbook of early literacy research* (Vol. 2) (pp. 211–28). New York: Guilford.

Levelt, W. (1989). *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. (1993). Language use in normal speakers and its order. In G. Blanken, H. Dittman, H. Grimm, J. Marshal and C. Wallesch (eds), *Linguistic disorders and pathologies* (pp. 1–15). Berlin: de Gruyter.

Levine, G. (2003). Student and instructor beliefs and attitudes about target language use, first language use, and anxiety: report of a questionnaire study. *The Modern Language Journal*, *87*, 343–64.

Li, D. C. S. (2000). Cantonese-English code-switching research in Hong Kong: a Y2K review. *World Englishes, 19*, 305–22.

Liebscher, G., and Dailey-O'Cain, J. (2005). Learner code-switching in the content-based foreign language classroom. *The Modern Language Journal*, *89*, 234–47.

Linnarud, M. (1983). On lexis: the Swedish learner and the native speaker compared. In K. Sajavaara (ed.), *Cross language analysis and second language acquisition* (pp. 249–61). Jyväskylä: University of Jyväskylä.

Little, D., and Singleton, D. (1992). The C-test as an elicitation instrument in second language research. In R. Grotjahn (ed.), *Der C-test. Theoretische Grundlagen und praktische Anwendungen* (pp. 173–92). Bochum: Brockmeyer.

Littleton, K., Mercer, N., Dawes, L., Wegerif, R., Rowe, D., and Sams, C. (2005). Talking and thinking together at Key Stage 1. *Early Years: an International Journal of Research and Development, 25*, 167–82.

Liu, D., Ahn, G., Beak, K., and Han, N. (2004). South Korean high school English teachers' codeswitching: questions and challenges in the drive for maximal use of English in teaching. *TESOL Quarterly*, *38*, 605–38.

Liu, J., and Jiang, Y. (2004). An investigation and analysis of EFL classroom language: a case study. *Foreign Language Teaching and Research*, *36*, 285–91.

Llurda, E. (ed.). (2005). *Non-native language teachers: perceptions, challenges, and contributions to the profession*. Boston, MA: Springer.

Locke, J. (1997). A theory of neurolinguistic development. *Brain and Language, 58*, 265–326.

Macaro, E. (1995). Target language use in Italy. *Language Learning Journal*, *11*, 52–4.

Macaro, E. (1997). *Target language, collaborative learning and autonomy*. Clevedon: Multilingual Matters.

Macaro, E. (2000). Issues in target language teaching. In K. Field (ed.), *Issues in modern language teaching* (pp. 171–89). London: Routledge.

Macaro, E. (2001). Analysing student teachers' codeswitching in foreign language classrooms: theories and decision making. *The Modern Language Journal*, *85*, 531–48.

Macaro, E. (2005). Codeswitching in the L2 classroom: a communication and learning strategy. In E. Llurda (ed.), *Non-native language teachers: perceptions, challenges, and contributions to the profession* (pp. 63–84). Boston, MA: Springer.

Macaro, E., and Mutton, T. (2002). Developing language teachers through a co-researcher model. *Language Learning Journal*, *25*, 27–39.

McCarthy, P. M., and Jarvis, S. (2007). vocd: a theoretical and empirical evaluation. *Language Testing, 24*, 459–88.

Mace-Matluck, B. J. (1982). *Literacy instruction in bilingual settings: a synthesis of current research*, Los Alamitos, CA: National Center for Bilingual Research.

McGill-Franzen, A., and Allington, R. L. (1991). The gridlock of low reading achievement: perspectives on practice and policy. *Remedial and Special Education, 12*, 20–30.

McKeown, M., and Beck, I. (2003). Taking advantage of read alouds to help children make sense of decontextualized language. In A. van Kleeck, S. Stahl and E. Bauer (eds), *On reading books to children: parents and teachers*. Mahwah, NJ: Erlbaum.

McNaughton, S. (2006). Considering culture in research-based interventions to support early literacy. In D. K. Dickinson and S. B. Neuman (eds), *Handbook of early literacy research* (Vol. 2) (pp. 229–42). New York: Guilford.

MacWhinney, B. (2000). *The CHILDES project: tools for analyzing talk* (3rd edn). Mahwah, NJ: Erlbaum.
MacWhinney, B. (2001). Emergence from what? *Journal of Child Language*, *28*, 726–32.
Magnuson, K. A., and Waldfogel, J. (2005). Early childhood care and education: effects on ethnic and racial gaps in school readiness. *Future of Children, 15*, 169–96.
Malvern, D., and Richards, B. J. (1997). A new measure of lexical diversity. In A. Ryan and A. Wray (eds), *Evolving models of language* (pp. 58–71). Clevedon: Multilingual Matters.
Malvern, D., and Richards, B. J. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, *19*, 85–104.
Malvern, D., Richards, B. J., Chipere, N., and Durán, P. (2004). *Lexical diversity and language development: quantification and assessment*. Basingstoke: Palgrave Macmillan.
Marriot, H. (1995). Acquisition of politeness patterns by exchange students in Japan. In B. F. Freed (ed.), *Second language acquisition in study abroad context* (pp. 197–224). Amsterdam: John Benjamins.
Marsden, E., and David, A. (in press). Vocabulary use during conversation: a cross-sectional study of development from year 9 to year 13 amongst learners of Spanish and French. *Language Learning Journal*.
Mason, B., and Krashen, S. (1997). Extensive reading in English as a foreign language. *System*, *25*, 91–102.
Mauranen, A. (2004). Speech corpora in the classroom. In G. Aston, S. Bernardini and D. Stewart (eds), *Corpora and language learners* (pp. 195–211). Amsterdam: John Benjamins.
Mayer, M. (1969). *Frog where are you?* New York: Dial Books for Young Readers.
Mayer, M. (1974). *Frog goes to dinner*. New York: Dial Books for Young Readers.
Meara, P. (1997). Towards a new approach to modelling vocabulary acquisition. In N. Schmitt and M. McCarthy (eds), *Vocabulary: description, acquisition and pedagogy* (pp. 109–21). Cambridge: CUP.
Meara, P., and Bell, H. (2001). P_Lex: a simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect*, *16*, 5–19.
Meara, P., and Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing, 4*, 142–54.
Meara, P., and Jones, G. (1990). *Eurocentre's vocabulary size test: user's guide*. Zurich: Eurocentres.
Meara, P., and Milton, J. (2003). *X_Lex, the Swansea Levels Test*. Newbury: Express.
Medgyes, P. (1999). *The non-native teacher*. Ismaning: Hueber.
Mehan, H. (1979). *Learning lessons: social organization in the classroom*. Cambridge, MA: Harvard University Press.
Ménard, N. (1983). *Mesure de la richesse lexicale*. Geneva: Slatkine.
Meng, Q. (2005). The effects of codeswitching on vocabulary learning in the EFL classroom. Unpublished master's thesis, University of Oxford, UK.
Michaels, S. (1981). Sharing time: children's narrative styles and differential access to literacy. *Language in Society*, *10*, 423–42.
Michaels, S. (1986). Narrative presentations: an oral preparation for literacy with first graders. In J. Cook-Gumperz (ed.), *The social construction of literacy* (pp. 94–116). Cambridge: CUP.
Michaels, S., O'Connor, C., and Resnick, L. (2007). Deliberative discourse idealized and realized: accountable talk in the classroom and in civic life. *Studies in Philosophy and Education*. Retrieved 24 October 2008, from http://www.lrdc.pitt.edu/publications/pdfs/200807161458010.2007_Deliberative%20Discourse.pdf

Milton, J. (2006). Language lite? Learning French vocabulary in school. *Journal of French Language Studies, 16*, 187–205.

Milton, J. (2007a). The value of vocabulary size measures in a language framework of reference. In M. Martyniuk (ed.), *Towards a common framework of reference for language(s) of school education?* (pp. 107–16). Krakow: Universitas.

Milton, J. (2007b). Lexical profiles, learning styles and construct validity of lexical size tests. In H. Daller, J. Milton and J. Treffers-Daller (eds), *Modelling and assessing vocabulary knowledge* (pp. 45–58). Cambridge: CUP.

Milton, J. (2008). French vocabulary breadth among learners in the British school and university system: comparing knowledge over time. *Journal of French Language Studies, 18*, 1–16.

Milton, J., and Hales, T. (1997). Applying a lexical profiling system to technical English. In A. Ryan and A. Wray (eds), *Evolving models of language* (pp. 72–83). Clevedon: Multilingual Matters.

Milton, J., and Meara, P. (1995). How periods abroad affect vocabulary growth in a foreign language. *ITL Review of Applied Linguistics, 107–108*, 17–34.

Milton, J., Wade, J., and Hopkins, N. (forthcoming). Aural word recognition and oral competence in a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse, M. Torreblanca-López and M. López-Jiménez (eds), *Further insights into non-native vocabulary teaching and learning*.

Ministère de L'Education Nationale. (1993). *Anglais: classes de collèges, 6e, 5e, 4e, 3e* [English: classes in lower secondary school, 1st, 2nd, 3rd, 4th year]. Paris: Centre Nationale de Documentation Pédagogique.

Ministry of Education (PR China). (2001). *English curriculum standards for compulsory education and senior secondary education*. Beijing: Ministry of Education.

Mitchell, R., and Myles, F. (2004). *Second language learning theories* (2nd edn). London: Arnold.

Möhle, D., and Raupach, M. (1983). *Planen in der Fremdsprache: Analyse von Lernersprache Französisch*. Bern: Peter Lang.

Moon, R. (1997). Vocabulary connections: multi-word items in English. In N. Schmitt and M. McCarthy (eds), *Vocabulary: description, acquisition and pedagogy* (pp. 40–63). Cambridge: CUP.

Morris, L., and Cobb, T. (2003). Vocabulary profiles as predictors of the academic performance of teaching English as a second language trainees. *System*, *32*, 75–87.

Morris, L., and Tremblay, M. (2002). Attention to unstressed words and the acquisition of grammatical morphology by French-speaking ESL students. *Canadian Modern Language Review*, *58*, 364–85.

Multilingual Resources for Children Project (MRC). (1995). *Building bridges: multilingual resources for children*. Clevedon: Multilingual Matters.

Munns, G., Arthur, L., Downes, T., Gregson, R., Power, A., Sawyer, W., Singh, M., Thistleon-Martin, J., and Steele, F. (2006). *Motivation and engagement of boys: evidence-based teaching practices*. Canberra: ACT, DEST. Retrieved 24 October 2008, from http://www.dest.gov.au/sectors/research_sector/publications_resources/profiles/motivation_engagement_boys.htm

Muysken, P. (2000). *Bilingual speech: a typology of code-mixing*. Cambridge: CUP.

Myers-Scotton, C. (1993). *Duelling languages: grammatical structure and codeswitching*. Oxford: OUP.

Myles, F. (2004). From data to theory: the over-representation of linguistic knowledge in SLA. *Transactions of the Philological Society, 102*, 139–68.

Myles, F. (2005). The emergence of morpho-syntactic structure in French L2. In J.-M. Dewaele (ed.), *Focus on French as a foreign language: multidisciplinary approaches*. Clevedon: Multilingual Matters.

Myles, F., and Towell, R. (eds). (2004). The acquisition of French as a second language [Special issue]. *Journal of French Language Studies*, *14*.

Myles, F., Hooper, J., and Mitchell, R. (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning, 48*, 323–63.

Myles, F., Mitchell, R., and Hooper, J. (1999). Interrogative chunks in French L2: a basis for creative construction? *Studies in Second Language Acquisition, 21*, 49–80.

Nagy, W. E. (1997). On the role of context in first- and second-language vocabulary learning. In N. Schmitt and M. McCarthy (eds), *Vocabulary: description, acquisition and pedagogy* (pp. 64–83). Cambridge: CUP.

Nation, I. S. P. (1983). *Teaching and learning vocabulary*. Wellington: Victoria University.

Nation, I. S. P. (ed.). (1984). *Vocabulary lists: words, affixes and stems*. Wellington: English University of Wellington, English Language Institute.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, MA: Heinle and Heinle.

Nation, I. S. P (2001). *Learning vocabulary in another language*. Cambridge: CUP.

Nation, I. S. P. (2007). Fundamental issues in modelling and assessing vocabulary knowledge. In H. Daller, J. Milton and J. Treffers-Daller (eds), *Modelling and assessing vocabulary knowledge* (pp. 35–43). Cambridge: CUP.

Nation, I. S. P. Home page. Accessed 6 September 2008, at http://www.victoria.ac.nz/lals/staff/paul–nation/nation.aspx

Nation, I. S. P., and Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt and M. McCarthy (eds), *Vocabulary: description, acquisition and pedagogy* (pp. 6–19). Cambridge: CUP.

Nation, K., Adams, J. W., Bowyer Crane, C. A., and Snowling, M. J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child Psychology, 73*, 139–58.

Nation, K., Clarke, P., Marshall, C. M., and Durand, M. (2004). Hidden language impairments in children: parallels between poor reading comprehension and specific language impairment? *Journal of Speech, Language, and Hearing Research, 47*, 199–211.

National Center for Education Statistics. (2001). *Survey of classes that serve children prior to kindergarten in public schools* (FRSS 78). Washington, DC: Institute of Education Sciences, US Department of Education.

National Center for Education Statistics. (2007a). *Digest of education statistics: 2007* (Document No. 2008022). Washington, DC: Institute of Education Sciences, US Department of Education.

National Center for Education Statistics. (2007b). *National assessment of educational progress*. Washington, DC: Institute of Education Sciences, US Department of Education.

National Clearinghouse for English Language Acquisition. (2003). *Resources about early childhood education*. Washington, DC: US Department of Education.

National Reading Panel. (2000). *Report of the National Reading Panel: Teaching children to read. Reports of the subgroups*. Washington, DC: National Institute for Literacy.

Nelson, K. E. (1989). Strategies for first language teaching. In M. Rice and R. L. Schiefelbusch (eds), *The teachability of language* (pp. 263–310). Baltimore: Brookes.

NICHD Early Child Care Research Network. (2005). Pathways to reading: the role of oral language in the transition to reading. *Developmental Psychology, 41*, 428–42.

Norris, J., and Ortega, L. (2006). *Synthesising research on language learning and teaching*. New York: John Benjamins.

Oakey, D. (2005). Academic vocabulary in academic discourse: the phraseological behaviour of EVALUATION in economics research articles. In E. Tognini-Bennelli and G. D. L. Camiciotti (eds), *Strategies in Academic Discourse* (pp. 169–84). Amsterdam: John Benjamins.

Oakhill, J., Hartt, J., and Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing, 18*, 657–86.

OECD (Organization for Economic Cooperation and Development). (2002). *Reading for change: performance and engagement across countries. Results from Pisa 2000*. New York: OECD.

O'Loughlin, K. (1995). Lexical density in candidate output on direct and semi-direct versions of an oral proficiency test. *Language Testing, 12*, 217–37.

Opitz, C. (2008). L1 attrition in adult bilinguals: the impact of L2 proficiency, language use and linguistic aptitude. Paper presented at the 15th World Congress of Applied Linguistics, Essen University, August.

Orosz, A. (2007). Gondolatok az angol nyelvoktatásról a 9–14 éves korú tanulók angol nyelvi szókincs mérése alapján. *Apáczai Napok 2006 Tanulmánnykötet I*. Győr: Nyugat Magyarországi Egyetem Apáczai Csere János Kar.

Palinscar, A. S., and Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*, 117–75.

Pan, B. A., Rowe, M. L., Singer, J. D., and Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development, 76*, 763–82.

Pan, B. A., Rowe, M. L., Spier, E., and Tamis Lemonda, C. (2004). Measuring productive vocabulary of toddlers in low-income families: concurrent and predictive validity of three sources of data. *Journal of Child Language, 31*, 587–608.

Partington, A. (1998). *Patterns and meaning: using corpora for English language research and teaching*. Amsterdam: John Benjamins.

Patkowski, M. S. (1991). Basic skills tests and academic success of ESL college students. *TESOL Quarterly*, *35*, 735–38.

Pawley, A., and Syder, F. (1983). Two puzzles for linguistic theory: native-like selection and native-like fluency. In J. Richards and R. Schmidt (eds), *Language and communication* (pp. 191–226). London: Longman.

Pearsall, J. (ed.). (1998). *The new Oxford English dictionary*. Oxford: OUP.

Peck, S. (1987). Signs of learning: child nonnative speakers in tutoring sessions with a child native speaker. *Language Learning, 37*, 545–71.

Perry, T., and Delpit, L. (eds) (1997). *The real ebonics debate: power, language, and the education of African-American children*. Special issue of *Rethinking Schools*. Retrieved 24 October 2008 from http://www.rethinkingschools.org/publication/ebonics/ebotoc.shtml

Peterson, C., Jesso, B., and McCabe, A. (1999). Encouraging narratives in preschoolers: an intervention study. *Journal of Child Language, 26*, 49–67.

Phythian–Sence, C., and Wagner, R. K. (2007). Vocabulary acquisition: a primer. In R. K. Wagner, A. E. Muse and K. R. Tannenbaum (eds), *Vocabulary acquisition: implications for reading comprehension* (pp. 1–14). New York: Guilford Press.

Pierce, A. (1992). *Language acquisition and syntactic theory: a comparative analysis of French and English child grammars*. London: Kluwer.

Pollock, J.-Y. (1989). Verb movement, universal grammar and the structure of IP. *Linguistic Inquiry*, *20*, 365–424.

Poplack, S. (1980). Sometimes I'll start a sentence in English y termino en espanol. *Linguistics*, *18*, 581–616.

Postlethwaite, T. N., and Ross, K. N. (1992). *Effective schools in reading: implications for educational planners: an exploratory study*. The Hague: The International Association for the Evaluation of Educational Achievement.

Power Vocabulary Builder. Retrieved 7 March 2008, from http://www.powervocabularybuilder.com

Pressley, M., Disney, L., and Anderson, K. (2007). Landmark vocabulary instructional research and the vocabulary instructional research that makes sense now. In R. K. Wagner, A. E. Muse and K. R. Tannenbaum (eds), *Vocabulary acquisition: implications for reading comprehension* (pp. 205–32). New York: Guilford Press.

Prévost, P., and Paradis, J. (eds). (2004). *The acquisition of French in different contexts: focus on functional categories*. Amsterdam: John Benjamins.

Prévost, P., and White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research, 16*, 103–33.

Primary National Strategy. (2007a). *Leading improvement using the Primary Framework: improving reading*. London: DfES. Retrieved 24 October 2008 from http://www.standards.dfes.gov.uk/primaryframeworks/downloads/PDF/Lit_read_cpdoutline.pdf

Primary National Strategy. (2007b). *What can schools do to increase the vocabulary of children who start education with a limited vocabulary? A note of the research evidence*. London: DfES, Schools Research Team (Schools Analysis and Research Division). Retrieved 24 October 2008, from http://www.standards.dfes.gov.uk/primaryframeworks/downloads/PDF/lit_read_imp_paper_b.pdf

Proctor, C. P., August, D., Carlo, M. S., and Snow, C. (2006). The intriguing role of Spanish language vocabulary knowledge in predicting English reading comprehension. *Journal of Educational Psychology, 98*, 159–69.

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review*, *56*, 282–308.

Qualifications and Curriculum Authority. (2008). *English: the National Curriculum for England*. Retrieved 24 October 2008, from http://www.nc.uk.net

Raatz, U. (1985). The factorial validity of C-tests. *Fremdsprachen und Hochschule. AKS Rundbrief*, *13/14*, 42–54.

Radford, A. (1996). Towards a structure-building model of acquisition. In H. Clahsen (ed.), *Generative perspectives on language acquisition* (pp. 43–90). Amsterdam: John Benjamins.

Ramirez, J. D. (1991). *Longitudinal study of structured English immersion strategy, early-exit and late-exit transitional bilingual education programs for language-minority children. Final Report* (Vols 1 and 2). San Mateo, CA: Aguirre Internationalo.

Ransdell, S. (2002). The care and feeding of monolingual and bilingual university students in South Florida: implications for assessment and training. *Psychology Learning and Teaching*, *3*, 126–30.

Rapp, D. N., van den Broek, P., McMaster, K. L., Panayiota, K., and Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: a perspective for research and intervention. *Scientific Studies of Reading, 11*, 289–312.

Raven, J. C. (1960). *Guide to the standard progressive matrices*. London: H. K. Lewis.

Ravid, D. (2006). Semantic development in textual contexts during the school years: noun scale analyses. *Journal of Child Language, 33*, 791–821.

Ravid, D., and Tolchinksky, L. (2002). Developing linguistic literacy: a comprehensive model. *Journal of Child Language, 29*, 417–47.

Read, J. (2000). *Assessing vocabulary*. Cambridge: CUP.
Read, J., and Nation, P. (2004). Measurement of formulaic sequences. In N. Schmitt (ed.), *Formulaic sequences: acquisition, processing, and use* (pp. 23–36). Philadelphia, PA: John Benjamins.
Reese, L., Garnier, H., Gallimore, R., and Goldenberg, C. (2000). Longitudinal analysis of the antecedents of emergent Spanish literacy and middle-school English reading achievement of Spanish-speaking students. *American Educational Research Journal, 37*, 633–62.
Regan, V. (1995). The acquisition of sociolinguistic native speaker norms: effects of a year abroad on second language learners of French. In B. F. Freed (ed.), *Second language acquisition in a study abroad context* (pp. 245–68). Amsterdam: John Benjamins.
Regan, V. (1998). Sociolinguistics and language learning in a study abroad context. *Frontiers, 4*, 61–90.
Richards, B., and Malvern, D. (2004). Investigating the validity of a new measure of lexical diversity for root and inflected forms. In K. Trott, S. Dobbinson and P. Griffiths (eds), *The child language reader* (pp. 81–9). London: Routledge.
Richards, B. J., and Malvern, D. D. (2007). Validity and threats to the validity of vocabulary measurement. In H. Daller, J. Milton, and J. Treffers-Daller (eds), *Modelling and assessing vocabulary knowledge* (pp. 79–92). Cambridge: CUP.
Richards, B. J., Malvern, D. D., and Graham, S. (2008). Word frequency and trends in the development of French vocabulary in lower intermediate students during Year 12 in English schools. *Language Learning Journal*, *36*, 199–213.
Richards, J. C., and Rogers, T. S. (1986). *Approaches and methods in language teaching*. Cambridge: CUP.
Rizzi, L. (1994). Early null subjects and root null subjects. In T. Hoekstra and B. Schwartz (eds), *Language acquisition studies in generative grammar* (pp. 151–76). Amsterdam: John Benjamins.
Robinson, P. (2001). Task complexity, task difficulty, and task production: exploring interactions in a componential framework. *Applied Linguistics*, *22*, 27–57.
Robinson, P., and Gilabert, R. (2007). Task complexity, the Cognition Hypothesis, and second language learning and performance. *International Review of Applied Linguistics in Language Teaching*, *45*, 161–76.
Rolin-Ianziti, J., and Brownlie, S. (2002). Teacher use of learners' native language in the foreign language classroom. *Canadian Modern Language Review*, *58*, 402–26.
Romaine, S. (1995). *Bilingualism* (2nd edn). Oxford: Blackwell.
Rondal, J. A., Ghiotto, M., Bredart, S., and Bachelet, J. (1987). Age-relation, reliability, and grammatical validity of measures of utterance length. *Journal of Child Language*, *14*, 433–46.
Rosenfeld, M., Oltman, P. K., and Sheppard, K. (2004). *Investigating the validity of the TOEFL test: a feasibility study using content and criterion-related strategies* (ETS report no. 71). Princeton, NJ: Educational Testing Service. Retrieved 24 October 2008, from http://www.ets.org
Rowe, M. L. (2008). Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language, 35*, 185.
Sandora, C., Beck, I., and McKeown, M. (1999). A comparison of two discussion strategies on students' comprehension and interpretation of complex literature. *Reading Psychology, 20*, 177–212.
Saunders, W. M., and Goldenberg, C. (1999). Effects of instructional conversations and literature logs on limited- and fluent-English-proficient students' story comprehension and thematic understanding. *The Elementary School Journal, 99*, 277–301.

Saunders, W. M., and O'Brien, G. (2006). Oral language. In F. Genesee, K. Lindholm-Leary, W. M. Saunders and D. Christian (eds), *Educating English language learners: a synthesis of research evidence*. New York: CUP.

Saville-Troike, M. (1988). Private speech: evidence for second language learning strategies during the 'silent' period. *Journal of Child Language, 15*, 567–90.

Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis)abilities. In S. B. Neuman and D. K. Dickinson (eds), *Handbook of early literacy research* (pp. 97–110). New York: Guilford Press.

Scarborough, H. S., Rescorla, L., Tager-Flusberg, H., Fowler, A. E., and Sudhalter, V. (1991). The relation of utterance length to grammatical complexity in normal and language-disordered groups. *Applied Psycholinguistics*, *12*, 23–45.

Schachter, J. (1974). An error in error analysis. *Language Learning*, *27*, 205–14.

Schleppegrell, M. J. (2001). Linguistic features of the language of schooling. *Linguistics and Education, 12*, 431–59.

Schleppegrell, M. J. (2004). *The language of schooling: a functional linguistics perspective*. London: Erlbaum.

Schmidt, R. (1983). Interaction, acculturation and the acquisition of communicative competence. In N. Wolfson and E. Judd (eds), *Sociolinguistics and second language acquisition* (pp. 137–74). Rowley, MA: Newbury House.

Schmitt, N., and Carter, R. (2004). Formulaic sequences in action: an introduction. In N. Schmitt (ed.), *Formulaic sequences: acquisition, processing, and use* (pp. 1–22). Philadelphia, PA: John Benjamins.

Schmitt, N., Grandage, S., and Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In N. Schmitt (ed.), *Formulaic sequences: acquisition, processing, and use* (pp. 127–52). Philadelphia, PA: John Benjamins.

Schmitt, N., Schmitt, D., and Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*, 55–88.

Schwartz, B., and Sprouse, R. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research, 12*, 40–72.

Sealey, A., and Thompson, P. (2004). 'What do you call the dull words?' Primary school children using corpus-based approaches to learn about language. *English in Education*, *38*, 80–91.

Sealey, A., and Thompson, P. (2006). 'Nice things get said': corpus evidence and the National Literacy Strategy. *Literacy*, *40*, 22–8.

Sénéchal, M., Ouellette, G., and Rodney, D. (2006). The misunderstood giant: on the predictive role of early vocabulary to future reading. In D. K. Dickinson and S. B. Neuman (eds), *Handbook of early literacy research* (Vol. 2) (pp. 173–82). New York: Guilford Press.

Shanahan, T., and Beck, I. (2006). Effective literacy teaching for English-language learners. In D. August and T. Shanahan (eds), *Developing literacy in second-language learners: report of the National Literacy Panel on language-minority children and youth*. Mahwah, NJ: Erlbaum.

Siegal, M. (1995). Individual differences and study abroad: women learning Japanese in Japan. In B. F. Freed (ed.), *Second language acquisition in a study abroad context* (pp. 225–44). Amsterdam: John Benjamins.

Sigma Assessment Systems. Accessed 25 August 2008, at http://www.sigmaassessmentsystems.com/default.asp

Sigott, G. (2004). *Towards identifying the C-test construct*. Frankfurt: Peter Lang.

Sigott, G. (2006). How fluid is the C-test construct? In R. Grotjahn (ed.), *Der C-Test: Theorie, Empirie, Anwendungen* [The C-test: theory, empirical research, applications] (pp. 139–46). Frankfurt: Peter Lang.

Silverman, R. D. (2007). Vocabulary development of English-language and English-only learners in kindergarten. *Elementary School Journal, 107*, 365–84.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: OUP.

Sinclair, J. M. (2004). New evidence, new priorities, new attitudes. In J. M. Sinclair (ed.), *How to use corpora in language teaching* (pp. 271–99). Amsterdam: John Benjamins.

Singleton, D., and Singleton, E. (2002). The C-test and L2 lexical acquisition/processing research. In J. A. Coleman, R. Grotjahn and U. Raatz (eds), *University language teaching and the C-test* (pp. 143–68). Bochum: AKS Verlag.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: OUP.

Skehan P., and Foster, P. (1997). The influence of planning and post-task activities on accuracy and complexity in task based learning. *Language Teaching Research, 1*, 185–211.

Skehan, P., and Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning, 49*, 93–120.

Skehan, P., and Foster, P. (2005). Strategic and on-line planning: the influence of surprise information and task time on second language performance. In R. Ellis (ed.), *Planning and task performance in a second language*. Amsterdam: John Benjamins.

Skehan, P., and Foster, P. (2007). Complexity, accuracy, fluency and lexis in task-based performance: a meta-analysis of the Ealing research. In S. Van Daele, A. Housen, F. Kuiken, M. Pierrard and I. Vedder (eds), *Complexity, accuracy, and fluency in second language use, learning, and teaching* (pp. 207–26). Brussels: University of Brussels Press.

Smith, J.A., and Kelly, C. (2002). Stylistic constancy and change across literary corpora: using measures of lexical richness to date works. *Computers and the Humanities*, *36*, 411–30.

Snow, C. (1983). Literacy and language: relationships during the preschool years. *Harvard Educational Review, 53*, 165–90.

Snow, C. E., Tabors, P. O., and Dickinson, D. K. (2001). Language development in the preschool years. In D. K. Dickinson and P. O. Tabors (eds), *Beginning literacy with language: young children learning at home and at school* (pp. 1–25). Baltimore, MD: Paul H. Brookes Publishing.

SocioTechnical Research Applications Incorporated. (1996). *Report on the ACYF bilingual/multicultural survey*. Washington, DC: Head Start Bureau.

Spira, E. G., Bracken, S. S., and Fischel, J. E. (2005). Predicting improvement after first-grade reading difficulties: the effects of oral language, emergent literacy, and behavior skills. *Developmental Psychology, 41*, 225–34.

Spycher, P. (2007). Academic writing of adolescent English learners: learning to use 'although'. *Journal of Second Language Writing, 16*, 238–54.

Staehr, L. (2009). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal, 36*, 139–52.

Stemmer, B. (1992). An alternative approach to C-test validation. In R. Grotjahn (ed.), *Der C-Test. Theoretische Grundlangen und praktische Anwendungen* (pp. 97–144). Bochum: AKS Verlag.

Storch, S. A., and Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: evidence from a longitudinal structural model. *Developmental Psychology, 38*, 934–47.

Strickland, D. S. (2001). Early intervention for African American children considered to be at risk. In S. B. Neuman and D. K. Dickinson (eds), *Handbook of early literacy research* (pp. 32–332). New York: Guilford Press.

Stubbs, M. (2001). *Words and phrases: corpus studies of lexical semantics*. Oxford: Blackwell.

Swan, M., and Walter, C. (1990). *The new Cambridge English course 2, student's book*. Cambridge: CUP.

Tabors, P. O., and Snow, C. E. (2001). Young bilingual children and early literacy development. In S. B. Neuman and D. K. Dickinson (eds), *Handbook of early literacy research*. New York: Guilford Press.

Tabors, P. O., Porche, M., and Ross, S. (2003). Predicting reading comprehension in a low-income sample. Paper presented at the Society for Research in Child Development, Tampa, FL, April.

Tarullo, L. B., and Zill, N. (2002). FACES: an assessment battery to track children's cognitive development in Head Start and early elementary school. Paper presented at the NICHD Workshop on Cognitive Development Measures for Large-Scale Studies, Washington, DC, May.

Tavakoli, P., and Foster, P. (2008). Task design and second language performance: the effect of narrative type on learner output. *Language Learning*, *58*, 439–73.

Tavakoli, P., and Skehan P. (2005), Planning, task structure, and performance testing. In R. Ellis (ed.), *Planning and task performance in a second language*. Amsterdam: John Benjamins.

Taylor, L., and Falvey, P. (2007). *IELTS collected papers* (Studies in Language Testing No. 19). Cambridge: CUP.

The Economist Intelligence Unit (TEIU). (2008). *The future of the international student market place: focus on China*. Retrieved 28 July 2008, from http://cemis.britishcouncil.org

Thomas, M. (1994). Assessment of L2 proficiency in second language acquisition research. *Language Learning, 44*, 307–36.

Thompson, P., and Sealey, A. (2007). Through children's eyes? Corpus evidence of the features of children's literature. *International Journal of Corpus Linguistics*, *12*, 1–23.

Tidball, F., and Treffers-Daller, J. (2005). Exploring measures of vocabulary richness in a corpus of oral narratives of native and non-native speakers of French. Paper presented at the annual meeting of the British Association for Applied Linguistics, Bristol University, September.

Tidball, F., and Treffers-Daller, J. (2007). Exploring measures of vocabulary richness in semi-spontaneous French speech. In H. Daller, J. Milton and J. Treffers-Daller (eds), *Modelling and assessing vocabulary knowledge* (pp. 133–49). Cambridge: CUP.

Tidball, F., and Treffers-Daller, J. (2008). Analyzing lexical richness in French learner language: what frequency lists and teacher judgements can tell us about basic and advanced words. *Journal of French Language Studies, 18*, 299–313.

Tizard, B., and Hughes, M. (1984). *Young children learning*. London: Fontana.

Towell, R., Hawkins, R., and Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, *17*, 84–119.

Treffers-Daller, J., Daller, H. M., Malvern, D. D., Richards, B. J., Meara, P. M., and Milton, J. (eds). (2008). Knowledge and use of the lexicon in French as a second language [Special issue]. *Journal of French Language Studies, 18*.

Treffers-Daller, J., Özsoy, S. A., and Van Hout, R. (2007). (In)complete acquisition of Turkish among Turkish-German bilinguals in Germany and Turkey: an analysis of complex embeddings in narratives. *International Journal of Bilingual Education and Bilingualism*, *10*, 248–76.

Trudgill, P. (1975). *Accent, dialect and the school*. London: Edward Arnold.

Tulaswicz, W., and Zajda, J. (eds). (1998). *Language awareness in the school curriculum*. Melbourne: James Nicholas.

Tunmer, W. E., and Hoover, W. A. (1992). Cognitive and linguistic factors in learning to read. In P. B. Gough, L. C. Ehri and R. Treiman (eds), *Reading acquisition* (pp. 175–214). Hillsdale, NJ: Erlbaum.

Turnbull, M. (2006). Employons 'le français' en français de base. *The Canadian Modern Language Review*, *62*, 611–28.

Tyler, A., and Nagy, W. (1990). Use of derivational morphology during reading. *Cognition*, *36*, 17–34.

Tymms, P. B., Merrell, C., and Henderson, B. (1997). The first year at school: a quantitative investigation of the attainment and progress of pupils. *Educational Research and Evaluation, 3*, 101–108.

Ultimate Vocabulary. Retrieved 7 March 2008, from http://www.ultimatevocabulary.com

Underwood, G., Schmitt, N., and Galpin, A. (2004). The eyes have it: an eye-movement study into the processing of formulaic sequences. In N. Schmitt (ed.), *Formulaic sequences: acquisition, processing, and use* (pp. 153–72). Philadelphia, PA: John Benjamins.

Unsworth, S. (2005). Child L2, adult L2, child L1: differences and similarities. A study on the acquisition of direct object scrambling in Dutch. Unpublished doctoral dissertation, Utrecht University, The Netherlands.

Ure, J. (1971). Lexical density and register differentiation. In G. E. Perren and J. L. M. Trim (eds), *Applications of linguistics*. Cambridge: CUP.

Vainikka, A., and Young-Scholten, M. (1996). Gradual development of L2 phrase structure. *Second Language Research, 12*, 7–39.

van Daele, S., Housen, A., and Pierrard, M. (2007). Psycholinguistic mechanisms underlying the manifestation and development of second language fluency, accuracy and complexity. Paper presented at the Symposium on Complexity, Accuracy and Fluency in Second Language Use, Learning and Teaching, Brussels, March.

van den Broek, P., Young, M., Tzeng, Y., and Linderholm, T. (1999). The landscape model of reading: inferences and the online construction of memory representation. In S. R. Goldman and H. van Oosendorp (eds), *The construction of mental representations during reading* (pp. 71–98). Mahwah, NJ: Erlbaum.

van Ek, J. A. (1990). *Waystage English*. London: Pergamon Press.

van Ek, J. A., and Trim, J. L. M. (1990). *Threshold 1990*. Strasbourg: Council of Europe Publishing.

Van Hout, R., and Vermeer, A. (2007). Comparing measures of lexical richness. In H. Daller, J. Milton and J. Treffers-Daller (eds), *Modelling and assessing vocabulary knowledge* (pp. 93–115). Cambridge: CUP.

VARGA (Vocabulary Acquisition Research Group Archive). Accessed 25 October 2008, at http://www.lognostics.co.uk/varga/

Vassiliu, P. (2001). Lexical input and uptake in the low level EFL classroom. Unpublished doctoral dissertation, University of Swansea, Wales.

Vaughn-Shavo, F. (1990). Using story grammar and language experience for improving recall and comprehension in the teaching of ESL to Spanish dominant first-graders. Unpublished doctoral dissertation, Hofstra University, Hempstead, NY.

Verhallen, M., and Schoonen, R. (1998). Lexical knowledge in L1 and L2 third and fifth graders. *Applied Linguistics*, *19*, 452–70.

Verhoeven, L., Narain, G., Extra, G., Konak, O. A., and Zerrouk, R. (1995). *Diagnostische toets tweetaligheid* [Test for bilingual development]. Arnhem, The Netherlands: CITO.

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics, 22*, 217–34.

Vermeer, A. (2004). The relation between lexical richness and vocabulary size in Dutch L1 and L2 children. In P. Bogaards and B. Laufer (eds), *Vocabulary in a second language* (pp. 173–89). Philadelphia, PA: John Benjamins.

Vertovec, S. (2006). The emergence of super-diversity in Britain (Working Paper No. 25). Oxford: Centre for Migration, Policy and Society, University of Oxford. Retrieved 25 October 2008, from www.compas.ox.ac.uk/publications/papers/Steven%20Vertovec%20WP0625.pdf

Walsh, K., Glaser, D., and Wilcox, D. D. (2006). *What education schools aren't teaching about reading and what elementary teachers aren't learning*. Washington, DC: National Council on Teacher Quality.

Wasik, B., and Bond, M. (2001). Beyond the pages of a book: interactive book reading and language development in preschool classrooms. *Journal of Educational Psychology, 93*, 243–50.

Wasik, B., Bond, M., and Hindman, A. (2006). The effects of a language and literacy intervention on Head Start children and teachers. *Journal of Educational Psychology, 98*, 63–74.

Watson, R., and Olson, D.R. (1987). From meaning to definition: a literate bias on the structure of word meaning. In R. Horowitz and S.J. Samuels (eds), *Comprehending oral and written language* (pp. 329–53). New York: Academic Press.

Weekes, B. S., Hamilton, S., Oakhill, J. V., and Holliday, R. E. (2008). False recollection in children with reading comprehension difficulties. *Cognition, 106*, 222–33.

Wei, L. (1994). *Three generations, two languages, one family*. Clevedon: Multilingual Matters.

Weizman, Z. O., and Snow, C. E. (2001). Lexical output as related to children's vocabulary acquisition: effects of sophisticated exposure and support for meaning. *Developmental Psychology, 37*, 265–79.

West, M. (1953). *A general service list of English words*. London: Longman.

White, L. (2003). *Second language acquisition and Universal Grammar*. Cambridge: CUP.

Whitehurst, G. J., and Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development, 69*, 848–72.

Whitehurst, G. J., and Lonigan, C. J. (2001). Emergent literacy: development from prereaders to readers. In S. B. Neuman and D. K. Dickinson (eds), *Handbook of early literacy research* (pp. 11–29). New York: Guilford Press.

Widdowson, H. (1989). Knowledge of language and ability for use. *Applied Linguistics, 10*, 128–37.

Wilcox-Herzog, A., and Kontos, S. (1998). The nature of teacher talk in early childhood classrooms and its relationship to children's play with objects and peers. *Journal of Genetic Psychology, 159*, 30–44.

Wingate, U. (2002). *The effectiveness of different learner dictionaries*. Tübingen: Niemeyer.

Wolf, M. K., Crosson, A. C., and Resnick, L. B. (2004). Classroom talk for rigorous reading comprehension instruction. *Reading Psychology, an International Quarterly, 26*, 27–53.

Wolf, M. K., Crosson, A. C., and Resnick, L. B. (2006). *Accountable talk in reading comprehension instruction* (CSE Technical Report 670). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Wray, A. (1999). Formulaic language in learners and native speakers. *Language Teaching, 32*, 213–31.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: CUP.

Wray, A., and Namba, K. (2003). Formulaic language in a Japanese–English bilingual child: a practical approach to data analysis. *Japan Journal for Multilingualism and Multiculturalism, 9*, 24–51.

Wray, A., Mollet, E., Fitzpatrick, T., Wright, M., and Wray, N. (in preparation). Towards comprehensive profiles of written language for comparative analysis: choosing the best tools (working title).

Wright, S. (2004). *Language policy and language planning: from nationalism to globalization*. Basingstoke: Palgrave Macmillan.

Yule, G., and Hoffman, P. (1990). Predicting success for international teaching assistants in a US university. *TESOL Quarterly*, *24*, 227–343.

Zill, N., and Resnick, G. (2006). Emergent literacy of low-income children in Head Start: relationships with child and family characteristics, program factors, and classroom quality. In D. K. Dickinson and S. B. Neuman (eds), *Handbook of early literacy research* (Vol. 2) (pp. 347–71). New York: Guilford.

Index