Cambridge Studies in Linguistics 144

Elementary Syntactic Structures

Prospects of a Feature-Free Syntax

Cedric Boeckx

ELEMENTARY SYNTACTIC STRUCTURES

Most syntacticians, no matter their theoretical persuasion, agree that features (types or categories) are the most important units of analysis. Within Chomskyan generative grammar, the importance of features has grown steadily, and within minimalism, it can be said that everything depends on features. They are obstacles in any interdisciplinary investigation concerning the nature of language, and it is hard to imagine a syntactic description that does not explore them.

For the first time, this book turns grammar upside down and proposes a new model of syntax which is better suited to interdisciplinary interactions, and shows how syntax can proceed free of lexical influence. The empirical domain examined is vast, and all the fundamental units and properties of syntax (categories, Parameters, Last Resort, labeling, and hierarchies) are rethought.

Opening up new avenues of investigation, this book will be invaluable to researchers and students of syntactic theory, and linguistics more broadly.

CEDRIC BOECKX is Research Professor at ICREA (Catalan Institute for Advanced Studies) and a member of the Department of Linguistics at the University of Barcelona.

111. SAMUEL D. EPSTEIN and T. DANIEL SEELY: *Derivations in minimalism*
112. PAUL DE LACY: *Markedness: reduction and preservation in phonology*
113. YEHUDA N. FALK: *Subjects and their properties*
114. P. H. MATTHEWS: *Syntactic relations: a critical survey*
115. MARK C. BAKER: *The syntax of agreement and concord*
116. GILLIAN CATRIONA RAMCHAND: *Verb meaning and the lexicon: a first phase syntax*
117. PIETER MUYSKEN: *Functional categories*
118. JUAN URIAGEREKA: *Syntactic anchors: on semantic structuring*
119. D. ROBERT LADD: *Intonational phonology second edition*
120. LEONARD H. BABBY: *The syntax of argument structure*
121. B. ELAN DRESHER: *The contrastive hierarchy in phonology*
122. DAVID ADGER, DANIEL HARBOUR and LAUREL J. WATKINS: *Mirrors and microparameters: phrase structure beyond free word order*
123. NIINA NING ZHANG: *Coordination in syntax*
124. NEIL SMITH: *Acquiring phonology*
125. NINA TOPINTZI: *Onsets: suprasegmental and prosodic behaviour*
126. CEDRIC BOECKX, NORBERT HORNSTEIN and JAIRO NUNES: *Control as movement*
127. MICHAEL ISRAEL: *The grammar of polarity: pragmatics, sensitivity, and the logic of scales*
128. M. RITA MANZINI and LEONARDO M. SAVOIA: *Grammatical categories: variation in Romance languages*
129. BARBARA CITKO: *Symmetry in syntax: merge, move and labels*
130. RACHEL WALKER: *Vowel patterns in language*
131. MARY DALRYMPLE and IRINA NIKOLAEVA: *Objects and information structure*
132. JERROLD M. SADOCK: *The modular architecture of grammar*
133. DUNSTAN BROWN and ANDREW HIPPISLEY: *Network morphology: a defaults-based theory of word structure*
134. BETTELOU LOS, CORRIEN BLOM, GEERT BOOIJ, MARION ELENBAAS and ANS VAN KEMENADE: *Morphosyntactic change: a comparative study of particles and prefixes*
135. STEPHEN CRAIN: *The emergence of meaning*
136. HUBERT HAIDER: *Symmetry breaking in syntax*
137. JOSÉ A. CAMACHO: *Null subjects*
138. GREGORY STUMP and RAPHAEL A. FINKEL: *Morphological typology: from word to paradigm*
139. BRUCE TESAR: *Output-driven phonology: theory and learning*
140. ASIER ALCÁZAR and MARIO SALTARELLI: *The syntax of imperatives*
141. MISHA BECKER: *The acquisition of syntactic structure: animacy and thematic alignment*
142. MARTINA WILTSCHKO: *The universal structure of categories: towards a formal typology*
143. FAHAD RASHED AL-MUTAIRI: *The minimalist program: the nature and plausibility of Chomsky's biolinguistics*
144. CEDRIC BOECKX: *Elementary syntactic structures: prospects of a feature-free syntax*

Earlier issues not listed are also available

CAMBRIDGE STUDIES IN LINGUISTICS

General Editors: P. AUSTIN, J. BRESNAN, B. COMRIE,
S. CRAIN, W. DRESSLER, C. J. EWEN, R. LASS,
D. LIGHTFOOT, K. RICE, I. ROBERTS, S. ROMAINE,
N. V. SMITH

ELEMENTARY SYNTACTIC STRUCTURES
Prospects of a Feature-Free Syntax

ELEMENTARY SYNTACTIC STRUCTURES

PROSPECTS OF A FEATURE-FREE SYNTAX

CEDRIC BOECKX

ICREA/University of Barcelona

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107034099

© Cedric Boeckx 2015

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2015

Printed in the United Kingdom by Clays, St Ives plc

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Boeckx, Cedric, author.
Elementary syntactic structures : prospects of a feature-free syntax / Cedric Boeckx.
pages ; cm
Includes bibliographical references.
ISBN 978-1-107-03409-9 (Hardback)
1. Grammar, Comparative and general–Syntax. 2. Minimalist theory (Linguistics) I. Title.
P291.B634 2015
415–dc23 2014014935

ISBN 978-1-107-03409-9 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

For Youngmi,
my alpha and omega

Contents

List of figures *page* ix
Preface xi
Acknowledgements xv
Abbreviations xvii

1 Biolinguistic concerns 1
1.1 Approaching Universal Grammar in medias res 1
1.2 Infrequently asked questions 4
1.3 Interdisciplinary measures 16

2 Syntactic order for free: Merge α 26
2.1 Bare Merge 26
2.2 Lexical precursor cells 27
2.3 Regulating Merge α: Santa Fe-Style Syntax 35
2.4 More emergent asymmetries 62
2.5 Additional remarks on the present model 79
2.6 Conclusion: syntax as grammatical pre-pattern 88

3 Trusting in the external systems: descent with modification 96
3.1 Opening remarks 96
3.2 Adding content, or making sense of syntax 99
3.3 Making content public, or how to put syntax into words 109
3.4 Conclusion 118

4 Elaborate grammatical structures: how (and where) to deal with variation 119
4.1 Parameters: caught between GB and minimalism 119
4.2 If not Parameters, then what? 140
4.3 Conclusion 144

5 Interdisciplinary prospects **147**

Appendix 1: **Déjà vu all over again?** **149**

Appendix 2: **Switching metaphors: from clocks to sand piles** **157**

Appendix 3: **More on the loss of syntactic variation** **165**

References 174
Index 199

Figures

4.1 Classical parameter hierarchy *page* 122
4.2 Parametric web (entangled hierarchy) 124
4.3 Acquisition algorithm 145

Preface

At the heart of this book lies a simple and fundamental question – rarely raised – concerning the human language faculty:

(i) How much does syntactic structuring depend on lexical information?

To facilitate the investigation of this question I propose we adopt the following idealization:

(ii) Assume that all units providing the input to syntactic structure building (call these 'lexical items' for now) have a common, minimal internal organization (i.e., lexically speaking, they are indistinguishable, and internally unstructured, atomic, 'flat').

Based on (ii), how much syntax, and of what kind, can we obtain? Put another way, can syntactic structure emerge in the absence of the usual suspects, the "words" and their varying properties?

The idea of a completely homogeneous lexicon feeding the syntactic component of our language faculty is likely to strike many as untenable, but I would like to make the rationale behind it clear from the beginning. It is generally assumed that, aside from relatively narrow phenomena such as haplology or semantic incongruence, the combinatorial possibilities available in natural language depend on the specific properties (lexico-syntactic 'features') of lexical items and their internal organizations (lexico-syntactic 'feature bundles'): verbs typically combine with nouns, and not with other verbs, because verbs require arguments to meet their lexical thematic requirements, and nouns have all the lexical ingredients to qualify as arguments, while verbs lack these. This 'Swiss cheese' or 'lego' model of syntax, where lexical items reach the syntactic component with 'holes' in them (aka subcategorization frames), which other lexical items have to fill with their own specifications, is widely shared across theoretical frameworks. Gottlob Frege and many other logicians thought of semantic composition in essentially those terms as well. But I will argue in the following pages that despite its near-universal adoption, this picture of the

relationship between the lexical and the syntactic components of natural language is mistaken. I will indeed call for an inversion of reasoning and claim that syntax is completely free of lexical influence, and that it is the lexicon that depends on syntax and not the other way around. As a reviewer of this book put it, in such a model, syntax is 'feature-free.' And although we are used to thinking that without lexical instruction syntax cannot even take place, I hope to demonstrate that in fact syntactic structure building can proceed unhindered.

This fact alone seems to me to be interesting enough to warrant attention, but I want to make clear the specific reason that pushed me to question the standard relation between the lexicon and syntax. For as noteworthy as the absence of lexical influence on syntax may be, if the shape of the feature-free syntactic component does not change much from the standard, lexically influenced picture, one may well ask, "Why bother?" In fact, behind the technical discussion of many of the pages of this book lies a more foundational consideration, which one might call the 'biolinguistic imperative.' The biolinguistic imperative has both a pragmatic and a more substantive side to it, which I will seek to clarify. By 'biolinguistics' I intend to refer to the general discipline aiming at uncovering what Eric Lenneberg called the biological foundations of our species-specific ability to develop a language. It stands to reason that this aim requires genuine and sustained interdisciplinary collaboration. Such an effort is doomed to fail if (among other things) linguists keep insisting on relying on fundamental properties that are (by definition) so domain-specific as to make interdisciplinary dialog break down. The pragmatic side of the biolinguistic imperative is to do everything one can (without, of course, sacrificing description and explanation) to facilitate cross-disciplinary interactions. In my experience, there is nothing as damaging as mentioning modular, informationally encapsulated features in the middle of an interdisciplinary exchange, and insisting on their traditional 'driver' role.

The more substantive side of the biolinguistic imperative derives from the fact that biological reflections on evolutionary novelties – and the human language faculty surely is one – strongly suggest that innovations do not find their origin in the appearance of *de novo* material. Innovations arise through recombinations of 'old' material, which sometimes give rise to unpredicted, 'emergent' properties, but which should never be mistaken for structures that require radically specific sources. It is generally agreed upon that within language syntax is the most innovative aspect, but I fear that the standard treatment of syntax in terms of something as domain-specific as lexical instructions moves us away from biologically plausible scenarios concerning not only the emergence of syntax, but also its implementation in the brain. In other words,

the lexicon as standardly conceived of prevents linguists from meeting their biology-oriented goals.

The main task of the first part of this book is to expand on the preceding paragraphs and show in detail how syntax is standardly taken to depend on lexical instruction (a theoretical position I dub 'lexicocentrism'), and how this has made it difficult, not to say impossible, for linguistically informed biolinguistic investigations to flourish.

The aim of the second part of the book is to contrast this state of affairs with the promises offered by a decidedly anti-lexicalist alternative for biolinguistics. This last qualification is important, for the 'prospects' referred to in the subtitle of the book are meant to be confined to the enterprise aiming at revealing the biological roots of the human language faculty. This is the descriptive and explanatory scope for the proposal to be developed in the forthcoming chapters. Crucially, the goal is not to account for the detailed properties of the grammars of specific languages – a valuable endeavor in its own right, but quite distinct, or so I will argue, from biolinguistic inquiry. To highlight this difference, I have entitled this book '*Elementary* Syntactic Structures,' which will be the target of investigation here, as opposed to what one might call '*elaborate* grammatical structures,' which are the focus of more philology-oriented approaches.

To articulate the argument just outlined at a reasonable level of explicitness, I have chosen to contrast the model I wish to put forth with the so-called Principles & Parameters (P&P) framework, with special emphasis on its minimalist articulation. There are several reasons why the P&P approach is a desirable foil to exploit. First, it has been developed in detail over thirty years, and as such it offers a rich body of doctrines to wrestle with. Second, the framework will be familiar to many readers, having been taken as point of departure for many popular textbooks in the field. Third, this is the model that I know best, which enables me to illustrate certain claims more readily than with other frameworks. And, fourth, the P&P approach is often portrayed as being concerned with problems of language development and design, oriented towards the sort of biological considerations that I am interested in. It therefore constitutes an ideal testing ground for my claim that lexicocentrism and biolinguistics are bound to be uneasy bedfellows.

Some readers may well wonder if the argument in the pages that follow is to be considered minimalist. This is a somewhat tricky – and from where I sit, rather minor – question. I certainly believe that the core proposal in this book converges with certain lines of minimalist investigation (especially the idea of 'approaching Universal Grammar from below'), but it also departs rather

sharply from standard minimalism, which is lexicocentric through and through. I should say that ultimately whether the approach pursued here is deemed minimalist or not is not a major concern of mine. What matters most is that it advances the biolinguistic enterprise.

Let me take this opportunity to let all the readers of this preface know that this is primarily a book about the syntactic component of the language faculty. Although I very much hope that scholars whose fields of expertise lie outside this particular empirical domain will read on and appreciate the arguments I make, perhaps even draw conclusions for their own research programs, I realize that some chapters, especially Chapter 2, will be demanding. Even if I have tried to make the discussion accessible to as many potential readers as I could think of, at times I have not been able to avoid getting down to the technical details. Inevitably, this has sometimes led me to add notes that ended up longer than one might wish. Because the central line of argument pursued in this book touches on many issues, I have also decided to relegate some material to appendices. A reviewer of this book urged me to incorporate all that material in the main text, as, according to him/her, readers tend not to read appendices. This would be a shame, as I think that the material discussed in the three appendices included here bears on central themes of this book. But because I could not find a satisfactory way of putting it all in the body of the book while at the same time keeping the main narrative as straightforward as I could, I am left to kindly ask readers not to ignore the considerations relegated to the periphery. Last, but not least, readers will find very few natural language examples in the pages that follow. It's not because I have a disdain for data, but, as I point out later on, the anti-lexicocentric stance I take forces me to focus on more elementary considerations than specific sentences in one language or another. As I wrote at the beginning of this preface, I am asking readers to study the language faculty before words come into the picture. Syntax is there.

Acknowledgements

The present book is the result of two converging lines of investigation I embarked on several years ago – the first one focused on the nature of the putatively primitive status of the combinatorial operation known as "Merge" (see, e.g., Boeckx (2009e)) and the second, on the nature of linguistic variation and the adequacy of parametric proposals (Boeckx (2011a, 2014c, forthcoming)). But the convergence did not crystallize into a coherent line of attack on lexicocentrism until I spent long hours talking to my former student Dennis Ott at Harvard. Conversations with my other students there, Hiroki Narita and Bridget Samuels, were also of critical importance at this early stage of 'book thinking.' The encouraging results of their own works also gave me confidence that the general approach I intended to pursue had at least a fighting chance of being on the right track, or at least that it was moving in an interesting direction. Equally important was my interaction with Paul Pietroski, who provided me with key insights from semantics that I have tried to take advantage of here, though I am sure insufficiently so. Marc Richards' work on phases also played an important role in making me reconsider the role of phases in grammar. A draft of his (2011) paper was very inspirational, as was his Ph.D. dissertation (2004) and his engagement with Boeckx and Grohmann (2007).

I also want to acknowledge my indebtedness to Dick Lewontin, Marc Hauser, Koji Fujita, Alec Marantz, Ray Jackendoff, Fritz Newmeyer, Juan Uriagereka, and, of course, Noam Chomsky, for many important discussions over many years. Massimo Piattelli-Palmarini deserves special mention for organizing interdisciplinary meetings where I've felt more at ease than in strictly linguistic gatherings, and where I've invariably learned a lot. Thanks, too, to Anna Maria Di Sciullo for counting on me in the context of her Biolinguistics International Network, and to all the members of the Biolinguistics Initiative Barcelona, especially my students Evelina Leivada and Pedro Tiago Martins.

The seed of the present work was planted as a talk delivered at the Program from Evolutionary Dynamics at Harvard University, led by Martin Nowak, in 2004. The talk was intended to be a reflection on the nature and role of Parameters in current linguistic theory (of a minimalist orientation), aimed at an audience consisting primarily of non-linguists. My focus back then was Mark Baker's parameter hierarchy, a concept which Baker had succeeded in making accessible to non-linguists in his *Atoms of Language* (2001). Since it dealt with the nature of variation at an accessible level, this was a perfect topic for biologists and other non-linguists in the audience. At the time I did not realize that this event would be the beginning of my questioning many of the key assumptions routinely made in syntactic theory (including in my own work), not only in the context of Parameters (variation), but, I later came to realize, also in the context of principles (invariance). Thanks to Martin for giving me this opportunity.

In writing this book I've come to realize in retrospect that the argument presented here is nothing more than my modest attempt to come to grips with Carlos Otero's assessment of the legacy of generative grammar ("we haven't yet appreciated Noam's revolution fully"), which he shared with me back in 2006. At the time I did not know my research would take me where I think Carlos had already gone a long time ago (see Appendix 1).

Several portions of the book have been presented at numerous conferences and workshops, too numerous to list individually here, so let me collectively thank all the organizers of these venues for the valuable opportunities they have offered me to present work in progress to audiences that never failed to respond constructively to what I had to say. Thanks, too, to two anonymous readers for Cambridge University Press, and to Helen Barton for her faith in this project.

Since I moved back to Europe I have come to realize the importance of funding agencies to sustain the sort of research I wanted to pursue, and it is with great pleasure that I acknowledge the unconditional support of my employer, the Catalan Institute for Advanced Studies (ICREA), as well as grants from the European Union (Marie Curie International Reintegration Grant PIRG-GA-2009-256413), from the Spanish Ministry of Economy and Competitiveness (FFI-2010-20634), and research funds from the Fundació Bosch i Gimpera.

Last, but not least, words of gratitude fall far short of what I'd like to be able to express towards my wife Youngmi, to whom this book is dedicated. All I can say is that I certainly don't deserve as much as she has given me, but she undoubtedly deserves much better. I love her deeply, and I'm very sorry for all the trials that this thing called academia gives rise to.

Abbreviations

C	Complementizer
CED	Condition on Extraction Domain
C-I	Conceptual–Intentional
CP	Complementizer Phrase
D	Determiner
DM	Distributed Morphology
DP	Determiner Phrase
EPP	Extended Projection Principle
FL	Faculty of Language
FLB	Faculty of Language – Broad Sense
FLN	Faculty of Language – Narrow Sense
GB	Government-and-Binding
HPSG	Head-Driven Phrase Structure Grammar
LCA	Linear Correspondence Axiom
LF	Logical Form
LFG	Lexical–Functional Grammar
LI	Lexical Item
MP	Minimalist Program
P&P	Principles-and-Parameters [framework]
PF	Phonological Form
PHON	Phonological Component
PIC	Phase Impenetrability Condition
QP	Quantifier Phrase
SEM	Semantic Component
S-M	Sensory-Motor
Spec	Specifier
T	Tense

TP Tense Phrase
UG Universal Grammar
V Verb
VP Verb Phrase

1 *Biolinguistic concerns*

1.1 Approaching Universal Grammar *in medias res*

The road leading theoretical linguistics beyond explanatory adequacy,[1] that is, towards a naturalistic, biologically grounded, better-integrated cognitive science of the language faculty – i.e. a flourishing *biolinguistics* – is chock full of obstacles. One can distinguish between external and internal obstacles. External obstacles are very familiar to the student of modern cognitive science. They are the remnant of behaviorist proclivities, the result of our seemingly innate bias towards dualism, and of our traumatic encounter with grammatical prescriptivism, to say nothing of our extrapolations based on our failed attempt to master another language as an adult. All of these factors invariably tend to keep us away from the biological nature of the language faculty, making us believe that this thing called language is a cultural invention you've got to learn painstakingly, full of non-sensical arbitrariness, nothing like the language next door. Old habits really do die hard. Although I think that compelling arguments can be made (and have been made) against overly empiricist, cultural views of language, these views are part of our nature, and one has to be aware of them (and keep them in check) at all times when attempting to delineate the neurobiology of the language faculty. Internal obstacles are more difficult to deal with, for those are habits that were adopted early during the practice of

[1] Chomsky defines "beyond explanatory adequacy" thus: "we can seek a level of explanation deeper than explanatory adequacy, asking not only *what* the properties of language are but *why* they are that way" (2004, 105). Much earlier, Chomsky (1965, 63) defined explanatory adequacy by contrasting it with descriptive adequacy, as follows:

> a grammar that aims for descriptive adequacy is concerned to give a correct account of the linguistic intuition of the native speaker; in other words, it is concerned with the output of the device; a linguistic theory that aims for explanatory adequacy is concerned with the internal structure of the device; that is, it aims to provide a principled basis independent of any particular language, for the selection of the descriptively adequate grammar of each language.

linguistics-as-cognitive-science, that rendered initial progress possible, but that can subsequently prove an impediment to further progress. In this contribution I want to examine one such factor, which I will refer to as 'lexicocentrism.'

By 'lexicocentrism' I mean the theoretical stance, shared across many frameworks, that appeals to the 'lexicon' (I put the term in scare quotes because of the many definitions it has received in the literature; for my purposes here, understand lexicon as the repository of elements on which syntax feeds) to account for most, if not all of what many would regard as core properties of the language faculty (detailed illustrations of this stance will follow momentarily, and will be found throughout the book). If one is interested in what Saussure called the arbitrariness of the sign, appeal to the lexicon is, of course, in order. It is a brute lexical fact that Catalan speakers say *gos* to refer to DOG, but French speakers say *chien* to refer to the same thing. But if one is interested in more grammatical facts, such as the ability for a noun to combine with a verb, or even, I will suggest, the very existence of categories like 'noun' and 'verb,' or in the patterns of cross-linguistic variation, then lexicocentric accounts retain their arbitrary character, and leave unanswered – indeed, they often make it hard to ask – certain questions that are well worth reflecting upon.

Take, for example, the currently standard treatment of displacement in transformational generative grammar. In situations like *Who did Mary kiss?*, we learn that *who* appears pronounced where it is (say, SpecCP) because the relevant functional head (C) in English has a particular lexical need that can only be satisfied through the presence of a lexical item with the appropriate lexical specification ([wh]-feature) in its specifier. Such an account is then generalized to the following condition ('principle'): displacement takes place only to satisfy the lexical demands of the host.[2] This in turn leads researchers to posit lexical properties on functional heads (e.g., "EPP" features) just because an element appears to have been displaced in their vicinity. I am certainly not the first to have noticed the arbitrary character of this kind of explanation. Perhaps no one said it as well as George Lakoff in the following passage, quoted in Kibort and Corbett (2010, 31):

> So linguists fudge, just as has been done in the reflexive rule, by sticking on the arbitrary feature +REFL. Such a feature is a fudge. It might as well be called +CHOCOLATE, which would in fact be a better name, since it would clearly reveal the nature of the fudge.

[2] In the minimalist literature, this is known as the "Attract" principle (Chomsky (1995, chap. 4)), or sometimes also "Suicidal Greed" (Chomsky (2000a)). Alternative accounts of movement ("Enlightened Self-Interest" (Lasnik (1999)) or "Greed" (Chomsky (1993))) proposed in the literature are equally lexicocentric.

In the early 1980s, Muysken and van Riemsdijk (1986, preface) correctly pointed out that

> Even basic questions [concerning features] such as 'how many are there?', 'what are they?', 'how do they distribute over syntactic structures?' were hardly addressed, let alone answered. Nevertheless, it is clear that syntactic features do play an important role in syntax . . . It would appear to be high time, therefore, to examine the theory of syntactic features in a more systematic way.

In their introduction to the volume, Muysken and van Riemsdijk write that "too little is known about [features]." More than twenty-five years later, similar remarks still apply. In their volume on "features as key notions in linguistics," Kibort and Corbett (2010) write that "the impact of features has increased steadily . . . features [are] essential to how we do linguistics" (p. 2), but quickly add that "there is much more to be understood about features" (p. 3). Strikingly, though, what one repeatedly finds in the literature is an attempt to replace one obviously stipulative lexicocentric account with another, perhaps at first less obviously stipulative, but equally lexicocentric account (see, for instance, Fukui and Speas' early (1986) treatment of the "Extended Projection Principle (EPP)" in terms of structural case, or the opposite suggestion in Marantz (1991)).

Readers may well wonder why lexicocentrism has dominated linguistic theorizing for so long. I think that part of the answer lies in the fact that for all its explanatory limitation lexicocentrism has revealed important grammatical facts and patterns and has made it possible to formulate interesting generalizations that (in the words of Eric Reuland) are "too good to be false." We all have to start somewhere. But I think there is another reason why lexicocentrism has maintained such a strong foothold in theoretical linguistics. For all the emphasis on the biological substrate of the language capacity in Chomsky's writings since the beginning of the so-called cognitive revolution of the mid 1950s, most theoretical linguists, even those of a Chomskyan persuasion, remain fundamentally interested in languages as objects of study,[3] and for languages it makes sense to start with the lexicon since this is clearly the most

[3] As evidence for this assertion, consider the following samples of replies provided by alumni/visitors of the MIT linguistics program on the occasion of its 50th anniversary, when they were asked "What was the broad question that you most wanted to get an answer to during your time in the program?" (http://ling50.mit.edu/category/replies):

> Since the beginning of my student career in the early 1970s I had been fascinated with the issue of how aspects of grammatical diversity cluster across languages (G. Longobardi).

distinctive thing about them. But what if the focus of inquiry is the language faculty (that which makes the acquisition of specific languages possible)? Here I think the dominance of lexicocentrism is down to the weight of tradition, of thinking of Universal Grammar as a grammar, organized like the traditional grammars of specific languages, where syntax is thought to be the study of how words are put together. Of course, if that is what syntax is, it makes eminent sense to start with the words, just like in semantics, one starts with lexical meaning to capture compositionality. As Caponigro and Polinsky (2011) state matter-of-factly, "the lexicon of a language is expected to shape its syntax."

1.2 Infrequently asked questions

Most linguists, I suspect, would endorse Polinsky and Caponigro's statement, certainly those pursuing a standard minimalist approach. This is in fact the view enshrined in most textbooks (see, e.g., Adger (2003); Hornstein *et al.* (2006)), the view that lies behind such notions as "Last Resort" and "triggered Merge" in current minimalism, and that makes it possible to claim that "labels can be eliminated" (Collins (2002)),[4] that "syntax is crash-proof" (i.e., driven

> I was preoccupied to know what should be the correct relationship between linguistic theory and language description (A. Salanova)
>
> As a student who had been strongly attracted by grammars of L (= English, Latin, German, Greek, French) and holder of a mathematics MA, what attracted me to the MIT program, via Chomsky's writings, was the sense that at least preliminary explicit formulations of these grammars of *L* were in sight—not during my stay at MIT, but in say a couple of decades.
>
> With almost everyone else, I was convinced from the first of '... the necessity for supplementing a "particular grammar" by a universal grammar if it is to achieve descriptive adequacy.' (Aspects of the Theory of Syntax: 6). Thus, I understood,
>
> (1) Grammar of L = UG + G_1 (= a Particular Grammar of L_1)
> These grammars, supplemented by UG, were to generate all and only grammatical sequences of the L_1. So, the broad question had two parts: what was UG, perhaps the hardest part, and what were the (formalized, explicit) Particular Grammars, a supposedly easier question. Nonetheless, the second part also seemed intriguing and puzzling, since, beyond some generalities, exact aspects of e.g. English and French grammars had little in common. (Kayne's dissertation, his later *French Syntax*, didn't seem to be a book about English grammar.) Thus in addition to UG, "the broad question for which I most wanted to get an answer to" was:
>
> (2) What exactly is the form of particular grammars that UG can then 'supplement'? [J. Emonds]

Felix (2010) contains a deeply insightful discussion of how these goals differ from Chomsky's (biolinguistic) motivation, and of the unfortunate consequences this can lead to.

[4] Collins's work on labels is often misunderstood, I think. He crucially did not argue for the elimination of labeling or headedness, but rather for a representational change: a replacement

by featural needs of 'pivots'; Frampton and Gutmann (2002)), and also the one that facilitates the endorsement of the "Borer–Chomsky Conjecture" concerning parametric variation (the idea that all variation reduces to lexical choices). It is indeed *the* cornerstone on which our modern conception of Principles-and-Parameters rests. A quick examination of other frameworks reveals that it is also an aspect of language design that is widely shared. (Accordingly, critics of minimalism, of whom there are quite a few, would be ill-advised to use what follows as evidence for the superiority of their own view of syntax.) And yet, I will argue that lexicocentrism is wrong. Deeply wrong. In fact, it may be the biggest (internal) obstacle that lies on the road towards a level of explanation that Chomsky has referred to as "beyond explanatory adequacy."

A common criticism of minimalist syntax is that it simplifies syntax by dumping everything it cannot deal with or does not like onto the external systems with which it interfaces. But I think that minimalist syntacticians commit an even bigger mistake – one that is rarely if ever highlighted (perhaps because it's shared across frameworks and also because it's so deeply intuitive) – by coding virtually everything they should explain as lexical traits, better known as features. Although it is true that minimalist syntacticians relegate a lot of standard syntactic phenomena to post-syntactic components, I do not think that this is necessarily a bad thing, given that we are finally coming to terms with the fact that these systems have powerful resources (see Hauser *et al.* (2002) on "the Faculty of Language in the Broad sense (FLB)"). I think that a lot of what makes minimalist analyses unconvincing, and certainly what makes them fall short of going beyond explanatory adequacy, is that by the time such analyses begin, all the action has already taken place, as it were. It has been carefully pre-packaged (pre-merged) into lexical entries. And once in the lexicon, it's taken for granted. It's not derived, it's not constructed. It is simply assumed as a matter of virtual conceptual necessity. But I take it that Epstein and Seely are right when they say that "if you have not 'grown' [i.e., derived, constructed] it, you have not explained it" (2006, 7). Instead of "approaching syntax (and UG) from below" (Chomsky (2007)), minimalist syntacticians approach it from the lexicon, *in medias res*, and as such they do not depart at all from pre-minimalist practice (or, for that matter, from the practice of traditional grammarians).

of standardly labeled nodes by a lexically/featurally defined set of asymmetric prominence relations. See Adger (2013b) for an unusually clear statement of Collins's approach. See also Boeckx (2008b).

Newmeyer (2004, 226 n.10) is certainly right when he points out that the lexicon is all-important in the minimalist program (MP):

> [I]n no framework ever proposed by Chomsky has the lexicon been so important as it is in the MP. Yet in no framework by Chomsky have the properties of the lexicon been as poorly investigated.

But I do not agree that the lexicon is more important in minimalism than before. It may be more conspicuous, due to the constant appeal to lexical features, but the lexicon has virtually always been seen as central. The problem is that if minimalism is to genuinely seek to move beyond explanatory adequacy (i.e., if minimalism is to do what makes it worth doing), it will have to explain, as opposed to encode, most of the properties that it now assumes as "given by the lexicon." It will have to break free of a long tradition of linguistic practice. In this sense, minimalism has so far failed to distinguish itself from previous transformational accounts, which relied on a principle made explicit in the Government-and-Binding era: the Projection Principle. As is clear from popular textbooks such as Haegeman (1994), "the projection principle: i.e. the idea that all syntactic structure is projected from the lexicon" was taken as "a basis for the organization" of the grammar. It is the central dogma regarding the flow of information in the grammar. Whereas minimalist syntacticians insisted on the demise of "government" as proof of the difference between minimalism and the models preceding it (see, e.g., Bošković and Lasnik (2007)),[5] they remained extremely conservative when it comes to the relationship between syntax and the lexicon (arguably, a more fundamental architectural property of grammar than 'government').

It is clear that minimalism suffers from featuritis (to borrow a term from computer science that nicely conveys the ad hoc character of feature-creation), and often syntacticians hide away all the interesting problems by convincing themselves that (as the saying goes) it's not a bug (an imperfection), it's a feature. These days, we have features for everything: structure-building features/merge-features (aka edge-features), agree-features (aka unvalued/uninterpretable features), move features (aka EPP-features), to say nothing of all the fine-grained featural distinctions ('flavors') brought about by the intensive cartographic projects that currently dominate syntactic inquiry. The problem is clear: in the absence of any realistic, grounded, cognitively

[5] In the end, the lack of appeal to "government" turned out to be a relatively superficial difference, given that the Agree-relation (Chomsky (2000a) is quite similar to it.

sound, biologically plausible theory of what counts as a possible feature, it is too easy to come up with a feature that will do the job. But it should be clear that features and the way we manipulate them syntactically are the problem, not the solution. It's where the investigation should end, not where it should start. As Šimík (2011) correctly states

> It is commonly assumed that formal features on syntactic categories are essentially descriptive devices—remnants of construction-based approaches to grammar. They are often introduced into the grammatical model in order to bridge the gap between the empirical facts we face and the assumptions which we believe to follow from independently motivated principles. In that respect, the postulated inventory and properties of formal features provide a useful overview of what we do not understand.

(In their volume on features, Kibort and Corbett (2010) seem to share Šimík's view when they write that "features are fundamental to linguistic description" (p. 1), but then puzzlingly add that "linguists frequently turn to them as they try to understand . . . the complexity of natural language." I say puzzlingly because I agree with Šimík that features obscure understanding, they label our ignorance.)

The problem with lexicocentrism is in fact even more severe once we realize that the basic units manipulated by syntax (lexical items/categories) are defined not as single features but as "feature-bundles" (see, among many others, Sprouse and Lau (2013): "we believe it is fair to say that there is some degree of consensus that the basic units are bundles of features"). Chomsky (2007, 6) makes it very explicit in the following passage: "In addition to Merge applicable without bounds, UG must at least provide atomic elements, lexical items LI, each a *structured* array of properties (features) to which Merge and other operations apply to form expressions" (my emphasis).[6] Such bundles are nothing more than little syntactic trees. How such treelets are constructed is left unaddressed, and in fact asking the question quickly leads to a paradox: if such treelets are built by merge, much like the regular syntactic trees that they resemble so much, why is merge at the sentential level said to require featural triggers, but merge at the lexical level is not? As we will see in detail shortly, featural triggers are nothing other than feature bundles, but then we are stuck

[6] The idea that syntactic features are internally organized is far from new or exclusive to minimalism. Although the references just given in the text are recent ones, many syntacticians adopted this idea long ago. As Muysken and van Riemsdijk (1986, 19) pointed out, "eventually, one . . . expect[s] there to be a full-fledged theory in which features are grouped into hierarchically-ordered classes and subclasses, like in phonology."

in the following loop: merge requires feature bundles to apply, and feature bundles require merge to exist! One could, of course, appeal to a new operation to construct feature bundles – call it Bundle – but, all else equal (I return to this issue in Chapter 3), this would duplicate entities in a way that flatly violates minimalism's favorite tool, Occam's razor. If Bundle constructs structures that look like those constructed by Merge, if Bundle swims like Merge, and quacks like Merge, then Bundle is Merge. The more so, if Merge operates on features, as Chomsky compellingly argued for in Chomsky (1995) when he introduced the notion of Feature-movement.

Tellingly, although the argument for moving 'just' features was a rather bold one, Chomsky did not go all the way. Although he recognized that moving just F (F a feature) was the most natural hypothesis within the framework he was considering, as the following passage reveals:

> So far I have kept to the standard assumption that the operation Move selects α and raises it, targeting K, where α and K are categories constructed from one or more lexical items. But on general minimalist assumptions, that is an unnatural interpretation of the operation. The underlying intuitive idea is that the operation Move is driven by morphological considerations: the requirement that some feature F must be checked. The minimalist operation, then, should raise just the feature F. (1995, 262)

Chomsky asked in the following paragraph "when F is raised to target K, why does F not raise alone . . .?" He went on to write:

> The answer should lie in a natural economy condition.
> (26) F carries along just enough material for convergence.
> The operation Move, we now assume, seeks to raise just F. Whatever "extra baggage" is required for convergence involves a kind of "generalized pied-piping" . . . For the most part—perhaps completely—it is properties of the phonological component that require such pied-piping.

But the pages following the passage just quoted make it clear that the situations Chomsky is considering here are situations where phonological (and perhaps) semantic features must raise along with syntactic features. Crucially, for purposes of the present discussion, Chomsky never in fact considered breaking the syntactic feature bundle to raise just F (F a syntactic feature). As he writes on p. 265:

> When the feature F on the lexical item LI raises without pied-piping of LI or any larger category α, as always in covert raising, does it literally raise alone or does it automatically take other formal features along with it? There are strong empirical reasons for assuming that Move F automatically carries along FF(LI), the set of formal features of LI. We therefore understand the

> operation move F in accord with (28), where FF[F] is FF(LI), F a feature of the lexical item LI.
> (28) Move F "carries along" FF[F].

Although unfortunately he does not elaborate on which "strong empirical reasons" he has in mind, Chomsky here makes it clear that he never in fact fully entertained the most natural hypothesis (Move *just* F, F a feature). Although he argued for splitting the syntactic feature bundle from the phonological and semantic feature bundles that together make up a lexical item, he never went as far as breaking the syntactic feature bundle itself.

By not doing so, Chomsky kept the nature of this bundle shrouded in mystery. In effect, bundles in minimalism retain the status of constructions in non-transformational frameworks: they are templates whose origins one is not supposed to discuss.[7]

Our blind reliance on the lexicon has had serious detrimental effects. This is nowhere as clear as in the context of the logical problem of language acquisition ("Plato's problem"). Consider the following quotes:

> Parametric variation is restricted to the lexicon. (Chomsky, 2001, 1)

> The availability of variation [is restricted] to the possibilities which are offered by one single component: the inflectional component [of the lexicon]. (Borer, 1984, 3)

[7] The only explicit passage regarding this issue that I have been able to find in Construction-friendly approaches is the following, from Jackendoff (2011, 602), who, after four or five books and many articles praising the superiority of "Unify" over Merge as the central operation in the grammar (including in the very article from which this passage is drawn!), acknowledges the limitations of such an operation (one can only unify structures that have been created beforehand, but what is responsible for this?; see Boeckx and Piattelli-Palmarini (2007)):

> I should make clear that Unification alone cannot create constituent structure: it only creates a Boolean combination of pre-existing features and structures. In order to build structure, one needs a skeletal constituent structure that can be unified with two or more items. Such a skeleton is of course already richly present in cognition: the part-whole schema. One formal realization of this schema is a set {x, y} with variable elements *x* and *y* as parts. This can be unified with specific elements A and B to form the set {A, B}—in effect the output of Merge. Similarly, a linearly ordered constituent [A∧B] can be licensed by the unification of A and B with a linearly ordered schema [x∧y], which is also ubiquitous in nonlinguistic cognition. Thus the effects of Merge can be constructed from Unification and one of these schemas.
>
> One might say then that these schemas are nothing but constraint-based counterparts of Merge, and this would be partly correct.

This passage makes it clear that Unify and Merge both fail to get to the heart of the matter, since they assume the existence of preformed structures. Jackendoff hints at the possibility that such structures may have non-linguistic origins, but if that is the case, then why are we the only species that has the kind of syntax we do?

> A parameter is an instruction for a certain syntactic action expressed as a feature on a lexical item and made operative when the lexical item enters syntax as a head . . . In this conception, the size of the set of parameters is not determined by the number of principles, but by the size of the (functional) lexicon. (Rizzi, 2010) (for a similar statement, see Rizzi (2009))

Although such statements give the appearance of a very restrictive theory of language variation – indeed, they have been argued to provide such a restrictive theory – in the absence of a theory behind terms like "lexical item," "the (functional) lexicon," or its "inflectional component," they amount to little more than disguised statements of ignorance, or wishful thinking.[8] This feeling is reinforced when we consider the fact that virtually throughout the generative period, the lexicon has been taken to be "really an appendix of the grammar, a list of basic irregularities" (a conception already expressed in Chomsky (1965), and reiterated in Chomsky (1995); a conception ultimately going back to Bloomfield and the structuralists). If that is the lexicon, surely we cannot claim to have understood the nature of variation by placing it (variation) there (in the 'lexicon').

But the gap between our understanding of the lexicon and the intensive use we make of it is by no means limited to the domain of parametric variation. It is equally damaging on the other side of the 'Principles-and-Parameters' model, in the domain of principles. In minimalism, all syntactic operations are currently assumed to be feature-driven. That is to say, as Epstein (2003 [2007], 43) has correctly pointed out, "the most fundamental operation ceases to be structure-building (Merge) and becomes structure-licensing (Check/Agree/Value)." Epstein's statement makes it clear that contrary to the rhetoric often used in minimalist circles, it is not true that "all you need is Merge" (see Berwick (2011)). As Chomsky himself makes it clear in a passage already quoted above, "[i]n addition to Merge applicable without bounds, UG must at least provide atomic elements, lexical items LI, each a structured array of properties (features) to which Merge and other operations apply to form expressions" (Chomsky, 2007, 6). But saying that Merge merely "applies" to LIs does not capture the fact that Merge is subordinated to the lexicon. Short of features triggering it (the "vehicle requirement" imposed on Merge in Pesetsky and Torrego (2007)), Merge can't apply. No wonder

[8] Consider the notion of "head" in Rizzi's statement. Why do we regard bundles of features as (minimal) heads, and other collections of features as (maximal) phrases? Where is the dividing line? (The problem is of course even more severe in a Bare Phrase Structure framework like Chomsky (1995).)

Pinker and Jackendoff (2005) criticized Hauser *et al.* (2002) for being silent on the evolution of the lexicon:

> [What is t]he source and nature of lexical entries, which do considerable work in the theory (defining phrase structures, triggering movement), and which therefore are far more abstract and language-specific than mere sound–meaning pairings[?] (p. 220)

It is quite possible that deep down Pinker and Jackendoff had primarily the size of the human lexicon in mind as a fundamental attribute of the language faculty, and that for them the lexicon is more than just the input to the syntactic component (something like the set of all sound–form–meaning triplets, ranging from single morphemes to phrasal idioms). But minimalist practice, if taken seriously, indicates that, more than the size, it is the nature of the formal features making up the lexicon that was the key evolutionary event (I return to this at length below), for, if standard minimalist analyses are taken seriously, this is what made syntax possible in the first place. But, just like in the context of parametric variation, no hypothesis concerning how (that aspect of) the lexicon came about is forthcoming.

Perhaps the clearest example of how much the lexical properties dictate syntactic behavior is to be found in Adger (2010). Adger (2010) (further elaborated in Adger and Svenonius (2011)) is the first and most explicit discussion of the nature of lexical entries (pre-syntactic feature bundles) within minimalism that I know of. The paper opens with a challenge. Its goal is to "explore the consequences of the idea that structure embedding in human language is only ever syntactic (that is, that there is a single engine for the generation of structure and the engine is the syntax" (an idea familiar in recent works by Marantz, and Borer, and directly related to the proposal in Hauser, Chomsky, and Fitch (2002)). As Adger correctly observes "if structure embedding is only syntactic, then the feature structures that are the basic atoms of syntax (i.e., lexical items) cannot involve embedding of one feature inside another." In so doing, Adger notes, this minimalist approach "contrasts rather starkly with work in other approaches which take lexical items to have rich featural structure" (all feature-unification frameworks, such as HPSG, and LFG). A more accurate statement would be 'in so doing, this minimalist approach *would* contrast rather starkly with other frameworks,' for, as a matter of fact, Adger ends up being forced to propose a fair amount of embedding inside his minimalist lexical entries. Consider his final proposal concerning the shape of feature bundles:

a. A lexical item is a set $\{K, F_1, \ldots, F_n\}$,
   (i) K is an ordered pair ⟨Cat, N⟩
   (ii) Cat is drawn from the set [known as the functional sequence or Cinque-hierarchy]{C, D, T, Num, Asp, . . . }–Ø, and N is drawn from the set of natural numbers above 0
b. F_1 is a pair ⟨Att, Val⟩, where Att is drawn from a finite set of M[orpho]S[yntactic]-features and Val from a finite set of values [say, +/–]
c. Hierarchies of Projections: these are sequences of Ks whose second member is ordered by the relation $\prec$ [see, e.g., Cinque (1999)]

The above would, for example, correspond to the following concrete lexical entry:

(1) *was* = {T, #8 [on the functional sequence], tense:past, pers.:3, num:sing, case:nom}

I am sure that the reader recognizes a fair amount of structuring here: ordered pairs are nothing other than embedded structures ($\langle\alpha,\beta\rangle = \{\alpha,\{\alpha,\beta\}\}$). There is a clear sense in which an attribute stands in a subset/superset (i.e., embedded) relation. In his paper, Adger is silent about the EPP/strength property of some features. If this property is thought of in the standard way (e.g., Pesetsky and Torrego (2001), Carstens (2005)), as a feature of a feature, it will require a further enrichment of lexical entries; i.e., further embedding. This is indeed the conclusion reached in Adger and Svenonius (2011), who speak of features like EPP as "second-order features," opening the door to "recursion into the feature structure" (p. 39).

The reason for this irreducible lexico-structural complexity is clear: Adger takes as his starting point the view that "much like categorial grammars, minimalist grammars can be seen as lexically driven combinatory systems" (a point of view well articulated in his textbook, Adger (2003)). True, Adger claims that "minimalism is unlike other 'lexicalist' theories in that almost all the interesting structure is syntactic," but then quickly adds "although the information which leads to the building of that structure is entirely lexical," thus recognizing the lexicocentric character of minimalist syntax.

Adger in fact recognizes the limits of his own proposal. In the concluding section of his paper, he writes: "I should stress that the theory is certainly not as 'minimal' as one might like . . . " But I think that Adger's theory of feature structure is probably as minimal as it could be given the

set of assumptions he adopts – in particular his claim that it is "features [that] drive the various syntactic operations. We can think about this in an essentially Fregean way: the features have some property that needs to be satisfied." As an example, consider the fact that Adger takes Merge, as he does in Adger (2003), to be licensed by a matching of categorial features on the selector and the selectee, where the feature of selector is unvalued. Accordingly, for lexical entries of lexical categories, one cannot get around distinguishing between a valued categorial feature and an unvalued categorial feature.

The outline of my general conclusion should already be clear at this point: the best-case scenario for an approach that seeks to tackle UG from below (i.e., for an approach with any hope of biological adequacy) is for there to be no need for feature-bundles driving syntax, for bundles are syntactic structures, and syntactic structures demand syntax to rise. Pre-syntactic lexical entries must be flat. The objects syntax manipulates must be atomic. The pre-syntactic lexicon must be structureless.[9] It should be clear that for this to be the case, Merge must be trigger-free (i.e., completely optional), processes like Agree as a syntactic operation must be rethought, functional hierarchies must be emergent, not preformed. Various other properties must be contextual/configurational, and yet other aspects of grammar, such as variation (Chapter 3) must be outsourced completely. If all of this obtains (yes, I know, it's a big *if*), then the lexical input to Narrow Syntax can be kept to a bare minimum, and Adger's challenge (essentially a biolinguistic challenge, as it ultimately relies on evolutionary plausibility) against embedding inside lexical items can be met. I will call this approach "Merge α," for, as many will probably recognize, it is reminiscent of the spirit behind the "Move α/Affect α" that dominated Government-and-Binding approaches (see Chomsky (1977b), Chomsky and Lasnik (1977), Lasnik and Saito (1992)).

What this amounts to is a call for a Gestalt shift regarding syntax and its relation with the lexicon. It will also require a reconception of the relation between

[9] This is, in fact, the view expressed in Marantz (1996, 3): "The syntax starts with simple, non-branching constituents." A view that is, unfortunately, never really put into practice, even in post-syntactic, realizational, exoskeletal approaches such as Distributed Morphology or Borer's (2005) exoskeletal approach. As Marantz (1996, 1) points out, "The lexicon as input to the computational system is always considered generative in some sense[, e]ven within DM." Nanosyntacticians (see the papers collected in Svenonius *et al.* (2009)) also rely heavily, though implicitly, on an organized, syntactically structured pre-syntactic lexicon, as they crucially assume a language-specific functional sequence that guides syntactic structuring, even if the latter manipulates single features. Borer's exoskeletal model is also at least partially lexicocentric, as it relies on the existence of a functor lexicon, with functional elements that "project."

syntax and grammar, as we will see. The key issue was already identified by Adger in the paper I have been quoting from:

> When two items merge, is there some constraint on what they can/must be? Empirically, we clearly want to say that there is, to distinguish, for example, *the many men* from **many the men*, so the question is how this kind of distributional constraint is modelled theoretically. (p. 15)

Adger is right; constraints of this sort abound in natural languages. But how strong is the argument for a (lexico-)syntactic treatment of these constraints? The unacceptability of *many the men* is a fact about a particular state of the language faculty ("English"). Even if universal, the constraint need not necessarily be viewed as a fact about the faculty of language in the narrow sense. It could be a fact about the faculty of language in the broad sense, the result of many components, something like an interaction effect. It is true that constraints of this sort have been at the heart of cartographic projects (witness the adverb ordering restrictions discussed in Cinque (1999), or the facts discussed in Rizzi (1997)), where they are reduced to lexical (c-selectional) restrictions. But this type of reduction does not mean explanation. Saying that it is *the*P that c-commands *many*P (or DP, QP), and not the other way around, is not very different from saying that *the many men* is okay but *many the men* is not. Fortuny (2008, 112f.) is certainly correct when he writes:

> [T]he empirical results of the cartographic project do not lead us to a primitive element of the syntactic component, but rather to the study of the C-I system[s] . . . [These ordering restrictions are] clearly relative to the levels of interpretation of the C-I system[s], and hence one may be skeptical about coding them in the theory of grammar in the form of derivational devices or in the form of universal hierarchies . . . The theory of the syntactic component of the faculty of language need not – and therefore must not – encode devices that translate the kind of external requirements to be satisfied: the syntactic component does not have to be defined to avoid the generation of [structures violating vacuous quantification]; similarly, if cartographies derive from the Full Interpretation condition, cartographies do not reflect the knowledge of grammar. [I would prefer the term 'syntax' here instead of 'grammar' – CB]

In the words of Harbour (2006): "Combinatorial restrictions have to be explained, not reified." To drive the point home, let me illustrate this by means of the following paradigm.

(2) a. They have a ton of money
b. They have tons of money
c. *They have seven tons of money

Watanabe (2009) seeks to explain the badness of the (c) example, and in so doing commits what I like to call "the grammarian's fallacy." His explanation is cartographic in nature, as he accounts for the fact that the presence of a numeral is incompatible with the idiomatic "very large [vague] quantity" by saying that "Vague quantity expressions are merged in the Spec of QP, which is above #P." But this structural generalization, if true, is not what accounts for the badness of the example in question in a deep way (in a way that goes beyond explanatory adequacy). It is what is in need of explanation (why exactly is QP above #P?). I wager that such an explanation lies in the fact that lexical items serve as instructions to activate certain concepts, which reside in dedicated (core) knowledge modules (Spelke and Kinzler (2007)), and that activating "very large vague quantity" (appealing to one of the components of our number sense (Boeckx (2009b); Dehaene (1997)), namely large approximate quantity estimation by means of *tons (of)* clashes with the instruction to activate a different component of our number sense, the one dealing with precise quantities, by means of numerals like *seven*.

Fortuny (2008), whom I quoted above, suggests an approach along similar lines to account for various ordering restrictions uncovered in Cinque (1999). Although a lot of work remains to be done in this area (see Chapter 3 for additional discussion), this kind of interface-oriented approach has the dual benefit of diminishing the appeal to pre-syntactic lexical design (lexicocentrism) and of forcing linguists to pay attention to properties of other cognitive modules, thereby enhancing the quality of interdisciplinarity that is badly needed to advance the biolinguistic program. Short of that, it is hard to disagree with Koster's (2010) "dissatisfaction with current Minimalism":

> My concerns are not about Minimalism as a program. On the contrary, I subscribe to the overall goal to construct a theory that makes grammar look as perfect as possible and that relegates as much as it can to "third factor" principles. My dissatisfaction is about how this program is carried out in practice. Others disagree, but my personal feeling is that little *theoretical* progress has been made since the 1980s. I emphasize *theoretical*, because empirically speaking the progress has been impressive. One can hardly think of any topic nowadays of which it cannot be said that there is a wealth of literature about it. All of this progress, I claim, is mainly "cartographic" and therefore compatible with pre-minimalist generative grammar and even certain forms of pre-generative structuralism. Part of the theoretical stagnation is due to the fact that some key problems of earlier versions of generative grammar, as they arose for instance in the GB-period, are either unresolved or ignored. But there are deeper problems, it seems, that involve the very foundations of the field.

The term 'biolinguistics' has now appeared a few times in the preceding paragraphs, and I think it's time to stress again that my attempt to defeat lexical design/lexicocentrism is ultimately motivated by biological plausibility.

1.3 Interdisciplinary measures

A growing number of passages such as those that follow can now be found in the linguistics literature.[10]

> How much should we ask of Universal Grammar? Not too little, for there must be a place for our unique ability to acquire a language along with its intricacies and curiosities. But asking for too much won't do either. A theory of Universal Grammar is a statement of human biology, and one needs to be mindful of the limited structural modication that would have been plausible under the extremely brief history of Homo sapiens evolution. (Yang, 2010, 1160)

> Given this, consider a second fact about F[aculty of]L[anguage]: it is of recent evolutionary vintage. A common assumption is that language arose in humans in roughly the last 50,000–100,000 years. This is very rapid in evolutionary terms. It suggests the following picture: FL is the product of (at most) one (or two) evolutionary innovations which, when combined with the cognitive resources available before the changes that led to language, delivers FL. This picture, in turn, prompts the following research program: to describe the pre-linguistic cognitive structures that yield UG's distinctive properties when combined with the one (or two) specifically linguistic features of FL . . . The short time scale suggests that the linguistic specificity of FL as envisaged by GB must be a mirage. FL must be the combination of operations and principles scavenged from cognition and computation in general with possibly small adventitious additions. In other words, despite appearances, FL is "almost" the application of general cognitive mechanisms to the problem of language. The "almost" signals the one or two innovations that the 50,000–100,000 year time frame permits. The minimalist hypothesis is that FL is what one gets after adding just a little bit, a new circuit or two, to general principles of cognition and computation. (Hornstein, 2009, 4)

These passages express well the point of view at the heart of Hauser *et al.* (2002), who, more than anything, want to draw attention to the richness of the Faculty of Language in the Broad Sense, and the many difficulties of assuming a high degree of linguistic specificity (a rich Faculty of Language in the Narrow Sense). This is the point where (as Jackendoff and Pinker

[10] Such passages are about what Richards (2008a) has called "phylogenetic adequacy"; equivalently, "natural adequacy" in Boeckx and Uriagereka (2007), "evolutionary adequacy" in Longobardi (2004) and Fujita (2009), and "biological adequacy" in Narita (2010b). On evolution imposing explanatory constraints on linguistic theorizing, see also Kinsella (2009), and the work of Ray Jackendoff from Jackendoff (1997) onwards.

correctly pointed out) the Hauser–Chomsky–Fitch vision meets minimalism, which takes a deflationist stance on Universal Grammar ("approaching it from below"), and which, according to Jackendoff (2005), gives minimalism its "empirical bite":

> At the time [pretty much throughout the history of generative grammar], it seemed that FL must be rich, highly structured, and substantially unique . . . Throughout the modern history of generative grammar, the problem of determining the character of FL has been approached "from top down": How much must be attributed to UG to account for language acquisition? The M[inimalist] P[rogram] seeks to approach the problem "from bottom up": How little can be attributed to UG while still accounting for the variety of I-languages attained? (Chomsky, 2007, 4)

The clearest expression of this "rich, highly structured, and substantially unique" hypothesis regarding UG – its apogee – was the modular view at the heart of Government-and-Binding (Chomsky (1981)). There, UG was assumed to consist of a variety of internal modules, of many distinct components that interact in complex ways. (This view remains at the heart of parametric models such as Baker's (2001) parameter hierarchy.) The problem with such a view is that this amount of internal modularity, and its complex organization can, biologically speaking, only have one source (if modules are assumed to be innate, which the relevant authors assume): natural selection. Everyone agrees on this much, from ultra-Darwinians (Dawkins, Pinker) to advocates of more nuanced positions (Gould, Fodor).[11] Fodor (1998) puts it best when he writes:

> If the mind is mostly a collection of innate modules, then pretty clearly it must have evolved gradually, under selection pressure. That's because . . . modules contain lots of specialized information about problem-domains that they compute in. And it really would be a miracle if all those details got into brains via a relatively small, fortuitous alteration of the neurology. To put it the other way around, if adaptationism isn't true in psychology, it must be that what makes our minds so clever is something pretty general.

[11] Here are some relevant quotes:

> [Natural selection], as far as we know, is the only process ultimately capable of generating complexity out of simplicity. (Dawkins, 2006)

> The key point that blunts the Gould and Lewontin critique of adaptationism is that natural selection is the only scientific explanation of adaptive complexity . . . the only explanation for the origin of organs with complex design is the process of natural selection. (Pinker and Bloom, 1990)

> I know of no scientific mechanism other than natural selection with the proven power to build structures for such eminently workable design. (Gould, 1997)

So, on the assumption that complexity requires natural selection and that natural selection requires time to work its magic (everyone's best guess, from Darwin onwards, is that it requires lots of it: say on the order of (at least) millions of years), the rapid rise of language in humans does not allow for this kind of complexity to develop. It is true that one may question the assumption that complexification requires a lot of time. After all, complex systems theorists demonstrate on a daily basis that complexity can emerge rapidly, but it is important here to distinguish between the kind of complexity that Complex Systems Science studies (which is complexity that emerges from simplicity, and is always generic in character) and the sort of complexity that we are talking about here (innate, module-internal, highly domain-specific complexity) (on various layers of complexity, see Deacon (2006)). Of the latter, we only have good examples that required a lot of time to come about.

One could, of course, also question the assumption that the human language faculty is of extremely recent vintage. If, contrary to what most scientists think, the language faculty has a long history, the internal complexity/modularity ascribed to it in previous models may have had enough time to emerge. So, how good is the evidence for this recent emergence?

There is no proof, of course, but it's a fairly good conjecture. In fact, I tend to think that the evidence is now stronger than many suspect. It's now generally agreed upon, based on converging evidence, that the guess we are talking about is as good as any: the emergence of new tools, cultural artifacts, signs of trade, cave paintings, and so on, that we find in the archeological record, first in Africa and then in Europe (McBrearty and Brooks (2000), Mellars *et al.* (2007)), points to a significant evolutionary transition. I tend to agree with Diamond (1992), Tattersall (1998), and many others that it is hard to imagine the emergence of these artifacts and signs of modern human behavior in the absence of the language faculty. But, as I pointed out in Boeckx (2011b), we can now make an even stronger case for the suddenness of the emergence of the language faculty. Recent genetic data suggest that *Homo sapiens* split into two sub-populations around 150,000 years ago, which remained separated for about 100,000 years (Behar *et al.* (2008)). If this interpretation of the data is correct, it suggests that the language faculty was already in place 150,000 years ago. If we combine this with the molecular evidence that suggests something(s) significant happened around 200,000 years ago (Piattelli-Palmarini and Uriagereka (2005) on FOXP2; Bufill and Carbonell (2004) on the epsilon3 allele; Williams *et al.* (2006) on Protocadherin11), the time of appearance of *Homo sapiens*, you get a window of time of just about 50,000 years (between 200,000 and 150,000 years ago). So all

in all, I think we should really try to keep the evolutionary add-ons for language to a bare minimum, with virtually no internal modularity (*ab initio*) to speak of.

I am not the first to make an argument against language-internal modules in the context of minimalism. Hornstein (2001), I think, was the first to see this issue clearly: the real target of minimalism is not government, Spec-head, or whatever else has been suggested in the literature. It should be language-internal modules. But although Hornstein expressed this admirably, it seems to me that he has not seen that this means we should also reduce our appeal to lexical features as much as possible. Curiously, Hornstein (2001, 215–216) addresses this issue briefly, and says that replacing modules by features (say, a Θ-module by θ-features requiring checking) is an advance. I think he is wrong, for features are nothing more than 'nano'-modules that in the hands of the cartographers naturally lead to massive modularity. Features have the very same degree of language-internal specificity and informational encapsulation as modules. All the talk about modules listed in Chomsky (1981) may have disappeared from the minimalist literature, but modules themselves live on disguised as features. It is easy to find featural counterparts to all the modules postulated in GB, which, to me, illustrates that much of minimalism today is really still GB in disguise. If, as I am trying to argue, going beyond explanatory adequacy means going beyond features/modularity, we have not begun to explore the space of possibilities afforded by minimalism. It is time for minimalism to evolve away from its precursors and become something new and distinctive on the theoretical linguistic scene. It is (high) time we take minimalism for what it is, a program to explore, not a set of assumptions to stick to, especially if it's one that assumes all the answers instead of asking the questions, as all lexicocentric approaches invariably do.

Of course, even programs have to be operationalized if one is to explore them. And at the beginning of minimalism, features did just that. They made it possible to identify properties of the computational system such as Last Resort and other economy principles. But appealing to case-features and Greed to explain the contrast between (3) and (4) is one thing. Appealing to EPP-features to account for the EPP effect, or to scrambling features to account for scrambling, is, I'm sure everyone will admit, another.

(3) a. John seems [*t* to have left]
b. It seems [John has left]

(4) a. *seems [John to have left]
b. *John seems [*t* has left]

From the use of features, minimalists quickly moved to the abuse of features. Cartographic representations (even fine-grained ones such as those assumed by proponents of nano-syntax; see Starke (2010)) encode notions that cannot possibly be taken as primitives: "ProcessP" or "EvaluativeMoodP" are almost certainly the output of syntax-dependent interpretive processes, not pre-syntactic primitives. They are the product of an interaction among various cognitive components, not the trigger of syntactic operations. In this sense, features in syntax have gone the way of genes in biology. From a plausibility argument to deal with certain facts ((3), (4) in the case of features, Mendelian data in the case of genes), they have become all powerful to the point of relegating syntax (in linguistics) and the organism (in biology) to the level of a mere vehicle of lexical/genetic expression – the selfish gene and (as one might call it) the selfish lexeme view. And just like geneticists hope to understand the organism by sequencing genomes, cartographists hope to understand syntax (and the syntax–semantics interface) by examining functional (lexical) sequences. Both clearly suffer from an unacceptable degree of preformationism and Platonism (a case of "misplaced concreteness" as A. N. Whitehead would have said). It's curious that linguists who have written about the evolution of language are able to lucidly identify the severe limitations of the selfish-gene worldview, but are not so quick to identify the very same limitations in their own linguistic practice. Shouldn't they take out the beam of wood from their own eyes before removing the speck that they see in their neighbor's?

Just like genocentrism led to the demise of the organism and development, lexicocentrism, in all its variants,[12] is leading to the end of syntax, to models where "syntactic structures are not generated by lexicon-independent rules . . . but as the spelling out of the contextual properties of lexical items ('valency')" (Jan Koster),[13] models where phrase structure (i.e. syntax) cease to exist, as in Collins and Ura (2001), who explicitly advocate a Word Grammar-style grammatical model, where lexical valency is the most important component (see also Carnie (2008), Osborne *et al.* (2011)).

The alternative, as is now clear in biology with the revival of embryology under the rubric of "Evo-Devo,"[14] is to place the emphasis on developmental

[12] Accordingly, the cartographic claim that functional sequences like Cinque's hierarchy (Cinque (1999)) are "syntacticizing semantics" (Cinque and Rizzi (2010)) is misleading, since the syntax presupposed is lexically driven, syntax merely projecting whatever information it receives from lexical entries or from Platonic functional sequences, as in nanosyntax.

[13] http://odur.let.rug.nl/~koster/resume.htm

[14] Being a program, "Evo-Devo" is pursued along many lines, some of which are still heavily gene-centered. The present book is more in line with more epigenetic pursuits within Evo-Devo. For discussion, see Benítez-Burraco and Longa (2010); Boeckx (2014b).

dynamics and phenotypic plasticity, epigenetics, and the emergence of theories such as niche construction, and focus on organismic processes as opposed to genetic blueprints or programs, and on interactions (the interactome as opposed to the genome; the triple helix as opposed to the double helix). As such it seems tailor-made for minimalist explorations, especially once these discard lexical blueprints or programs (i.e., numerations), and truly explore interface-based explanations. Much like the emerging extended synthesis in biology, linguistics will have to embrace pluralism, get rid of isolationist (i.e., modular, self-sufficient) tendencies, and revisit the works of old foes to treat them as friends.[15] Much like modern biology, modern linguistics will have to soften its stance on various issues, especially those touching on specificity and innateness (Massimo Piattelli-Palmarini would talk about this in terms of leaving behind the (necessary) age of specificity; see Piattelli-Palmarini (2010)). The range of processes explored are likely to be more abstract (less-task-dependent) and generic. My own view is that this is the only way linguists have to take on what one may call "Poeppel's challenge," on which I am about to expand.[16]

Like many others, Poeppel would like to relate mind and brain somehow, and like many he is not impressed with how far we've gotten until now. It's not just the linguists' fault, of course. But I tend to think it's in part our fault. It is fair to say that GB is cognitively exceptional in that its principles and operations are cognitively *sui generis* and very specific to language. As I already mentioned above, Eric Reuland once said that GB principles are "too good to be false," but he also added that they are "too (domain-)specific to be true." As Poeppel and Embick (2005) observe, this is a serious problem for those aiming to find brain correlates for the primitives of FL. They dub this the granularity

[15] For linguistics, this means taking into account the fact that many works of biologists of which generative grammarians approve are closely associated with psychological models that generative grammarians have prematurely rejected *in toto*. Witness the influence of Waddington on Piaget, or the influence of McCulloch and connectionism on Stuart Kauffman. For more on this point, see Boeckx (2014a).

[16] Fitch (2009, 298) formulates a similar challenge in his prolegomena to a science of 'biolinguistics':

> We need to distill what we know from linguistic theory into a set of computational primitives, and try to link them with models and specific principles of neural computation . . .
>
> Thus we need linguistic models that are explicit about the computational primitives (structures and operations) they require, and that attempt to define linguistic problems at a fine enough grain that one can discuss algorithmic and implementational approaches to their solution. We need a list of computations that linguistic theorists deem indispensable to solve their particular problem (e.g., in phonology, syntax, or semantics).

problem. The aim is to find those primitive operations that are at once empirically grounded and that could be embodied in neural wet-ware. Given this, the goal for the linguist will be to find a class of very basic primitive operations that plausibly underlie linguistic computations for consideration as candidates for possible neural circuits.

As Poeppel notes elsewhere (Poeppel (2005)), if anything is localized in nervous tissue, at least at the level of cellular ensembles or columns, it will be elementary computational functions. Poeppel goes on to say (and that's his challenge to the linguistic community)

> Linguists and psycholinguists owe a decomposition (or fractionation) of the particular linguistic domain in question (e.g. syntax) into formal operations that are, ideally, elemental and generic. The types of computations one might entertain, for example, include concatenation, comparison, or recursion. Generic formal operations at this level of abstraction can form the basis for more complex linguistic representation and computation.

I think this is a great challenge. And it's for the biolinguistic community as a whole. But I think minimalists in particular should take this challenge to heart because, as I will try to show below, by their emphasis (in theory, if not yet in practice) on "approaching UG from below," they are better equipped than most to meet this challenge. So, they would be missing a great opportunity if they didn't try. I have argued elsewhere (Boeckx (2006, 2009b, 2010a), Di Sciullo and Boeckx (2011)) that minimalism has contributed substantially to the re-emergence of interdisciplinarity in linguistics and thus to the return of biolinguistic discourse. The minimalist program forces linguists to reformulate previous findings in terms of elementary units, operations, and interface conditions; those that, according to our very best bets, have the character of conceptual necessity (those that are, in the passage by Fitch quoted in n. 16, "indispensable"). As one can reasonably anticipate, many of these will have such a generic flavor to them (combine, map onto a linear sequence, etc.) that they are plausibly not specific to the language faculty. This should be very good news to researchers in other areas, as the concepts articulated by minimalists may find an equivalent in their own field, or be more readily testable using familiar techniques, something that modern ethologists fervently desire (see Balari and Lorenzo (2013) for a similar point). At the same time, some of these generic operations will make it more plausible to entertain 'descent with modification' scenarios concerning the evolution of language,[17] and also, I suspect,

[17] On a particularly compelling call in favor of descent-based arguments, in a context very much related to the subject of the present discussion (modularity for him, features/nano-modules for us), see Marcus (2006).

reconnect with experts on psycholinguistics and language acquisition, who currently view our isolationist (read: modular) stance as counterproductive.

Some aspects of lexicocentrism have been highlighted by Distributed Morphologists like Alec Marantz,[18] or advocates of exoskeletal approaches to word structure like Hagit Borer, who correctly implicate syntax more into the construction of the lexicon. Nevertheless there is a crucial residue of generativity in their pre-syntactic lexicon (as Marantz (1996) acknowledges: see n. 9), as they do not seek to derive (syntactically) the functional lexemes (and the hierarchies they form) that they crucially rely on to deflate the content of lexical categories.

Ray Jackendoff, too, has identified part of the problem with lexicocentrism in various publications (1997, 2002, 2005, 2010), by drawing attention to the constant projectionist models of syntax, and also, like Marantz and Borer, to the phrasal nature of much of our lexical knowledge (what Construction Grammarians call "Constructions"), but I suspect that he too retains a lexicocentric view of syntax. Unfortunately, Jackendoff is not explicit at all here. He (in my view, correctly) takes language-specific lexical items to be mappings between different representations they are (small) interface rules ("correspondence rules"; see Marantz (1984)), but nowhere (as far as I know) does Jackendoff tell us how these different representations come about. Jackendoff claims that the key grammatical process is Unify (as opposed to Merge), but in order for Unify to apply, different representations must be constructed. Here too, we begin *in medias res*. In the context of Jackendoff's "Simpler Syntax" model, this is an important point, for the antidote to lexicocentrism is a heavier reliance on syntax (a heavier dose of "syntactocentrism," as Jackendoff would say), which is just the opposite of what Jackendoff claims: if one wishes to explain (or, better said, go beyond explanatory adequacy), one should be a Constructing Grammarian, not a Construction Grammarian.[19] One should adopt exo-lexicalism *in toto*, that is, do away with any remnant of lexicalism. Marantz (1995) is wrong, minimalism (seen as the attempt to go beyond explanatory adequacy) is not the end of syntax, it's the end of the all-powerful lexicon.

The solution is not to let syntax blindly follow lexical instructions, but rather to let syntax construct the lexicon, for it is only by constructing things that one can hope to explain them. The Inclusiveness guideline suggested by

[18] For example, Marantz (2013) correctly points out that it is wrong to talk about unaccusative vs. unergative verbs; rather, we should talk about unaccusative vs. unergative *structures*.

[19] As Adger (2013a) puts it, constructions are not explanations, the same way that brain maps are not explanations for how the brain works (see Poeppel (2012)).

Chomsky (1995) was a way to prevent the reification of relations in featural terms (e.g., indexing in the context of binding relations). This was not meant to be a ruling against emergent phenomena in syntax. Getting out as much (in syntax) as one puts in (the lexicon) simply begs the question of what syntax is for.[20]

Let me conclude by stressing again that lexicocentrism really threatens the very project of "biolinguistics." The more we rely on the lexicon, the more inescapable Koster's conclusion that linguistics is at best applied biology appears to be. As Koster notes[21]

> In [Koster's] overall theory of language [which resembles the mainstream more than he claims], the capacity for language, even in its narrowest sense, is not seen as a matter of biology but as applied biology, i.e., a technology belonging not primarily to individuals but to their shared culture. Invented words rather than syntax are at the essence of language in this view, while recursive syntax is seen as a successful extension of the properties of the cultural objects in question ("words"). The combinatorial potential of words is as cultural and non-individual as the words it belongs to and therefore first and foremost public property that individuals seek to adopt from the day they are born into a community.

If one wants to avoid this conclusion, as I think we should (see Balari *et al.* (2012)), we must try very hard to limit the scope of lexical supervision, and seek a level of representation of syntax that is appropriate for biolinguistic inquiry. As a matter of fact, the goal is not all too different from the one sketched in the opening paragraph of *Syntactic Structures*:

> linguists must be concerned with the problem of determining the fundamental underlying properties of successful grammars. The ultimate outcome of these investigations should be a theory of linguistic structure in which the descriptive devices utilized in particular grammars are presented and studied abstractly, with no specific reference to particular languages. (Chomsky, 1957, 11)

The remainder of this book is an attempt in this direction.

Let me close by saying that even if the particular alternative I articulate in the following chapters turns out to be on the wrong track, at the very least

[20] For a very good example of lexical overspecification, consider the treatment of locality in Müller (2010), where lexical entries consist of not one, but multiple structured stacks of features.

[21] Taken from http://odur.let.rug.nl/~koster/resume.htm; see also Koster (2009).

my questioning of lexicocentrism has what Chomsky sometimes refers to as "a certain heuristic and therapeutic value" in the context of minimalism:

> It is perhaps worth mentioning in this connection that the Minimalist Program, right or wrong, has a certain therapeutic value. It is all too easy to succumb to the temptation to offer a purported explanation for some phenomenon on the basis of assumptions that are of roughly the order of complexity of what is to be explained. If the assumptions have broader scope, that may be a step forward in understanding. But sometimes they do not. Minimalist demands have at least the merit of highlighting such moves, thus sharpening the question of whether we have a genuine explanation or a restatement of a problem in other terms. (Chomsky, 1995, 233)

By highlighting the tacit importance of operations like "Bundle" or preformed functional sequences driving syntactic derivations, the present investigation will reveal how much Merge can explain, and how much (or how little) we genuinely understand about the language faculty.

2 *Syntactic order for free: Merge α*

2.1 Bare Merge

The central thesis of this book, which I have tried to express and illustrate in the opening chapter, is that the widely shared theoretical stance that I have called lexicocentrism has dominated linguistic theorizing for far too long and that we would do well to give it up if the goal of theorizing is to "go beyond explanatory adequacy," "to approach UG from below," and shed light on the nature, origin, and development of the human language faculty.

Being so rampant, lexicocentrism takes on many guises in the literature, and I have highlighted two particularly salient forms of it in Chapter 1: (i) an unconstrained appeal to domain-specific, non-primitive features, poorly grounded in cognition and its evolution; and (ii) a constructionist/preformationist core that limits the (explanatory) scope of Merge. In this chapter I'd like to remedy this situation, and allow Merge to structure lexical items and rely on properties of syntactic units that are so primitive and minimal as to be unquestionable. In other words, I'd like the lexical component on which syntax feeds, which following Marantz (1997) I will refer to as the *narrow lexicon*, to be primitive and structureless (and thus, unlike in Marantz's model, not even minimally generative). At the same time I'd like Merge to be as basic and generic an operation as possible, in keeping with David Poeppel's desideratum for a genuine biolinguistics. Saying that Merge should be basic also means that it, too, should be 'structureless': it should not be composed of suboperations, as it has often been in the past. This is what Boeckx (2009e) and Hornstein (2009) meant when they called for "decomposing" merge. The same admission is found in Chomsky (2012a, 4) where it is said that until recently, "Merge was complicated to provide a label." The complication Chomsky is alluding to lies in the asymmetry imposed on Merge by what I have called in Chapter 1 the central dogma of transformational generative grammar: the Projection Principle. As is clear, for example, in the formulation of

Merge in Chomsky (1995, 244),[1] the symmetric act of combination had to be accompanied by an asymmetric act of projection, yielding an asymmetric head–complement relation, with the head providing the label for the set generated. As Chomsky (2012a, 4) remarks, this was just "by stipulation." Dropping the projectionist dogma for the time being (I will come back to the need for labeling below), the simplest, most desirable, symmetric formulation of Merge is as follows:

(1) Merge_{def}
Take two lexical items α and β and form the set $\{\alpha, \beta\}$:
$\mathbf{M}(\alpha, \beta) = \{\alpha, \beta\}$

Bare Merge, then, is, at bottom, unrestricted, unbounded set-formation.[2] The challenge ahead of us is to rebuild syntactic theory and with it, grammatical theory, from (1). In this sense, the challenge can be said to be 'minimalist,' if minimalism is understood as an attempt to "reconstruct syntactic theory around Merge as the central computational operation," as Jackendoff (2011) nicely put it.

Right away, it should be clear that taking (1) as a theoretical foundation will have important architectural repercussions. It will not only affect our view of syntax and of the (pre-syntactic, 'narrow') lexicon, it will also require us to rethink what we understand by the external systems with which syntax interacts to yield externalized and interpreted grammatical expressions. This is the topic of the rest of this book.

2.2 Lexical precursor cells

Let me begin with the pre-syntactic ('narrow') lexicon, and try to come up with a principled answer concerning what it is made of. The previous section made clear that minimally, we ought to expect the units that syntax manipulates to be 'combinable,' or 'mergeable.' Minimally, then, lexical items must be endowed with at least one property: the capacity to merge. This is the property that Chomsky (2008, 139) already identified, and dubbed the "edge feature" of a lexical item. Here is the relevant passage:

> For a L[exical] I[tem] to be able to enter into a computation, merging with some [syntactic object], it must have some property permitting this operation.

[1] The asymmetry of Merge is extensively discussed in Boeckx (2008b, chap. 3).

[2] It follows that Merge takes on its linguistic specificity when we consider the units it combines, how many of these there are, and in which context this combination operates.

> A property of an LI is called a feature, so an LI has a feature that permits it to be merged. Call this the edge-feature (EF) of the LI.

There are several things to say about this 'edge' property (e.g., why call this the 'edge' property?), and I will tackle these in due time, but for now let me make clear that Chomsky's passage does not say that this is the only feature lexical items have. But in light of what I have said in the previous chapter concerning lexicocentric tendencies in linguistic theorizing, suppose we adopt as a working hypothesis the idea that this is the only property of lexical items relevant for syntactic computation. Notice that if we make this assumption, from the perspective of syntax, all lexical items are alike, since the edge feature is something that by definition all lexical items must have. Of course, this does not mean that linguistically speaking, all lexical items are alike. They may be syntactically indistinguishable, but semantically or phonologically, they typically end up very different. Such differences could be modeled in terms of indices, as Harley (forthcoming) and others have suggested for 'roots' in frameworks like Distributed Morphology. Such indices could function as addresses: instructions for the syntax-external systems to 'fetch' ('bind'/activate/retrieve in long-term memory) the relevant concept on the semantic side or the relevant phonological matrix on the sound/sign side. So, lexical items could be seen as triplets: an edge property, coupled with two indices, one pointing to the system(s) responsible for interpretation, the other to the systems responsible for externalization. Crucially, such lexical items are not feature bundles in the traditional sense criticized in Chapter 1: the edge property and the indices do not form internally complex treelets and their grouping does not depend on merge. I take the indices not to be visible to the syntax (they could be late inserted, at the point of Spell-Out; see next chapter), which as far as it is concerned only sees singletons (atomic units endowed with the edge property). Put differently, the lexical entries of the narrow lexicon are flat.

There is another sense in which the lexical items considered here differ from the traditional ones. In standard models, the syntactic features that make up lexical items are either interpretable on the relevant item (say, [Person] on nouns) or not (say, [Case] on nouns); that is, they clearly have a meaning dimension to them. The edge property is a purely structural property, it has neither semantic nor phonological content. Perhaps because of these differences, readers may feel that it is misleading to call the units I have in mind 'lexical items.' Anticipating the discussion to follow, which will relate conceptually to how biological patterns can be reliably generated from a homogeneous medium by the sheer force of physico-chemical ("third factor") properties, as originally proposed by Alan Turing in his seminal (1952) paper, I am tempted to refer

to the units of the narrow lexicon as 'lexical precursor cells,' and will call them that, as opposed to (full-fledged) 'lexical items,' when I want to avoid confusion.[3]

The idea that the narrow lexicon is made up of elements about which one can say only one thing – that they are mergeable – highlights an important theme of the present book as a whole: the (radical) dissociation of form and substance. Chomsky (1965) introduced the distinction between formal and substantive universals, but form (syntax) has never been kept really separate from its "substantive foundation," a term that Muysken and van Riemsdijk (1986) use to collectively refer to "features." This has led to rather curious situations, at times verging on the paradoxical. Consider, for example, the idea that the mechanism underlying recursion in language (by hypothesis, Merge) is also responsible for the human-specific nature of our musical and mathematical abilities (an idea that Chomsky has expressed on several occasions, at least since Chomsky (1988)), two realms that also exhibit the property of discrete infinity. Chomsky (2008) proposes that the natural numbers can be generated by iteration of Merge applied to a singleton lexical item (Ø or { }).[4] But notice that for this to be tenable, Merge must be free of what Pesetsky and Torrego (2007) call its "vehicle requirement": it cannot be triggered, for if it is allowed to apply to a single element, that element cannot satisfy its own need. Likewise, in their detailed examination of structure in music, Katz and Pesetsky (2009) put forth the following "Identity Thesis for Language and Music": "All formal differences between language and music are a consequence of differences in their fundamental building blocks (arbitrary pairings of sound and meaning in the case of language; pitch-classes and pitch-class combinations in the case of music). In all other respects, language and music are identical." Specifically, they argue that music, like language, contains a syntactic component in which structures are built by merge. Here too, then, merge is allowed to apply to units that are quite distinct from traditional lexical items. Again, for this to be tenable, merge must be freed of lexical influence.

[3] The term 'lexical precursor cell' corresponds to what I called 'conceptual address' in an early, incomplete draft of this book circulated online ("Defeating lexicocentrism," made available on Lingbuzz). I have abandoned the term 'conceptual address,' as it gave rise to misunderstandings (e.g., in Acquaviva and Panagiotidis (2012)) that in retrospect were perhaps to be expected due to my use of the loaded adjective 'conceptual.'

[4] Here is the relevant passage in Chomsky's recent work:

> Suppose that a language has the simplest possible lexicon: just one LI, call it "one." Application of Merge to the LI yields {one}, call it "two". Application of Merge to {one} yields {{one}}, call it "three." Etc. (2008)

In a similar vein, Martina Wiltschko in a series of presentations and as-of-yet unpublished work (e.g., Wiltschko (2009)) has argued (in part addressing the Evans and Levinson (2009) challenge) that grammatical "categories defined via substantive content cannot be universal." She shows on the basis of a wide and deep cross-linguistic survey that certain concepts (e.g., "number") map onto different nodes in different languages, whereas at the same time, the same node (e.g., INFL) is filled or 'substantiated' in different languages by different elements (Person, Location, or Tense).[5] In sum, "categories are not intrinsically associated with substantive content"; syntactic categories must be defined more abstractly/generically (a point I return to below in the context of cartography, as it eventually vitiates the idea of a unique functional sequence of the sort argued for by Cinque (1999), Starke (2010), and others). Wiltschko's idea, taken to its logical conclusion, seems to me to indicate that syntax operates free of lexical content, leading to a homogeneous narrow lexicon and indistinguable lexical precursor cells of the sort advocated here.[6]

The lexical precursor cells populating the narrow lexicon in the present framework have many, though I insist, not all, the properties that have come to be associated with roots in frameworks like Distributed Morphology (Halle and Marantz (1993)) or lexemes in the exoskeletal morphology model of Borer (2005). For Borer, Marantz, and others, roots are said to be 'syntactically deficient,' in the sense that in these frameworks they cannot take complements (they cannot head phrasal constituents/project), nor can they impose selectional requirements on structure: in Borer's terms, roots are acategorial, monomorphemic, and lack argument structure. It's precisely because of this syntactic underspecification that roots must occur in the context of a (functional) categorizing head (Marantz's "little *x*": *n*, *a*, or *υ*). It's also thanks to this underspecification that roots show the sort of flexible valency extensively illustrated in Borer (2005). The only thing roots can do is merge, which is exactly the computational scope I give to lexical precursor cells. Harley (forthcoming) has been most explicit in calling roots "pure units of structural computation,

[5] See also Miyagawa (2010) for a partially similar conclusion regarding agreement in agreement/ϕ-feature-less languages.

[6] Wiltschko appears to resist this conclusion because she holds onto the idea that grammatical categories can be defined, and distinguished, in terms of generic functions such as 'typing,' 'anchoring,' 'viewpoint,' 'classification,' whereas I am inclined to believe that these various functions are abstractly the same, with the differences arising configurationally. See next chapter for discussion. (It should also be said that Wiltschko continues to assume that merge requires abstract featural triggers, and in this sense her proposal retains a lexicocentric character.)

lacking both semantic content and phonological features."[7] The reason I did not call my lexical precursor cells roots is because in studies on roots such as those found in Distributed Morphology, or in Borer (2005), or De Belder and van Craenenbroeck (2011), roots acquire their characteristics when contrasted with the functional elements that also inhabit the pre-syntactic, narrow lexicon.[8] For Marantz, Borer, and others, roots are special precisely because of the asymmetry in the lexicon between roots and functional units. But I agree with Harley (forthcoming) that in fact "roots behave like normal syntactic elements."[9] For Harley, this means that roots, like functional elements, can project. For me, since no element projects in the syntax, roots are no different from functional units. The lack of a pre-syntactic lexical distinction between lexical and functional units is in fact an important property of the present model (see Manzini and Savoia (2011) for a partially converging view), and an important difference with the models like that of Sigurdsson (2011), which otherwise seek to reduce the content of the pre-syntactic, narrow lexicon.

In many ways, the lexical precursor cells of the present book are the syntactic equivalent of the very minimal units of semantic computation in natural languages advocated by Paul Pietroski in a series of papers and forthcoming book (Pietroski (2007, 2008, 2010, 2011, 2012, forthcoming)). Pietroski departs from a lot of logic-oriented work in formal semantics, and builds on Chomsky's remarks on meaning over the years that take semantic computation to amount to the generation of instructions to language-external, but mind-internal systems of thought and action. Specifically, Pietroski argues that phrasal meanings are instructions for how to build conjunctive monadic concepts whose conjuncts correspond to the phrasal constituents. Inspired by Chomsky's minimalist program for linguistic theory, Pietroski's aim is to "reduce the stock of composition operations that semanticists regularly appeal to if only to make it more plausible that our innate endowment supports these operations, and perhaps to help identify the uniquely human aspects of this endowment." Put in other words, Pietroski's semantic theory

[7] Harley's paper shows that roots cannot be individuated on the basis of their phonological content (contra Borer (2005)), nor on the basis of their conceptual meaning (contra Arad (2005)). For this reason she adopts an index notation to individuate them syntactically. Such an index can then serve as instruction to the semantics and the phonology.

[8] For De Belder and van Craenenbroeck (2011), roots actually don't reside in the pre-syntactic, narrow lexicon; only functional elements do.

[9] In an early implementation of Distributed Morphology (Marantz (1996)), roots were also treated on a par with functional items in that both were inserted late (i.e., post-syntactically). This continues to strike me as correct.

is biolinguistically driven. I will have more to say about his general approach in the next chapter, when I deal with the systems of interpretation with which syntax interfaces, but here I would like to devote some space to an important aspect of Pietroski's enterprise concerning the units of semantic computation.

In Pietroski's system, lexicalization plays a special role, similar to what the edge property achieves for lexical precursor cells here. Pietroski advocates a model of semantic computation that is so minimalist (consisting of a conjunction operation and a limited version of existential closure) that it imposes a very specific and simplistic format on the units of combination: they should all be simple, monadic predicates. Pietroski demonstrates that if the units combined were allowed to vary in adicity, semantic composition would not be as systematic as we find it in natural language. Given that concepts clearly vary in adicity (certainly not all concepts we have are monadic; think of polyadic concepts like SELL), Pietroski (2012) suggests we conceive of lexicalization as "a formally creative process" in which nonmonadic concepts are turned into monadic predicates which can then be systematically conjoined with others. In other words, lexical concepts "exhibit less formal variation than the concepts absent lexicalization." It is as if lexicalization "dumbs concepts down," to achieve a stock of concepts that are systematically combinable via the simple composition operations that Pietroski assumes. This, in turn, means that for Pietroski "lexicalization is a large part of what makes humans linguistically [and, I would add, cognitively] special" (for an explicit statement regarding this point, see Pietroski (2008), which agrees with Boeckx (2011b, c); see also Ott (2009c)).

Pietroski's view contrasts with the familiar position that takes "lexical items simply [to] label the concepts they lexicalize, and composition of lexical meanings [to] mirror composition of the labeled concepts, which exhibit diverse adicities." Like Borer (2005), Pietroski is impressed by the range of argumental frames lexical items can fit into, which is surprising if they inherit the adicities of the concepts they lexicalize. If that were so, we would expect lexical items to be much 'pickier' (less flexible) than they patently are.[10] In sum, by "effac[ing] conceptual adicity distinctions," lexicalization yields a form of uniformity to the units of semantic composition that is very reminiscent of what the edge feature achieves with lexical precursor cells. Pietroski correctly notes

[10] To be sure, Pietroski does not

> deny that verbs are associated, at least statistically, with a canonical number of arguments. These associations presumably reflect, in part, the adicities of lexicalized concepts. But they may also reflect complicated interactions of grammatical principles with various contingencies of actual language use.

that this makes the 'lexicon' less varied and, philologically speaking, far less interesting, but, as he points out, "if the goal is to describe biologically implementable composition operations ... lack of sophistication may be a virtue." It also allows one to think of the core combinatorial operation (Conjoin, for Pietroski) as something that may, in some primitive form, be available to other animals, much like Merge, in the present proposal, which, at bottom, is nothing more than set-formation.

Both my lexical precursor cells and Pietroski's lexicalized concepts, though not identical, highlight the fact that a more homogeneous vision of lexical items may shed light on core biolinguistic aspects of the language faculty. Saussure was certainly right when he stated that "dans la langue il n'y a que des différences" [in a language, there are only differences], but on the basis of what we have discussed so far, one could say that "dans le langage, il n'y a que des similarités" [in language (i.e., the language faculty), there are only similarities]. (This contrast of opinion will be important when addressing the topic of language variation ('logodiversity') in Chapter 4.)

Since several passages drawn from Pietroski's works touch on what may be specifically human, I'd like to say a few words here about the treatment of the differences between birdsong and human language offered by Berwick *et al.* (2012).

Berwick *et al.*'s discussion also bears on the lexicon because it takes as its point of departure famed ethologist Peter Marler's well-known contrast between "phonological syntax" and "lexical syntax." On Marler's account, songbirds exhibit only phonological syntax, that is, the stringing together of elements, sounds, according to some well-defined pattern, but without the meaning of the resulting sequence as a whole dependent on the meaning of its individual parts. In contrast, Marler argues that only human language exhibits lexical syntax, that is, changes in meaning resulting from different combinations of elements such as word parts, words, or phrases. Put another way, Marler notes that while both birdsong and human language are combinatorial, in the sense that they both assemble larger structures out of more basic parts, only human language is compositional, in the sense that the meaning of a word or sentence changes as we change its component parts. Marler's use of the term "lexical syntax" may cause one to think that the key difference lies in the existence of lexical items ("words").

But Berwick *et al.* (2012) note that "Marler's notion that it is "lexicoding" – words – that completely characterizes the division between human language and birdsong captures part, but not all, of the necessary distinctions." In particular, "[i]t does not account for the inherent asymmetry of human language

structure, and falls short when it comes to describing human language structures that have no associated lexical meanings, such as the metrical or prosodic structure associated with human language."

Elsewhere, they write, "we should emphasize that it would be a mistake to conclude that all birdsong–human differences result simply from the lack of words in birdsong ... For example, even though birds lack words, there is nothing that logically blocks birdsong syntax from relying on syllable groupings or other features that could themselves be labeled by properties of their constitutive parts."

Berwick and colleagues conclude that "it is not the lack of words alone that blocks the possibility of more complex birdsong syntax. Rather, this gap is due to a fundamental deficiency in a very particular computational ability, namely, the lack of the combinatorial operation of the sort found in human language." (From their article, it is clear that they have something like 'Merge' in mind; see Berwick (2011).)

However, Berwick *et al.* end up partially agreeing with Marler, because the notion of Merge they advocate is not the one corresponding to (1), but the asymmetric version that fails to separate the act of combination from the asymmetric labeling of the output of merge. They write

> The key difference is the use of a verb or noun's features to label an entire word sequence with a single label ... the selection of a privileged element in this way renders the underlying structure fundamentally asymmetric. Note that there is no analog to this in birdsong ... Consider as an example the birdsong motif described earlier, consisting of seven particular syllables. This motif is not "labeled" by selecting just one of these syllables and its properties to name the entire motif; none of the syllables takes priority ... Neither is the resulting structure asymmetric as it is in human language. This is true precisely because birds apparently do not have words or manipulate word features at all. This is one difference between the human language syntactic system and birdsong. We noted earlier that this does not in principle bar the possibility of birdsong making use of features of song elements, for example, syllables and their acoustic features, and assembling them in a similar hierarchical fashion. However, current evidence suggests that this does not occur in birdsong. Rather, the combinatorial operator itself is absent.

By stressing the notion "labeling dependent on word features," Berwick *et al.* effectively end up "anchor[ing]" the combinatorial operator "on words or more precisely, word features," and conclude that

> It is this operation [Asymmetric Merge] that is apparently absent in birds, so far as we know. However, even though birds seemingly lack words, it does not follow that the combinatorial operator is necessarily absent in birds. For

> example, the combinatorial operator could still work on other elements, for example, syllables, in this way yielding the distinctive metrical patterning of sound melodies, rhythmic patterns ... However, for whatever reason, the operator does not appear to have been exploited this way in birds. It remains an open question as to whether a similar analysis would apply to birdsong metrical patterns; this then is a possibly crucial open research question where a non-human model might (speculatively) provide insight into its counterpart in human language. If birdsong were found to operate in a similar way to human metrical structure, this might provide precisely the required evolutionary "bridge," in the sense that the combinatorial operator was present in the common ancestor of both species, but full-fledged language required in addition words and their features, an ability present in the human lineage, but not in any bird species. It follows that it is precisely here that one might look for key evolutionary innovations that distinguish humans from birds.

My own view is quite different. For me, it's not the projection of (word) features, but the presence of an unrestricted combinatorial operator (corresponding to the notion of 'edge feature'; or, equivalently, the absence of selectional restrictions) that was the big evolutionary breakthrough. It was the formation of a homogeneous pre-syntactic, narrow lexicon that made it possible for humans to make infinite use of finite means, and construct "finite-yet-unbounded" systems (in the sense of systems yielding endlessly many complex expressions that can be characterized recursively), which encompass language, mathematics, and music. Rather than seeing the evolutionary breakthrough as a constraint on merge (labeling), we should think of it as the lifting of a constraint – the removal of selectional restrictions that block systematic combination of the sort we find in language. (I return to this characterization of "humaniqueness" in the next chapter.) This, of course, is not to deny that there are asymmetric, 'labeling' effects in grammar. Such effects exist, but they do not require special word features to emerge. Instead, I will show in the next section that such effects can be made to follow almost automatically from more primitive and generic cognitive properties, in accordance with Poeppel's desideratum for biolinguistics.

2.3 Regulating Merge α: Santa Fe-Style Syntax

The concept looming large in Berwick *et al.* (2012), the one for which they needed to resort to 'word features' (lexicocentrism), is Asymmetry. The pre-syntactic, narrow lexicon proposed here is completely homogeneous and uniform; all the lexical precursor cells are, for syntactic purposes, identical, and completely interchangeable. And the unique process discussed so far is

Merge, which, stated in its simplest form as in (1), amounts to a symmetric set-formation procedure. On the face of it, then, the present approach has no chance of 'saving' what one could call the Asymmetry effects captured by the notion of Projection in the traditional $\overline{X}$-schema. Everywhere you look in language, it seems that Saussure was right in saying that "a language is a system of differences with no positive terms." Think not only of features, which are nothing but expressions of differences ("distinguishers," as Muysken and van Riemsdijk (1986) aptly call them), but also of thematic asymmetries (the external vs. internal argument distinction at the heart of the unergative/unaccusative distinction), extraction asymmetries ("Condition on Extraction Domain [CED]-effects"), binding asymmetries (binder vs. bindee), and, perhaps the one underlying all the preceding ones, asymmetric c-command. These syntactic asymmetries also find correlates at the level of interpretation (scope asymmetries) and externalization (incorporation asymmetries, prosodic prominence asymmetries, and so on). The challenge for us is to find a way to capture these without resorting to an extra operation like 'Project' or relying on the asymmetries made available by features, since both of these strategies, familiar from the literature, are, at bottom, nothing more than ways of coding the facts. The same holds of recent proposals concerning labeling algorithms.

For example, as part of his attempt to eliminate phrase-structural residues from UG, Chomsky (2007, 2008), and most explicitly Chomsky (2013), suggests that endocentricity (the core fact that Projection tried to capture) is the result of a simple head-detection algorithm. Inspection of a syntactic object K of the form $\{\alpha, \beta\}$ determines α to be the head ('label') of K if α is an "atomic lexical item." (Chomsky refers to this search procedure as Minimal Search.) As Chomsky notes, this algorithm fails to return a single label in several contexts (situations where both α and β are atomic lexical items, or where both are phrasal). In such contexts, Chomsky resorts to an idea from Moro (2000), where structures that are too symmetric are 'asymmetrized' as a result of movement of one of the members causing the symmetry. As Moro (2000) shows, there are local domains, such as small clauses, which cannot contain two elements of the same type/category, as represented in (2) (with an illustration from Italian).

(2) a. **pro* copula [$_{SmallClause}$ DP DP]
b. *sono molte foto del muro la causa della rivolta
are many pictures of-the wall the cause of-the riot
'many pictures on the wall are the cause of the riot'

As can be seen in the next example, Moro shows that one of the two elements of a small clause (it does not matter which one) must vacate the relevant local domain.

(3) a. molte foto del muro sono la causa della rivolta
b. la causa della rivolta sono molte foto del muro

(For the sake of completeness, I should mention that Chomsky (2008, 145) also adds a special clause to his labeling algorithm to handle structures formed by movement: "If α is internally merged to β, forming $\{\alpha, \beta\}$, then the label of β is the label of $\{\alpha, \beta\}$.")

As should be obvious, Chomsky's labeling algorithm merely restates the traditional idea that the head of a phrase projects.[11] It does not explain why it does so. Moreover, it does not properly define the notion 'head' (or 'atomic lexical item'). As far as I can see, Chomsky understands the latter roughly as 'element taken from the (pre-syntactic) lexicon' (as opposed to syntactic unit formed in the syntax), but as the previous chapter makes clear, this leaves unaddressed the issue of how the internal structure of lexical items (which Chomsky assumes to be present) comes about. If the traditional notion of a lexical item as a bundle of features is adopted, then all lexical items are in some sense phrasal, meaning that the Minimal Search algorithm will fail to return a proper label in virtually all instances of Merge.[12] Last, but not least, the fact that the labeling algorithm proposed by Chomsky contains a disjunction (one clause for external merge, the other for external merge) strongly suggests that a generalization is being missed.[13]

2.3.1 *Enter phases*

The alternative I would like to pursue will seek to anchor the asymmetries mentioned above in a property of language that I have not touched on so far but that arguably falls within what Chomsky (1993) labeled 'virtual conceptual necessity.' The property is called *phase* (equivalently, *cyclic spell-out* or *cyclic transfer*), and the general idea to be pursued is to adopt a more

[11] As Chomsky (2008) admits, his algorithm contains ideas "carried over from X-bar-theoretic approaches."

[12] This remark applies with equal force to the labeling algorithm proposed in Narita (2010a), which stipulates that "Merge must take at least one L[exical] I[tem] as its input" (what Narita calls the "H-α schema"), and takes LI to be the label of the relevant structure. On the stipulative character of other labeling algorithms proposed in the minimalist literature see Boeckx (2008b, 80–84).

[13] For additional discussion and criticism of Chomsky's labeling algorithm, see Richards (2009b).

dynamic, or derivational perspective on 'Asymmetry.' Instead of seeing asymmetric labeling as merely the register of an asymmetric input relation (be it a featural one, as in Collins (2002), or a 'lexical' one, as in the traditional $\overline{X}$-schema or in Chomsky's recent labeling algorithm), I will argue for a notion of asymmetric labeling that emerges (in the course of the syntactic derivation) from a symmetry-breaking process. This process will amount to a *partial* (and therefore asymmetric, or, better said, symmetry-breaking) transfer of the structure formed by merge to the external systems responsible for interpretation and externalization. This process has been independently argued to take place periodically (cyclically), and will be said here to give rise to a regular structuring rhythm (transferred structure, non-transferred structure, transferred structure, non-transferred structure) that will essentially regulate ('self-organize' might actually be a better term) what would otherwise be a totally unconstrained merge process. This cyclic transfer will give rise to enough points of asymmetry (the points at which a distinction arises between transferred and non-transferred portion of structure) on which to anchor all the asymmetries that were thought to require projection or features. Put differently, the hypothesis pursued here can be stated as follows (see already Boeckx (2009c, 43(3))):

(4) Phases are the sources of all asymmetries found in Universal Grammar

Such a statement in turn invites the following conjecture:

(5) Because all asymmetries in Universal Grammar are rooted in phases (cyclic transfer), all of them will be 'interpretive'

That is to say, the effects of all these asymmetries will manifest themselves only once transfer has taken place, i.e., post-syntactically (in the interpretive components, where 'interpretive' here refers to both semantic interpretation and phonological interpretation). Put differently, syntax (merge) will be free of these asymmetries. Within syntax, symmetry will reign.

Before examining this conjecture more closely, let me focus on the notion of 'phase,' since it will play an important role in virtually everything that follows.

Chomsky (2000a) elevates the old idea (in generative grammar)[14] that linguistic computations proceed in a cyclic (i.e., compositional) fashion to a postulate, according to which syntactic derivations are divided into *phases*. In

[14] Tobias Scheer has pointed out in a number of publications that the principle of the Cycle in phonology and syntax, first formulated in Chomsky *et al.* (1956), appears to be a genuinely new idea in linguistics. For reviews of the treatment and scope of the Cycle over the years, see Freidin (1999), Lasnik (2006).

Chomsky's writings, phases consist of a designated element (the *phase-head*), whose complement (the so-called complement domain) gets transferred to the interfaces (either immediately before a new head is introduced (Chomsky, 2000a), or else as soon as the next phase-head is introduced (Chomsky, 2001), a point I return to below). Everything apart from the complement domain of the phase (i.e., the phase-head, as well as the specifiers and adjoined material above it) constitutes the edge domain.

Intuitively speaking, phases impose a locality condition on syntactic computation (the same way the principle of compositionality imposes a locality condition on semantic computation): cyclic Spell-Out allows the syntactic component to keep a minimum amount of information (the edge) in 'working' memory. The rest (the complement domain) can be ignored for purposes of syntactic computation.

Over the years, Chomsky has offered a variety of suggestive arguments in favor of a phase-based architecture of narrow syntax (for an exhaustive list, see Gallego 2011b). To my mind the most compelling argument provided by Chomsky comes from the 'logic of feature-valuation' (as one would expect from a lexicocentric approach). This logic relies on two assumptions adopted in Chomsky (2000a). First, featural interpretability is not a relevant (i.e., accessible) dimension within narrow syntax. (This seems desirable, if look-ahead is to be avoided.) Second, feature-value (by hypothesis, a property accessible to narrow syntax) provides an instruction to the SEM component not to interpret the feature in question on the element bearing it. Chomsky (2001) observes that because features can come to be valued in the course of the derivation via *Agree*, it is imperative to synchronize valuation and transfer[15] to ensure that the right instructions reach the external systems.[16] This in effect requires Transfer to take place every time valuation does (effectively forcing unvalued features to be confined to phase-heads; Chomsky 2007, 2008, Richards 2007).[17] (Note

[15] Actually, it is also important to make sure that valuation and transfer be mapped to both SEM and PHON at the same time. The logic of valuation thus argues against proposals like Marušič (2005) or Richards (2008b), where not the same portion of the tree has to be transferred to both SEM and PHON. I return to this issue below.

[16] Note that the logic of valuation and inheritance does not leave room for options like those explored in Ouali (2007) or Legate (2011) (situations of underinheritance, and the like), which in addition rely on the notion of feature bundle, unavailable in the present framework.

[17] One should not conclude from this relation between Transfer and Agree that all phase-heads must carry unvalued features or that Transfer only takes place when Agree does. Other considerations may require cyclic Transfer (e.g., see Uriagereka 1999), creating a distinction between phase heads with and without unvalued features, which may relate to the strong–weak phase distinction introduced in Chomsky (2001). For relevant discussion, see the passages below in the main text; see also Chomsky (2012a, 6), where it is said that "among the phases are the syntactic objects in which structural case and unvalued ϕ-features are valued" (the use of the

that because the process of valuation is asymmetric, it would have to take place at the phase-level, in accordance with the hypothesis (4) above.)

As befits a lexicocentric model like the one assumed by Chomsky, the argument for phases based on valuation relies on the syntax manipulating feature bundles, and as such it is not literally available here (some version of the argument will turn out to be salvageable, though, as we will see below). The argument in favor of phases I'd like to put forward is simply that if cyclic Spell-Out is assumed, asymmetries that used to require projection, features, or stipulative algorithms, can be made to follow naturally in a feature-less framework like the present one. Put differently, in the present context, phases provide a rather natural mechanism of symmetry-breaking.[18]

Quite apart from all the asymmetries mentioned above, there is a good reason to break the symmetry of merge expressed in (1). As extensively discussed in Richards (2010), the external systems with which syntax interfaces do not appear to tolerate elements that are too similar to one another to be too close to one another.[19]

It is also worth noting that van Riemsdijk (2008) relates this Anti-Identity Avoidance ban to Relativized Minimality, which, interestingly, as Richards argues for Anti-Identity, also appears to hold at the phase-level; cf. Chomsky (2001). According to Richards, English quotative inversion offers a relevant example of this ban. In quotative inversion, a quote appears at the beginning of the sentence, and the subject is postverbal, as in

(6) "It's cold," said John.

In such examples, the verb may be followed by material other than the subject:

preposition 'among' entails that there are 'other' phases besides those associated with unvalued features).

18 The use of symmetry-breaking advocated here can be seen as a radical extension of Moro (2000), who was the first to investigate the possibility of symmetry-breaking in syntax, but who limited its use to situation of small clauses (merge partners bearing an identical label), and who resorted to movement to break the symmetry of Merge – an option that we should be suspicious of.

It may be useful to point out that symmetry-breaking is by now a well-established and much-studied process, central at fundamental levels of physical explanations, and essential in the context of complex, dynamical, dissipative systems. As such, it certainly qualifies as a third-factor effect, to use the terminology of Chomsky (2005).

19 van Riemsdijk (2008) reaches conclusions similar to Richards, although he is less specific about the domain over which the ban applies. I should also point out that the "Identity avoidance" condition, as van Riemsdijk calls it, may well be a reflex of the more general Anti-locality condition put forth and discussed in Grohmann (2003). See Boeckx (2008b, c) for remarks along these lines. See also Arsenijević and Hinzen (2012) for the formulation of a constraint that I would argue reduces to Richards's ban.

(7) "It's cold," said John to Mary.

However, the postverbal material may not consist of multiple DPs:

(8) *"It's cold," told John Mary.

As this example shows, the ban at issue is more abstract than typical situations of haplology, which bans identical-sounding elements from being adjacent. Richards shows that his ban is about elements of the same abstract syntactic category. It's not the case, of course, that elements that are quite similar to one another can never appear next to one another (think of double object constructions like *John gave [Mary a present]*). To allow for this, Richards argues that the ban is local: it only holds 'within a phase'; more accurately, within the transferred portion of the phase (the complement domain). In a phase-based model like Chomsky's briefly summarized above, this makes sense: the external systems with which syntax interfaces only receives a portion of the syntactic derivation at any given time. It makes sense that whatever condition they impose on the syntax holds only for the portions they receive. Accordingly, Richards's ban can be formulated as follows:

(9) No two elements of the same syntactic category can be transferred at the same time

My reason for discussing Richards (2010) here is that it suggests that if it were not for phases, all outputs of the syntax we have considered until now would crash, since all lexical items are syntactically identical (one could say that they are of the same category: all are lexical precursor cells). Thanks to phases, it is at least possible to allow for some combinations of identical elements not to crash. Specifically, Richards's ban suggests that phase boundaries are a way to make identical elements distinct (for purposes of the external systems). To put it differently, the external systems will tolerate a syntax that combines α and β (two elements of the same category) so long as a phase boundary intervenes between them. This is another way of saying that transferring both α and β as soon as they have been merged would lead to a crash.

By means of this reflection, we have reached two important conclusions about transfer in a model of syntax where the pre-syntactic, narrow lexicon is homogeneous:

- *A conclusion about timing of transfer; "when"*:
  Transfer must happen more than once to avoid crashes caused by Richards's ban. In other words, transfer must be cyclic.

- *A conclusion about the amount of material transferred; "what"*:
  Transfer should not be too greedy; it should not transfer more elements than can be distinguished by the external systems at each Spell-Out point (i.e., cycles must be small).

Conclusion two is, of course, more specific than conclusion one. It says, for example, that upon merging the first two elements in a derivation, α and β, if these are to count as distinct elements (a natural recoverability condition on the derivation), a phase boundary must exist between the two. A natural conclusion, given Chomsky's conception of phase (specifically, his distinction between phase complement and phase edge), would be to take the element transferred (say, β) as the complement domain, and the other element, α, to correspond to the phase edge of what one could call the initial phase. Since minimally the phase edge consists of a phase-head (specifiers and adjuncts being optional), we could easily reconstruct what Narita (2010a) has called the H-α schema (the traditional X^0–XP distinction) as follows: upon transfer, β is the complement, α is the head, not in virtue of their lexical properties (both are atomic units, indistinguishable lexical precursor cells), but solely on the basis of the fact that one of the elements is transferred and the other not.

Notice that just like transferring both α and β is not an option (it would not allow the external systems to distinguish them), transferring neither is not an option either, for delaying transfer would unavoidably lead to a violation of Richards's ban at a subsequent stage. If it is hard enough to keep two elements distinct, imagine what would happen with more elements involved. This is another way of saying that on the basis of what we have seen so far, merge should be binary, as it is the only way to let phase boundaries reproduce distinctions that were lost once the pre-syntactic, narrow lexicon is as impoverished as I assume here.

Incidentally, the fact that Richards's ban prevents syntax from transferring the entire output of merge makes independent sense. It stands to reason that if Transfer means 'no longer active for further computation' (the only meaning of Transfer that would not void phases of any content, or rob the concept of phase of any desirable consequences regarding computational complexity reduction), the entire output of Merge should not be transferred as it is formed. If it were, syntax would then be stillborn and Markovian (Merge would be equivalent to Concatenate/string-formation). At least one member of the Merge set must remain accessible for further computation (external Merge), if syntax is to be able to construct hierarchies.[20]

[20] This conclusion agrees with Chomsky (2007) against taking all phrases to be phases (a position entertained in Bošković (2007); Epstein and Seely (2002); Müller (2010)).

2.3.2 Categorization

Returning to the first merge situation considered above, $\{\alpha, \beta\}$, I said that a phase boundary between the two could easily reproduce the X^0–XP distinction, with α the head and β the phrasal complement. But this head–non-head distinction is not completely equivalent to the one expressed in the traditional $\overline{X}$-schema, because in the present system, both merge partners lack categorial information. Assuming that the external systems demand at least some minimal categorial specification, as seems plausible, it would be desirable to find a way to assign a label to the spelled-out unit. Fortunately, the relevant mechanism is already in place in the literature. To the best of my knowledge, Marantz (2000) was the first to suggest that roots (which readers will remember are a-categorial units in the pre-syntactic lexicon) could be categorized at the phase level (see also Marantz (2008)). Building on Chomsky's (1995) "little υ" hypothesis, according to which the traditional VP consists of two layers or shells – a υP dominating a VP – Marantz suggested that all lexical categories be reinterpreted as roots labeled by the functional, "little x" layer dominating them.[21] Accordingly, a root $\sqrt{X}$ will become an N(P) in virtue of being dominated by a n-(phase) head; or an Adjective(P) if dominated by a a-(phase) head.

This labeling-by-phase mechanism could be exploited in the present framework to assign a category label to β, with α taking on the little x role in Marantz's system: thanks to phases, not only would α and β count as distinct, but in fact, one could serve as the label of the other. But what sort of label would α give to β? In Marantz's system, the functional portion of the narrow lexicon is rich enough to provide the system with different labels[22] (υ, n, a, and perhaps others like p), but the pre-syntactic lexicon considered here does not allow us to make such distinctions. β being labeled by α is equivalent to β being labeled by, say, γ. If the pre-syntactic lexicon only contains one kind of element, it can only provide one kind of label, but a system with just one label is about as interesting as a system with no label at all. Labels, like features, are only useful if there are more than one of them, since the whole point of labels is to make distinctions. Minimally, there must be two labels for a labeling system to be worth existing. The issue, then, is how to squeeze at least two labels out of the present system.

If one is to avoid lexicocentrism at all costs, as I think we should, the only way to achieve the desired result is to rely not on features, but in fact on insisting along the same path we used to already reconstruct the minimal distinction

[21] Borer (2005) also proposes that her a-categorial lexemes are labeled by the functional structure associated with them, but Marantz was, I think, the first to combine this idea with phases.

[22] For Marantz, these are distinct functional feature bundles.

between head and complement, that is, build on the range of configurations offered by phases. To do this, it will be necessary to look a little bit beyond the first merge situation we have examined so far. Recall that by inserting a phase boundary between α and β in $\{\alpha, \beta\}$, both acquire an identity: β is labeled by α and in turn α is the label of β. Now consider a situation where a third lexical precursor cell γ is merged to α. Since α already has an identity (a label) in virtue of labeling β, it could count as sufficiently distinct (derivationally) from γ even in the absence of a phase boundary between them. We could then proceed to add a third lexical precursor cell δ, yielding $\{\delta, \gamma, \alpha\}$. Here a phase boundary must be inserted between δ and γ because we are back to the situation we faced when we first merged α and β: neither δ nor γ are labeled by or act as the label for another element. If we were to transfer the unit $\{\delta, \gamma, \alpha\}$, the derivation would crash due to Richards's ban. Inserting a phase boundary between δ and γ suffices to let the derivation proceed. Taking δ to be the phase-head, the rest of the unit $\{\delta, \gamma, \alpha\}$, *viz.* $\{\gamma, \alpha\}$, will count as the phase complement. Since α already has an identity, and identities are best kept unique and exclusive if they are of any use,[23] γ will be labeled by δ (in turn, δ receives its identity as the label of the set $\{\gamma, \alpha\}$).

Notice that whereas nothing forced merge to be binary in the case of merging δ, γ, and α the need to insert a phase boundary between δ and the two other units yields a binary branching structure. This may not be all that surprising, given that the method pursued here to comply with Richards's ban is one that relies on the logic of unambiguous paths put forth by Kayne (1984) to motivate binary branching in the first place. In the present context, it would be legitimate to talk of unambiguous categorization paths (which must be 'exclusive' and unique, i.e., unambiguous if they are to be of any use, in terms of unambiguous derivational paths/configurations).

So far, it looks like we have not learned anything more than what we already knew from examining the set $\{\alpha, \beta\}$, but I would like to argue otherwise. What the scenario considered in the previous paragraph tells us is that there can be at least two 'kinds' of phase-heads, if we take into consideration the configurations that they label. In the case of first merge, involving α and β, the phase-head α labels the singleton β. In the case of δ above, it labels a two-member set ($\{\gamma, \alpha\}$). Adopting the terminology I introduced in Boeckx (2012a), one could say that there are 'transitive' and 'intransitive' phases, based on the nature of the complement domains they give rise to (two-membered sets and singleton sets, respectively).

[23] See Watanabe (1995) on this undesirability of erasing/rewriting information gained at an earlier stage in the derivation in a cyclic system.

The two-category system that emerges from the present discussion is reminiscent of the conception of nouns vs. non-nouns in Kayne (2011). Kayne's starting point is the tension between a bare phrase structure system like the one in Chomsky (1994) or his own (1994) Linear Correspondence Axiom (LCA), which maps asymmetric c-command onto precedence. Kayne notes that situations where an atomic lexical item merges with an independently formed phrase poses no problem for the LCA, even outside $\overline{X}$-theory. A labeling algorithm that takes the atomic unit to be the head, as in Chomsky (2007, 2008, 2013), will yield enough asymmetry for the LCA to apply. The problem arises when two atomic lexical items merge. Such structures are too symmetric to map onto precedence. Instead of resorting to the solution in Moro (2000) in terms of movement of one of the atomic elements, Kayne suggests, building on a suggestion by M. Guimarães, we let one of the two atomic elements merge with itself (forming the singleton set {LI}) prior to combining with the other atomic element.[24] This self-merge step in effect turns one of the atomic elements into a phrase, allowing the merger of two atomic lexical items to comply with what Narita called the H-α schema, and with the LCA. Kayne (2011) goes on to point out that "in a given derivation, some lexical items x will appear as part of {x}, others (the 'y's) will not. That is, some lexical items will be involved in singleton set formation, others will not." By means of a series of assumptions, Kayne ends up concluding that the lexical items involved in singleton-set formation are of the noun category, which, according to him is the only truly open-class category in natural languages. Kayne's reasoning is as follows: assuming, with Chomsky (1995, 2001), that unvalued features must be valued as soon as an element enters into the derivation, elements involved in singleton-set formation cannot enter the derivation bearing unvalued features, for self-merge/singleton-set formation would be incapable of valuing these, violating the requirement of immediate valuation. Assuming, with Collins (2005), that parametric variation is confined to the domain of unvalued features, and that parametric variation is limited, hence confined to the domain of functional/closed-class items (see, among many other publications, Kayne (2005)), it follows that elements forming singleton sets must be open-class items. Kayne proposes that the category noun matches the

[24] Strictly speaking, Kayne departs from Guimarães's suggestion, as he does not consider the possibility of self-merge ("which leads to questions about how to distinguish occurrences of x and what to say about 3 or more x's merging all at once"), but rather assumes that "one option for merge, taken to be set-formation, is the direct formation of the singleton set {x}."

On self-merge, see also Adger (2013b).

characteristics of the class of elements forming singleton sets.[25] Notice that Kayne takes this proposal to entail that verbs belong to the closed-class category (equivalently, all verbs are, at bottom, light verbs), and points out that this is a rather natural conclusion, considering that many verbs are arguably the result of combining a noun and a light ('functional,' closed-class) verb; cf. Hale and Keyser (1993) on [$_V$ *laugh*] as [$_V$ DO [$_N$ LAUGH]]).

One need not resort to self-merge or singleton-set formation, nor make all the assumptions Kayne makes to let the syntax generate singleton sets for the external systems to handle. As we saw above, singleton sets naturally emerge from the phase boundaries required upon first merge between two lexical precursor cells. It does not follow that these single-membered complement domains of first phases will be treated as nouns (but notice that it does not follow under Kayne's reasoning either; that singletons are nouns is his conjecture). All that follows is that such singletons could form a natural class post-syntactically. It is, of course, quite natural to expect this class to be 'nominal' in character, if only because, unlike verbs or adpositions, nouns tend to lack argument structure specifications, and are thus the prototypical units that can appear in isolation (as singletons). (Many of them tend to correspond to what the psychology literature calls 'Spelke objects.') Ultimately, though, categories like nouns and verbs are notions that make most sense in the context of the specific languages, as such categories tend to have an irreducible morphological aspect to them, so I am not convinced that we should insist upon shoehorning the class of intransitive phase complements (singletons) into the language-specific category 'noun,' though I may agree that such singletons, given their derivational history, are, as it were, predisposed to manifest nominal behavior post-syntactically.[26]

2.3.3 More (on) categories

What the present category system also shares with Kayne's proposal is that it generates only two natural classes: the intransitive and the transitive phasal

[25] Here Kayne agrees with Hale and Keyser's (1993, 2002) treatment of nouns as non-relational elements (i.e., elements devoid of any argument structure). For a relevant overview of Hale and Keyser's system, see Mateu (2005).

[26] One may perhaps talk, as I did in the 2010 "Defeating lexicocentrism" manuscript version of this chapter, of "proto-noun" and "proto-adposition," in analogy with Dowty's (1989) "proto-θ-roles," which gave rise to a more elaborate range of thematic relations once the lexical meaning of specific predicates were taken into account (i.e., at a later stage in the derivation). But I am no longer sure that even that terminology is necessary.

complements.[27] Any super-transitive options necessarily reduce to the transitive one, given the need to define unambiguous structural contexts for categories, which yields binary branching, and, as is well known (see Larson (1988)), binary branching decomposes super-transitive structures into a succession of transitive 'shells' or layers. Accordingly, whereas intransitive phasal complements may be predisposed to behave like nouns, no such clear-cut prediction can be made when it comes to transitive phasal complements, given the wider range of traditional categories they could map onto: verbs, adpositions, adjectives, to say nothing about the many functional categories posited in cartographic proposals beginning with INFL and COMP in the early 1980s. Alternatively, this situation invites us to conceive of all these functional categories as abstractly the same entity. That is to say, the present system begs the following question: in the absence of (necessarily language-specific) morphological cues, could we tell apart a verb from a preposition, or a complementizer from a determiner, or a tense marker from an adjective? Put yet another way, perhaps the language faculty (Universal Grammar) does not make these finer-grained distinctions. Certainly, the proposal here is that the pre-syntactic, narrow lexicon does not make any such distinction, and that syntax can only generate a two-category distinction.

Before pursuing this issue of categorial impoverishment further, I should perhaps point out that the two-category system of the present framework differs from Kayne's interpretation of his two-category system in at least one important respect. By bringing the closed-class–open-class distinction into the discussion, Kayne seems to interpret his singleton vs. the rest distinction in terms of the traditional lexical vs. functional divide (with nouns for him forming the only truly lexical, open-class category). This is not how I wish to

[27] One empirical consequence shared by both proposals is that structures typically described as Noun–Noun compounds must be analyzed as consisting of two intransitive phases separated by a transitive phase, or of a structure where one Noun is adjoined to the other, as in (i).

(i) $[[p\ [\sqrt{P_{null}}\ [_{nP}\ n\ [\sqrt{N}]]]]\ n\ [\sqrt{N}]]$

(In this representation, I am using standard labels like *n* for the intransitive phase-head, N for the intransitive phase complement, and *p* and P for the transitive phase and phase complement, respectively.)

Note that the presence of a phonetically null adpositional layer may account for the semantic underspecification of many N–N compound relations (e.g., *banana box*: box *for* bananas? box *(made) out* of bananas? box *in the shape of* a banana?), for the adjectival/modifying role of one of the Ns, and also may provide a way to understand the link between availability of productive N–N compound and double object structures (Snyder (1995, 2001)), where covert Ps have also been posited (see, among others, Pesetsky (1995)).

understand the transitive vs. intransitive phase complement distinction. In the present system the distinction that somewhat replicates the lexical–functional divide is the distinction between the element in the phase complement being labeled by the phase-head, and the phase-head labeling the previously unlabeled unit in the phase complement. That is, I follow Borer and Marantz in taking functional elements to 'exoskeletally' label lexical material. Although I do not assume the existence of pre-syntactic differences like *n* vs. υ, there is a sense in which the present system replicates the (lexical) N vs. (functional) *n* distinction when it takes β to be labeled by α in the context of an intransitive phase. The same holds of the (lexical) V (or P, or A, ...) vs. (functional) υ (or *p*, or *a*, ...) distinction, when γ is labeled by δ in the context of a transitive phase. By contrast, Kayne's system does not resort to a functional nominal category like "little *n*" to generate Nouns; instead, it resorts to self-merge.

In a certain sense, the present proposal is more in line with traditional grammatical category systems than Kayne's, since the transitive–intransitive phase complement distinction generates the equivalent of a two-*lexical*-category system (cf., e.g., the [±N; ±V] system of Chomsky (1970)). Unlike Kayne, it does not collapse lexical category distinctions and the functional–lexical category distinction. Put another way, it allows for the existence of light (functional) nouns, alongside light non-nominal functional categories, which, I think, is empirically more adequate.

To approximate current grammatical category systems (earlier systems like Chomsky (1970) did not yet worry about the range of functional categories like C, T, and D that are now familiar to syntacticians), the present system needs to resort to what could be called the 'logic of grammaticalization.' The process of grammaticalization is one by which lexical categories over the course of the history of a language turn into[28] functional elements (presumably, due to repeated use, leading to semantic bleaching and phonological destressing). What is particularly interesting in the present context is the cross-linguistically well-established result in Heine and Kuteva (2007), where it is shown that most functional categories found in languages can be traced back to lexical items that once used to lexicalize 'nouns' and what they call 'verbs,' although I think that 'adposition' would be a better term for the latter, as Heine and Kuteva explicitly state that the verbs they have in mind are non-inflected for tense (a defining property of adpositions vs. verbs; see Svenonius (2007)).

[28] The term 'turn into' is not to be understood literally; 'come to lexicalize' may be more appropriate in a generative context. It is not the case that the new category is created *de novo*; rather, a lexical item that typically lexicalizes a lexical category comes to lexicalize another, more functional category.

Be that as it may, what is important in Heine and Kuteva (2007) is that patterns of grammaticalization reveal that a two-category system is sufficient to generate all the other categories. Based on this, I am tempted to propose that in the post-syntactic morphological distinctions, the two-category system made available by the bare-bones syntax examined so far suffices to give rise to enough flavors to yield morpho-syntactic categories like adjectives, complementizers, tense markers, pronouns, determiners, conjunction markers, and the like. In other words, I am led to propose that all the functional categories one finds in natural languages are to be defined derivationally, as configurational variants (specializations) of the two basic categories defined by phases in the syntax (transitive and intransitive phase complements). (I should note that Heine and Kuteva make this claim in the domain of diachrony, but the present claim pertains to the domain of synchrony, or, better put, the domain of derivational dynamics.)

Whereas this hypothesis differs significantly from cartographic proposals whose lexicocentrism leads them to assume that all their fine-grained categorial distinctions are equally primitive, the present proposal can be seen as a generalization of certain ideas already found in the literature. For instance, several authors (Amritavalli and Jayaseelan (2003); Mateu (2005); Kayne (2011)) have already argued in favor of regarding adjectives as non-basic, and viewing them instead as (possibly phoneticaly null) adposition + noun sequences (*angry* would be [$_P$ WITH [$_N$ ANGER]]). As mentioned above, others (e.g., Svenonius (2007)) have pointed out that morphological cues aside, it is hard to tell apart verbs from adpositions. As a matter of fact, it has been argued (see Aboh (2009)) that in some morphologically poor languages, serial verb constructions, traditionally claimed to consist of V–V sequences (with the first V possibly a light verb), are better analyzed as P–V sequences.[29] In his study on lexical categories, Baker (2003) in fact is led to argue that verbs may not be basic and may always bottom out as adjectives, which, given what I just said, may then turn out to be further decomposable into an adposition and a noun.[30] In a similar vein, Leu (2008) suggests that Demonstratives and, by

[29] Interestingly, non-finite verb forms, such as participles and gerunds, have been argued to be P-like (see Gallego (2009); Masullo (2008); see also Emonds (2008) for relevant discussion). It is also worth mentioning that the event/aspect structure and adpositional structures have been argued to match perfectly (Ramchand (2008); Tungseth (2008); see also Cuervo (2003) on applicative (prepositional) classes matching verbal aspectual classes).

[30] If indeed all verbs bottom out as non-relational categories, it follows that even internal arguments of verbs must be introduced by a light verb/applicative head. That is to say, all arguments of verbs are external. On this separationist hypothesis, see Lohndal (2012); Schein (1993, 2002); Williams (2009).

extension, other exponents of the (definite) "Determiner" category, may be part of the extended projection (in the sense of Grimshaw (1991)) of adjectives. Leu (2012) even goes further in treating indefinite determiners like German *ein* ('a'; 'one') as consisting of the preposition *in* + some nominal element. We could in fact argue that complementizers like *that*, which some (Kayne (2011), Manzini and Savoia (2011)) have argued to be demonstratives, are nothing more than extended adjectives, which would in part revamp (or turn around) early transformational treatment of adjectives as reduced relative clauses. Be that as it may, all of these proposals, and the many more I could have cited from the literature,[31] really suggest that the elaborate categories that figure in many generative studies may be reducible to a few (possibly, two) primitive ones.[32]

Chomsky himself has, in recent years, provided arguments for reducing the stock of primitive categories, beginning with V, which are nothing more than a root dominated by a light verb υ,[33] and on to T(ense) (or Infl), which he has argued gets its content from the category C dominating it.[34] If this is on the

The present system can certainly model this fully separationist approach, but does not require it. It is equally compatible with Chomsky's less articulated two-verbal layer structures.

[31] See, for instance, Martín (2012) on decomposing (dative) pronominal clitics (traditionally, of category D) into more primitive N and P categories.

[32] To add but one more example, it is certainly plausible to take coordinators (say, *and*) as morphologically disguised adpositions (say, commitative *with*).

[33] Several authors, including Pylkkänen (2008), Harley (2013), have argued that Chomsky's υ, which in his writings (beginning in Chomsky (1995)) fulfills two roles (categorizing the root of its complement as V, and closing off the verbal domain by introducing the external argument in its specifier), should be split into two categories, a root-categorizing head (υ), and a Voice head introducing the external argument. If this is indeed the case (Pylkkänen (2008) argues that this is actually a point of parametric variation), the present system would require the presence of an intermediate head, separating the phase-head corresponding to 'Voice' from the one corresponding to 'υ,' much like T separates C from υ in Chomsky's system. Perhaps one of the aspect heads identified in the literature on VPs would be this category (alternatively, Baker and Collins's (2006) Linker foreword).

[34] Here are relevant passages:

> [F]or T, ϕ-features and Tense appear to be derivative, not inherent: basic tense and also tenselike properties (e.g., irrealis) are determined by C (in which they are inherent: "John left" is past tense whether or not it is embedded) or by selecting V (also inherent) or perhaps even broader context. In the lexicon T lacks these features. T manifests the basic tense features if and only if it is selected by C (default agreement aside); if not, it is a raising (or ECM) infinitival, lacking ϕ-features and basic tense. So it makes sense to assume that Agree and Tense features are inherited from C, the phase head. (2008, 143–144)

> There are further reasons for expecting that TP is not a phase. T has the basic properties of uninterpretable features. It may yield a phonetic reflex, but its ϕ-features are determined by the context, so it should enter the lexicon without values for these features. T bears these features if and only if it is selected by C, hence it should inherit

right track, T could then be seen as a root categorized by a dominating functional category.[35] The fact that C often takes the form of an adposition (see complementizer *for (... to)* or complementizers introducing adjunct clauses like *after*, *before*, *without*, *since*, etc.) leads me to think that the [C [T ...]] sequence may be merely a higher-order[36] [p [$\sqrt{}$ ...]] sequence, i.e., a specific manifestation of a transitive phase.[37]

Viewing certain categories as merely higher-order versions of some lower categories (independently of their specific names) is a line of inquiry I began in Boeckx (2008b, chap. 4).[38] The idea there was that the fine-grained categories

> these from C (...). The biconditional holds of embedded clauses, but it would make no sense to hold that in root clauses T has different properties. It therefore follows that root clauses must have C, even if it is unpronounced.
>
> What is true of agreement features appears to hold as well for tense: in clear cases, T has this feature if and only if it is selected by C, though C never (to my knowledge) manifests Tense in the manner of ϕ-features in some languages. If that is basically accurate, then there are two possibilities. One is that Tense is a property of C, and is inherited by T. The other is that Tense is a property of T, but receives only some residual interpretation unless selected by C (or in other configurations, e.g., in English-like modal constructions). [footnote omitted] One advantage of the latter option is that T will then have at least some feature in the lexicon, and it is not clear what would be the status of an LI with no features (one of the problems with postulating AGR or other null elements). Another advantage would be an explanation for why C never manifests Tense in the manner of ϕ-features (if that is correct). (2007, 20)

35 Contra Chomsky (2007, 21). Chomsky's claim that T has the basic properties of uninterpretable features does not seem quite accurate. Uninterpretable features are not interpreted, whereas T is, albeit its interpretation depends on the functional structure around it.

36 In Chomsky's spare functional representation of the clause, it would be a second order sequence, right above the lower p [$\sqrt{}$ ...] defining the verbal domain.

37 On C as P, see Emonds (1985). After all, isn't it quite natural to treat C as an element providing a location (i.e., a p-element) on a time line for an event (Ditto MoodP (another flavor of C), which locates an utterance in a possible world)?

38 Part of the evidence I used came from Butler (2004), who showed that Cinque's (1999) hierarchy consisting of over thirty distinct functional heads making up the Inflectional layer of the clause can be 'reduced' to recurring layers of only three more generic functional units (temporal, aspectual, and modal heads).

Other types of evidence come from, for example, Pylkkänen's (2002, 2008) approach to applied arguments, where the same category (ApplP, a 'flavor' of υ/p) becomes specialized depending on its merge-site, hence the existence of low applicatives, medial applicatives, high applicatives, super-high applicatives, etc. (see Tsai (2010)). Likewise, the nominal category Person may be regarded as high n (low n corresponding to Gender/ [+animate] Class). For relevant material about this, see Picallo (2006, 2008), Martín (2012). Similarly, the distinction between P_{Place} and P_{Path} plausibly reduces to high p and low p. Ditto for Force and Finiteness: high C and low C. (As a matter of fact, Rizzi (1997) conjectures that Force and Finiteness form a syncretic category, and become split only under the presence ("activation") of the Topic/Focus field.)

identified in cartographic studies emerged as instances of duplication[39] and divergence, a well-established process in biology (a refinement of Darwin's 'descent with modification,' in fact). As the derivation grows, so does the number of phase heads (specifically transitive phase heads). As the syntactic context for these Ps grows with each instance of Merge, it can be exploited morphologically.[40] T, C, υ, and so on become akin to what geneticists call paralogs (genes related by duplication within a genome). Whereas so-called orthologs retain the same function in the course of evolution, paralogs evolve new functions, even if these are related to the original one.

These remarks point to the fact that the (post-syntactic) morphological component of the human language faculty, like natural selection, constantly tinkers with the spare resources made available to it (by the generator, narrow syntax in the case of language), recycling[41] the same categories and adding morphophonological flavors to them, which have sidetracked linguists into thinking that these emergent classes of categories are primitives.

A different kind of argument for the (derivational) recycling of basic categories comes from looking at recent studies questioning not the existence, but the role assigned to certain categories. The argument is, I think, clearest in the case of the category D(eterminer), so I will use D as my focus in what follows, but the discussion will lead me to extend the argument to several other categories. (To facilitate the reader's task, I will resort in what follows to standard labels like N, *n*, and D 'as if they existed syntactically,' but I ask the reader to bear in mind that this is something of a shortcut; all there is in syntax is a two-category system that knows nothing about nouns, verbs, and the like.)

39 For instance, it is not too far-fetched to view adverbials like *angrily* as double adjectives: [angrily] = [IN ANGRY MANNER] = [IN [WITH ANGER] MANNER]. (Note that the English adverbial suffix *-ly* is in fact historically derived from the adposition *like*.)

40 Notice that I say 'can,' not 'must.' Those languages that decide not to exploit these contexts morphologically will then recycle the same functional head over and over again. This appears to be what happens in languages that are said to have a poor inventory of functional categories, such as Mandarin Chinese.

It is in fact quite significant that in well-studied languages like English, loss of overt inflectional morphology blurs the distinction between many categories, such as C and P (again, think of non-finite complementizer *for*), T and P (think of non-finite T *to*), Aux and P (the auxiliary *of* discussed in Kayne (1997)) – all of these revealing the P-nature of all the other categories, exactly as the present system predicts.

Conversely, the richer the language is morphologically, the more salient these functional specialization of category occurrences will become.

41 The idea of morphological recycling is not new; see Longa *et al.* (1996, 1998). I would like to add that the tinkering character of morphology may have interesting implications for phylogenetic studies of the human language faculty.

The present system takes N/*n* to be the core nominal category ("which is what we intuitively always wanted to say" (Chomsky (2007), 25–26)). But what is the role of D in this system? If the theory sketched here is correct, D must be regarded as introducing a transitive phase.[42] It must be a species of "*p*," on a par with *v*, C, etc. There is nothing 'nominal' about it.

Although widely adopted, and deeply implicated in matters such as argumenthood (see Longobardi (1994)), the DP-hypothesis has always been problematic for selection and for concord (see Bruening (2008) for a careful survey of problems; see also Fukui and Zushi (2008)). The selection problem is particularly salient (although not confined to the nominal domain, as I will discuss momentarily; see Shlonsky (2006)): we want the Verb/Preposition to select N, not D (no verb is known to impose selectional restrictions on D). But if D dominates N, as in the DP-hypothesis, why doesn't it block selection of N?

Taking advantage of the hypothesized *p*-source of D and C, I would like to argue that D is in fact a kind of C. Although taking D to be a complementizer is not completely unheard of (see Szabolcsi (1984), among others), I would like to further argue that D is a relative complementizer, and adopt Kayne's (1994) revival of the raising analysis of relative clauses to allow for N/*n* to raise from within the DP and reproject[43] an NP/*n*P layer, rendering selection by V/P straightforward, as represented in (10).[44]

(10) $[_{nP}\ [_{DP}\ \mathrm{D} \ldots]\ [n\ [\ \sqrt{N}\ ]]]$, where DP = $[_{DP}\ \mathrm{D} \ldots [_{nP}\ n\ [\ \sqrt{N}\ ]]]$

Note that the derivation in (10) not only solves the selection problem, it also allows us to treat DP-internal concord (controlled by N/*n*, not by D) as a case of standard agreement, established in a manner similar to agreement at the clausal level. Upon raising inside what is now claimed to be a relative clause, *n*P established agreement/concord, much like a subject nominal triggers verbal agreement inside a clause. (Incidentally, the proposal also captures the oft-noted dependency between D and Person (see, e.g., Longobardi (2006) for the

[42] This means that standard treatments of clitics as intransitive D categories must be rethought. In the present model, clitics must be of category '*n*.' While some clitics may be 'bare' *n*Ps, others may contain a DP-layer inside them. This difference may be one way of understanding the contrasts between accusative and dative (or, more accurately, [−person] vs. [+person]) clitics put forth in Martín (2012), e.g., their doubling behavior.

[43] By 'reproject,' I mean that the identity of the label of the whole structures changes after movement of the relevant element at the edge of that structure.

[44] On NP-reprojection, from a different perspective, see Georgi and Müller (2010). For interesting material bearing on the present proposal, see also Koopman (2005) on Noun-formation via relativization in Masaai. See also Taylor (2009) for the claim that some instances of *the* in English (such as *the more, the merrier*) are complementizers.

claim that DP is really PersonP). If Person corresponds (as suggested above) to the higher occurrence of *n*, the one that reprojects out of the relative clause, the existence of Person (a higher *n*) entails the presence of C (the head of the relative clause out of which *n* reprojects). If PersonP is really the reprojected *n*P, it follows that DP-internal concord cannot involve person agreement, a robust generalization (see Baker (2008b)), one that is in fact often taken to militate against a single mechanism for concord and agreement.)[45]

It is worth noting in the context of the present proposal that the representation in (10) captures the well-established fact (see Heine and Kuteva (2007)) that determiners tend to grow out of (grammaticalize out of) demonstratives, which in turn tend to emerge from locatives. If D is C, and C is just a high occurrence of P, then D is ultimately P, a locative.

Leu (2008) has recently formalized this link between demonstratives and locatives by taking D(em.) to contain a locative element (which Leu takes to bottom out into an adjective, i.e., a PP for me). Leu's representation is reproduced for *this man* as follows:

(11) [$_D$ the HERE man] (with HERE silent in standard English)

Leu in fact suggests that phrases like *this man* contain a two-DP layer (roughly: [[*this* HERE] THE *man*], with HERE and THE phonetically null in most, but not in all languages. One could straightforwardly adapt Leu's proposal in the present framework, with Leu's locative element forming a low occurrence of P/D, and Leu's definite determiner forming a high occurrence of *n*. Treating (some) determiners as the Spell-Outs of the higher *n* may help us understand why in some languages classifiers (traditionally corresponding to a low projection, the most natural candidate being *n* in the present context, given that classifiers are light nouns) can function as high (definite) determiner categories (see Cheng and Sybesma (1999, 2005), Simpson (2005)).[46] The present proposal in fact anticipates differences among so-called classifier

[45] In the context of Person and a second occurrence of *n* (over D/P), it is worth noting the presence of a P-element (such as Spanish *a*) in the context of [+person/animacy] objects. Martín (2012) in fact explores the relationship between Datives and Locatives, and finds the presence of locative elements in a variety of person contexts, such as person inflection (Martín suggests that the final *y* of Spanish first person singular verb forms like *soy* '(I) am,' *voy* '(I) go,' etc. is really a Locative element, cognate of French *y* and Catalan *hi*), clitic composition (Latin first person singular dative clitic *tibi* as *t-ibi* 'you-there'), clitic syncretism (Catalan locative *hi* used in dative contexts), and so on. For additional discussion, see Boeckx and Martín (2013).

[46] They may even function as relative markers; see Jenks (2010).

languages, which appear to exist (see Jenks (2010); Jiang (2012); Ochi (2010); Saito *et al.* (2008)).[47]

It is tempting to extend the reasoning to solution to the selection problem that has plagued the DP hypothesis since its original formulation (Abney (1987)) to all the selection-related problems caused by the proliferation of functional categories, particularly in the context of cartographic proposals.

The problem is obvious, but rarely addressed in the literature. The only exception I know of is Shlonsky (2006), who discusses the uneasy relationship between cartography and minimalism (on the latter, see Boeckx (2008)). Let me quote the relevant passage in full:

> The maps drawn by cartography require some rethinking of the traditional division of the clause into a υ/VP, TP and CP domains. This is particularly relevant in light of the relatively simple structures that Chomsky's Minimalism works with. In part, this is a division of labor: Minimalism focuses on mechanisms of computation (Merge and Search) and the role of uninterpretable features, while the cartographic enterprise is primarily concerned with the inventory of interpretable features. Hence, under this view, minimalism needs an abbreviated structure, the C-T-υ-V system, while cartography explores the full representation (see (Chomsky (2001, n. 8)) for a comment to this effect). In practice, however, Minimalist research has adopted the C-T-υ-V system not merely as an "expository convenience" (Rizzi (2004, 7)), but as a substantive hypothesis for clause structure. The tension between Minimalism's impoverished structures and the richness of cartographic representations is a real one.
>
> Thus, phases (CP and υP) and their edges (i.e., their heads and specifiers) play a key role in the computation of locality in minimalist syntax. It is far from clear how to integrate these notions into the structural maps of cartography, in which the clause is typically seen as a homogeneous hierarchy of projections.
>
> In Cinque's system, for example, T dissolves into two distinct projections (Past and Future). Each should, in principle, have a specifier but which one corresponds to T? Similarly, what does "little *v*" [υ] correspond to in a cartographic articulation of lower aspect and event-type? Which one of these lower heads should be taken to constitute the edge of υP? The problem is just as acute in the CP domain, where the edge of CP is its (outer) specifier, but in a cartographic perspective, should it be equated with Spec/Fin, Spec/Force or perhaps Spec/Focus (which, according to Rizzi (1997), hosts wh operators)?
>
> Perhaps an even more nagging problem is that of selection. Minimalism inherits from previous approaches the view that selection is carried out under

[47] As Jeong (2011) shows, this casts serious doubt on the macro parametric treatment of DP/NP-languages advocated in recent works by Željko Bošković (see, e.g., Bošković (2005, 2008)).

> sisterhood. Thus, C selects T and V selects C. How is selection satisfied in e.g., an indirect question, if the head bearing the interrogative feature is Foc or Int (see Rizzi (2001)) and thus not a sister to V? Or take the familiar problem of how subjunctive features on an inflectional head can be selected by a higher predicate, given the number of intervening heads between V and the relevant mood head? The locality of selection plays a major role in Chomsky's most recent work (Chomsky (2008); see also Richards (2007)), in which C transfers Case features to T. Which C? Which T?
>
> The desirable goal of integrating the research agendas of Minimalism and Cartography requires, so it seems, modifications in the way structure, in the cartographic sense, is manipulated by the computational system. One possible direction would be to formally implement the notion of "abbreviated structures" by construing the cartographic structures not as a homogeneous hierarchy of equiponent projections, but as a structure composed of delimited "domains" or "extended projections," in the sense of Grimshaw (1991), a υP domain, a TP domain and a CP domain. Such a delimitation of structure is necessary for good empirical reasons, as well, since one needs to explain not only the clustering of similar features in the structure but also the delimitedness of verb-movement, NP movement and other operations which depend, in Minimalism, on uninterpretable features like Case or agreement. Why, one may ask, do these features typically fail to extend to the heads of left periphery? Cartographic works have, for the most part, implicitly assumed delimited structures or spaces but have not provided a formal implementation of domains.

Shlonsky's remarks are spot on. Take the nominal domain. Cartographic arguments have led to the explosion of the traditional NP into numerous projections, all dominated by a D-like layer (say, DP for the sake of concreteness; other labels would do just as well). This resulted in the material heading the traditional NP (the substantial Noun) being related to the most deeply embedded projection inside the DP, and yet everyone agrees that of all the elements contained in DP, it is the N that must be accessible for selection by, say, a verb. This is as clear a locality problem as one can find in syntax. Moreover, as Shlonsky correctly points out, cartographic representations have led to the scattering of traditional heads into several single-feature-based projections, making it hard for a particular head to be selected, for what one wants to say is that it is an ensemble of heads that is selected, a "domain," as Shlonsky calls it.

I would like to propose that all such selectional problems be solved by means of reprojection (head-raising analysis of relative clause), and that the multiple occurrences of the (re)projected head delimit the functional 'zone.'[48] Put differently, I propose we treat the potentially offending material for selection as

[48] On the typical size of such zones, see Boeckx (2008b), and Chapter 3.

being contained inside a relative clause, off the main projection line. To give an example, in addition to my reformulation of the DP-hypothesis, I would like to argue that when CP splits into ForceP and FinitenessP (i.e., two occurrences, and subsequent functional specializations, of CP), it is indeed, as Rizzi (1997) conjectured, due to the activation of the Topic/Focus field. The latter I would like to treat as relative clause material. Interestingly, Irurtzun (2007) has independently argued, following interpretive-based proposals by Herburger (2000), in favor of Focus-activation forcing reprojection of FinitenessP. Irurtzun's specific analysis is a bit too complex for me to go into here, as his treatment of focus, following Herburger, is not standard (though fully compatible with the interpretive properties discussed in Chapter 3), and would require too much of a digression. I merely want to note that Irurtzun offers an explicit rationale for why activation of the Topic/Focus field (Rizzi's insight) may lead to reprojection. I expect that other instances of split categories (e.g., the discussion of the splitting or bundling of υP and VoiceP in Pylkkänen (2008, chap. 3)) will be linked to the presence of intervening material that 'forced' reprojection. (I should note that in a Merge α framework, nothing is forced. Failure to reproject will lead to selectional problems and filtering at the interfaces. I return to this point below.) I also therefore predict that somewhere inside the intervening material one is likely to find relative-complementizer-like elements (ultimately, adpositions). This may well be how applicative morphemes, topic markers, and the like should be treated.

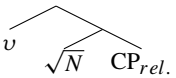
Splitting of a syntactic category via reprojection leads us to expect that languages that consistently do so will develop distinct morphological markers for the various occurrences of the category in question, much like the various occurrences of a given category will acquire distinct interpretive roles. At the interpretive level, we expect this morphological analyticity to force the activation of finer-grained conceptual distinctions: Person and Gender as opposed to Class. Ditto at the morphological level, where finer-grained morphological classes may fossilize (grammaticalize) over time. This ties in with the discussion in Béjar (2003), where values/attributes of higher-order feature types (say, Person and ϕ, respectively) are not treated as different, ontologically speaking. Attributes and values are simply finely cut features.

Apart from issues concerning selection, the discussion on D suggests that if D is C_{rel}, clausal complementation may also amount to relativization.[49]

[49] For relevant material and discussion, see Arsenijević (2009); Kayne (2010); Manzini and Savoia (2011); Stepanov and Stateva (2006). Caponigro and Polinsky (forthcoming), which is sometimes mentioned in the context of recent proposals arguing for a relativization analysis of clausal complementation, provides arguments that in some languages complementation is

Specifically, if verbs can really be light verbs, the root in their complement domain may be the root (nominal) being relativized. So, a verb like *claim* may easily be analyzed as a light verb MAKE and a root $\sqrt{CLAIM}$: (Traditional labels used)

(12)



The fact that it is a mere root that is relativized, as opposed to a functionally more loaded category (say, the equivalent of an *n*P or something even bigger) may account for why such instances of relativization do not give rise to Relativized Minimality/Complex NP Constraint-type islands,[50] and why verbs appear to impose selectional restrictions on (this type of) CP_{rel}.[51]

2.3.4 Interim summary

Taking stock of this long section, we can say that the approach to categories developed here sees morpho-syntactic categories (or features, which is the same thing, given that single features are now standardly taken to project their own phrases) as configurational flavors of very generic, substance-free, notions like transitive and intransitive phase complements. To a large extent, it agrees with the claim made by many functionalist linguists (see, e.g., the work of Martin Haspelmath such as Haspelmath (2010)) where morpho-syntactic categories in languages are not part of Universal Grammar. It also relies on a two-category distinction, not unlike the system adopted in cognitive linguistics (Langacker (1987, 1999)), although there categories have (cognitive) substance, whereas here they are purely syntactic/configurational, and only acquire substance post-syntactically.[52]

absent, and a relativization strategy must be used. Accordingly, their study should be seen as an argument that relativization and complementation should not be collapsed. But it seems to me that the cross-linguistic difference at the heart of Caponigro and Polinsky (forthcoming) can be easily accommodated even if we assume that complementation is relativization in all languages. The difference could lie in the nature of the nominal being relativized: a root nominal, or a functionally more complex nominal, with the latter giving rise to the phenomenon Caponigro and Polinsky (forthcoming) illustrate on the basis of Caucasian varieties. Among these, Arsenijević (2009) is the most explicit concerning the possibility that all instances of clausal complementation are instances of relativization.

50 Being least specified, roots are the most transparent items.

51 The selectional difference between the traditional categories CP and DP discussed in Bruening (2008) may then be a result of the type of category being relativized: $\sqrt{N}$ vs. *n*P.

52 In Chapter 3, I will agree in part with the cognitivist claim that the substance that these categories receive may not be specifically linguistic, and may be rooted in non-linguistic cognition.

Certainly the present approach ties in with Harbour's program (see Harbour (2006, 2009)), who has argued that aspectual systems, number systems, pronominal series, and more can be explained solely on the basis of very primitive (and, by hypothesis, non-human-specific) concepts like singularity,[53] addition, and augmentation, applied to the structures generated by Merge and restricted by the representational power of memory systems. Importantly, for this program to have a fighting chance, it is imperative to regard linguistic categories (number, person, case, etc.) as superficial manifestations of far more abstract, and conceptually more general categories. In Harbour's terms:

> a. In the domains I've explored (person, deixis, location, number), there turn out not to be any person/number/... features per se, but, rather, that the relevant features have a more general semantics of which person, etc., represent a usage that arises under a particular lattice embedding. (i.e., there's a degree of semantic underspecification in the predicates that the features denote, which permits the features to be applied to different semantic domains.) Particularly interesting case: there's near identity between aspectual features of Krifka (1992) and the 'number' features, and 'locative' features.
> b. Language design exhibits a type of economy envisaged by the Minimalist Program: not only do such traditional grammatical labels as inclusive and exclusive (dual and trial) dissolve into combinations of more abstract features (no news there), but so too do the categories of person and number themselves.
> c. [Feature] Geometries are otiose. Harbour (2006)

As I suggested above it is becoming clear that much of the richness, diversity, and specificity of our mental ontology should be approached in the way Darwin understood the richness, diversity, and specificity (adaptiveness) of life (bio-physical ontology): as the result of descent with modification (duplication and divergence) from primitive forms. Paraphrasing the closing passage of the *Origin of Species*, one could say that from such simple conceptual beginnings, endless features most beautiful have evolved. Crucially, these notions, generic, and distributed across nominal, verbal, adjectival dimensions, cannot all be ordered onto a unique functional sequence, thereby questioning the guiding idea of much current cartographic work.

Needless to say, many will doubt that such a minimal(ist) category system as the present one can reproduce the full range of categories that linguists have identified (not only C, T, but also C_{Force}, C_{Fin}, P_{Path}, P_{Place}). But it is important to remember that the justification for most of these categories is morphological in nature, and that they may have much in common with one

[53] On the non-human-specific character of the concept 'singular,' see Barner *et al.* (2008).

another syntactically. The very term 'flavor' used in the cartography literature in fact suggests that even these authors feel that there may be some underlying commonality behind the myriad instantiations one finds. It is also important to remember that no theory of what counts as a possible category is offered in the cartography literature. Saying that single features can project (Cinque and Rizzi (2010), Starke (2010)) won't do, until a theory of what counts as a possible syntactic feature is offered. Proceeding without a theory can be quite dangerous, as Borges warned us:[54]

> On exactitude in science
> In that Empire, the Art of Cartography attained such Perfection that the map of a single Province occupied the entirety of a City, and the map of the Empire, the entirety of a Province. In time, those Unconscionable Maps no longer satisfied, and the Cartographers Guilds struck a Map of the Empire whose size was that of the Empire, and which coincided point for point with it. The following Generations, who were not so fond of the Study of Cartography as their Forebears had been, saw that that vast Map was Useless, and not without some Pitilessness was it, that they delivered it up to the Inclemencies of Sun and Winters. In the Deserts of the West, still today, there are Tattered Ruins of that Map, inhabited by Animals and Beggars; in all the Land there is no other Relic of the Disciplines of Geography.

Let me take this opportunity to touch again on the cartography project some more.

In a very revealing interview (Cinque (2002)), Guglielmo Cinque contends that the major objection to cartography voiced at meetings or in linguistics articles ("Really, are there really so many projections?") is not very strong because "the important thing" is that "the number of projections is limited [as opposed to infinite]," and "non arbitrary." Cinque points out that when he began his investigations that led to Cinque (1999), "there was no overall idea of what the restrictions would be on how many projections you could propose, or what these projections actually were." He goes on to say that "When I started working on the issue, I decided to address the question in an empirical way." He asked himself "[w]hat kinds of notions get grammaticalized? I can imagine many semantic or cognitive notions that never get expressed grammatically (i.e., through suffixes, prefixes, particles, etc.)." Why aren't these notions grammaticalized? The cartography project aims at answering precisely this.

[54] Borges, of course, was not writing about syntactic cartography, but the general point he made is valid, I think. Borges's reflections seem to have been inspired by Lewis Carroll, who had one of his characters in *Sylvie and Bruno* saying "We now use the country itself, as its own map, and I assure you it does nearly as well."

As Cinque and Rizzi (2010) observe, "[t]here are precise limits to the observed cross linguistic variation, a fact which calls for a principled explanation." The cartographers want to find out "What the proper labeling of X and Y is" (X and Y being heads posited in the literature) and "How rich the 'right' functional structure of clauses could be."

Currently, the guiding hypothesis appears to be that "each morphosyntactic feature would correspond to an independent syntactic head with a specific slot in the functional hierarchy" (Cinque and Rizzi, 2010). That is, "[there is] a unique consistent order/hierarchy, imposed by UG." "Featurally complex heads may arise in syntax, but they cannot be 'atoms' of the syntactic computations: 'one (morphosyntactic) property – one feature – one head.'"

Following this reasoning, it should be no surprise to read the following: "The inventory of functional elements is much larger than is generally thought. To judge from Heine and Kuteva's four hundred or so independent grammaticalization targets, the number of functional elements must at least be of that order" (Cinque and Rizzi, 2010).

The question looming large, of course, in such enterprise is "Where does the hierarchy, and its universal properties, come from?" (Cinque and Rizzi, 2010). In their survey article, Cinque and Rizzi strike a cautionary note: "The question whether such universal hierarchies of functional projections are primitive objects of UG, or can be derived from interface or more general external conditions is important, but fundamentally orthogonal to the prior task of drawing their precise map, and perhaps not easily determinable at the present state of our knowledge."

Whereas no one ought to question the descriptive usefulness of any "attempt to draw maps as precise and detailed as possible of syntactic configurations" (Cinque and Rizzi, 2010), one wonders to what extent current cartographic approaches really explain what they want to explain. Consider the central statement that "it is not the case that any imaginable semantic property or distinction can be grammaticalized. There is a fairly restrictive universal set" (Cinque and Rizzi, 2010). I am not disputing this fact. But is there anything within the cartography project that makes it true? To answer this question, ask yourself whether we would feel the need to change *anything* in the theory (assumed by cartography) if we were to discover a language that has affixes related to olfaction? The answer seems to me that if we were to discover such a previously unexpected element we would turn it into a head and try to locate it inside the hierarchy. This, to me, suggests that the predictive power of cartography studies is very weak indeed, perhaps due to the fact, recognized by Cinque in the interview that I quoted from above, that Cartography is not theory-driven

(contra Cinque and Rizzi (2010)). As Cinque (2002) points out, the question of which morpheme gets grammaticalized is addressed "in an empirical way." In the context of the minimalist program and the attempt to go "beyond explanatory adequacy," such an empirical approach ought to be viewed with suspicion. As Harbour (2006) points out, the grammaticalization of semantic properties and the combinatorial restrictions at the heart of cartographic studies "have to be explained, not reified."

2.4 More emergent asymmetries

So far I have been concerned in this chapter with showing that the necessarily asymmetric character of Spell-Out can, if taken to apply cyclically, as in phase-based derivational theories, complement the symmetric nature of Merge and generate a seed asymmetry that can be used to define post-syntactically, as opposed to pre-syntactically (lexically), certain basic distinctions, such as those underlying basic categories.

I am the first to confess that capturing the distinction between, say, nouns and non-nouns may not be that impressive, but let me stress again that the goal of a biologically oriented linguistic theory should not be to account for all the sundry differences among languages (what I have suggested we call 'elaborate grammatical structures'), but the very basic, fundamental properties that are arguably grounded in our biology (call these 'elementary syntactic structures'). As Borer (2003, 33) correctly points out

> It is in the nature of things that an endo-skeletal [lexicocentric] approach, with its ability to associate idiosyncratic as well as unpredictable syntactic properties with atomic listed lexical items, is both less restricted and more redundant, but also, potentially, more capable, at least prima facie, of describing the wealth of phenomena attested in natural language.

Seeking to (re)construct syntax from the ground up, with such a minimal set of (lexical, pre-syntactic) instructions does not mean reconstructing every detail ever identified in syntactic analyses, for not all such details form a natural (biological) class. The attempt to reduce the scope of lexical restrictions, and to broaden that of generic principles ("third factor principles" in the sense of Chomsky (2005)), thereby enhancing the explanatory quality of syntactic analyses, will come at a cost, of course: loss of language-specific (in the sense of specific languages), morphological detail. But if the goal is to understand the fabric of the language faculty, the fundamental processes behind the generalizations that make up Universal Grammar, this loss must not only be tolerated,

but should in fact be welcomed. The explanatory goal, from a biolinguistic perspective at least, is not the final state of knowledge of language known as I-language, but the forces that make the attainment of this final state possible. I-languages are a mixed bag, they are not only biological entities, but also the products of cultural processes. This is clearly expressed in the following quote, taken from Chomsky (1995, Introduction, p. 11, n. 6):

> Thus, what we call "English" or "French" or "Spanish" and so on, even under idealizations to idiolects in homogeneous speech communities, reflect the Norman conquest, proximity to Germanic areas, a Basque substratum, and other factors that cannot be regarded as properties of the language faculty.
> ... Pursuing the obvious reasoning, it is hard to imagine that the properties of the language faculty – a real object of the natural world – are instantiated in any observed system. Similar assumptions are taken for granted in the study of organisms generally.

Koster (2009) is right in saying that language, even in its I-sense, is not strictly a matter of biology but of applied biology, i.e., a technology belonging not primarily to individuals but to their shared culture. But Koster errs in thinking that this invalidates the goals of biolinguistics. It simply means that as far as the biolinguistic enterprise is concerned, a level of representation at which the morpho-syntactic properties of words don't matter must be found. This is what I am trying to achieve in this work. I am essentially asking readers to entertain the following thought experiment: imagine syntax without morpho-lexical baggage, syntax freed of lexical influence. What does that syntax look like? Is it good enough (or better) to capture (in a naturalistic, explanatory fashion) the basic properties of the syntactic component?

The preceding discussion has shown that a theory of categories can be obtained from minimal resources. In what follows I'd like to show that more properties can be deduced from the present model.

2.4.1 Adjunction and Pair-Merge

Up until now, the discussion has revolved around head–complement relations. Taking as our starting point (Norvin) Richards's local ban on structurally adjacent elements that are too similar to one another, we have been able to reproduce the structural pattern that, to my knowledge, (Marc) Richards was the first to uncover (see Richards (2011)):[55] syntactic derivations proceed

[55] Richards (2011), however, went on to segment derivations as Non-Phase–Phase–Non-Phase–Phase, which I take to be wrong if derivations proceed bottom up. A combination of Richards's segmentation and mine may help us understand where left-to-right parsing meets bottom-up derivations, an issue I will leave for further research.

along the line of an alternation between phase-edges and phase-complements. In the present system, a phase-head labels the head in its phase complement, generating the following rhythm:

(13) ...{Phase head, {Non-Phase head, Phase head, {Non-Phase head, Phase head, {Non-Phase head}}}}

While this pattern provides the basis for the main spine of syntactic trees (it is easy for the reader to see that the sequence {Phase head, {Non-Phase head, Phase head, {Non-Phase head}}} reproduces the more familiar Chomskyan sequence $\{C_{Phase-head}, \{T_{Non-Phase-head}, \upsilon_{Phase-head}, \{\sqrt{V_{Non-Phase-head}}\}\}\}$), it is silent on the issue of specifiers and adjuncts, and *a fortiori* on internally merged elements. These are the issues I'd like to address now.

Both specifiers and adjuncts are, using standard terminology, phrasal material that connect to the main branch of the syntactic tree. Both form {XP, XP} structures that are problematic for labeling algorithms like the one recently proposed by Chomsky. They are equally 'problematic' for the present model, as both specifiers and adjuncts are minimally of the form {Phase head, {Non-Phase head}} (adjuncts are, in fact, typically even bigger: {Phase head, {Non-Phase head, Phase head, {Non-Phase head}}}). Inserting one of them in the lowest phase complement in (13) is not problematic, it simply adds another (branching) layer of Phase-head–Non-Phase-head, as shown here:

(14) ...{Phase head, {Non-Phase head, Phase head, {Non-Phase head, Phase head, {{Phase head, {Non-Phase head}}, Non-Phase head}}}}

Although in this example the lowest (clausal) complement now contains two non-phase-heads, they are still separated by a phase-head, and as such do not run afoul of Richards's ban. The problem with specifiers and adjuncts arises when even a single one of these is inserted in a phase edge. There, they necessarily give rise to domains containing two elements of the same kind (phase-heads) that are adjacent within what will become the next higher phase complement. Such structures should lead to a crash once transferred to the external systems, as they yield *[XX]-sequences. And yet there is no questioning the existence of two broad categories, specifiers and adjuncts, in language.

The solution I'd like to put forth here will not only allow us to generate specifiers and adjuncts, it will also enable us to distinguish specifiers from adjuncts, for I think that there are in fact two ways to get around the *[XX]-problem we face, giving rise to different kinds of entities. The first solution to

the *[XX] problem once again resorts to the logic of Spell-Out and the asymmetry induced by (partial) transfer. Consider the following structure, consisting of a phase edge and a phase complement.

(15) {Phase head, {Non-Phase head}}

Adding another structure of this kind inside the phase edge {Phase head, {Phase head, {...Non-Phase head}}, {Non-Phase head}} – corresponding to, say, a *p*P adjoining to a υ – need not result in a violation of Richards's ban if we insert a phase boundary between the two phase-heads. This essentially amounts to spelling out the entire structure inserted into the phase edge as soon as it is merged to the main branch. The solution is indeed the familiar one: make sure that elements of the same kind end up in different Spell-Out domains. The solution is essentially the same as Chomsky's (2004) treatment of adjoined structures in terms of Pair Merge, and the operation "SIMPL" (short for 'Simplify'), which removes adjuncts from the structure they are adjoined to, putting them on a separate plane.[56] According to Chomsky, whereas standard merge ('set-merge') yields structures of the kind $\{\alpha, \beta\}$, adjunction ('pair-merge') yields structures of the kind $\langle\alpha, \beta\rangle$. Chomsky claims that the explicitly asymmetric representation of Pair Merge captures the inherently asymmetric nature of adjunction: an adjunct is adjoined *to* the main branch (not the other way around). In set-theoretic terms, $\langle\alpha, \beta\rangle$ is equivalent to the structures with explicit (asymmetric) labels used for Merge in Chomsky (1995): $\{\alpha, \{\alpha, \beta\}\}$, and there is a sense in which in the present model, the element which the adjoined material adjoins to acts as the 'label,' since the phase-head being adjoined to is the one that remains after Spell-Out takes place, just as it did when its original phase complement formed and got spelled out. That is, the present analysis reproduces the traditional treatment of adjunction as adjunction to a maximal projection.

Let me stress that the elementary treatment of adjunction offered here does not require any new operation like Pair Merge (as opposed to, say, Set Merge).

[56] See also Raposo (2002) for the claim that adjuncts must be spelled out as soon as they are inserted into the main derivation.

Note, incidentally, that if adjuncts are spelled out as soon as they attach to the main spine, talking about movement (internal) merge of adjuncts may not be accurate, since they are removed from the main structure as soon as they enter the derivation. Perhaps it would be more appropriate to speak of remerge (multiple external merge) in such cases (not to be confused with the option of 'Internal Pair Merge' discussed in Richards (2009a), which concerns instances of internal merge not associated with agreement). For theory-internal arguments against allowing movement of adjuncts, see Boeckx (2003).

In all cases, merge takes place and Spell-Out takes place in the usual way. Let me also say that the forced nature of Spell-Out of the entire adjunct probably underlies why adjuncts that are part of phase edges ('high adjuncts') are universally opaque for extraction (the adjunct island). Only adjuncts inserted very low (in the lowest phase complement) need not be spelled out on their own, hence are predicted to be more transparent. On the accuracy of this prediction, see Boeckx (2003, 2012b); Narita (2010a); Truswell (2007); Uriagereka (2012) (see Browning (1987); Uriagereka (1988)). Relevant contrasts are of the following type:

(16) a. *What did you get upset [because Mary said *t*]?
b. *Who have you been really happy [since talking to *t*]?
c. *What does John work [whistling *t*]?

(17) a. What are you working so hard [in order to achieve *t*]?
b. Who did John go home [after talking to *t*]?
c. What did John drive Mary crazy [whistling *t*]?

The solution just offered for adjuncts could, of course, be adopted for specifiers. Collapsing the two kinds of entities is not unheard of (see, most notably, Kayne (1994)), but the present framework offers an alternative treatment for (most) specifiers, not available for (most) adjuncts.[57] The alternative lies in the fact that most specifiers are nominal in nature, whereas most (perhaps all) adjuncts are adpositional/clausal in character.[58] If this difference can indeed be maintained, then, the present approach could distinguish between specifiers and adjuncts in the following structural terms:

(18) a. {Phase head, {Non-Phase head}} (specifier)
b. {Phase head, {Non-Phase head, Phase head, {Non-Phase head}}} (adjunct)

Though seemingly minimal, the structural difference at issue could make a big difference when we consider what happens when a specifier or an adjunct is inserted into the phase edge corresponding to the main spine of the tree. Both will give rise to structures where two phase-heads will find themselves structurally adjacent to one another, but whereas the adjoined structure is

[57] I have nothing to offer regarding multiple specifier structures, other than claiming that these must involve adjunction of some kind, followed by immediate Spell-Out, for them not to run afoul of Richards's *[XX] ban.

[58] On the existence of a phonetically null adpositional layer in nominal adjuncts, see Larson (1987); see also Boeckx (2003).

structurally identical to the structure it adjoins to (both are transitive phases), specifiers added to the transitive phase corresponding to the clausal spine are in terms of the present category system *intransitive* phases. As such, it is conceivable that the external systems manipulating syntactic representations could make use of this difference and exploit the different derivational histories and regard the phase-head of the specifier as sufficiently distinct from the clausal phase-head it merges with, allowing the derivation to proceed without immediate Spell-Out of the specifier phrase, unlike in the adjunct situation discussed above. This would allow us to capture the intuition that specifiers are more integrated into the main spine of the tree than adjuncts – at least, as far as nominal specifiers are concerned. Clausal/adpositional specifiers would have to be regarded in the present model as adjuncts, and are expected to be as opaque as traditional adjuncts. This seems to be the case, judging from the island character of clausal subjects (Ross's (1967) sentential subject constraint) as well as inherently case-marked arguments (i.e., nominals enveloped in an adpositional layer; for evidence that these are islands, see Starke (2001)):

(19) a. *Who is [for John to beat *t*] easy? (cf. [For John to beat Harry] is easy)
b. *Who did [that John beat *t*] please us? (cf. [That John beat Harry] pleased us)

(20) a. *(?)Ktereho doktora to byla chyba*
Which.Gen doctor.Gen it was fault.Nom
'Which doctor's fault was it'
b. ?*Ktereho herce by sis rad koupil obrazek*
Which.Gen actor.Gen would you gladly buy picture.Acc
'Which actor's picture would you gladly buy'
c. **Ktereho herce bys rad vynadal priteli*
Which.Gen actor.Gen would.you gladly scold friend.Dat
'Which actor's friend would you gladly scold'
d. **Ktereho herce se bojis pritele*
Which actor you fear friend.Gen
'Which actor's friend do you fear'
(Czech data; from Starke (2001, 38f.))

Readers may question the basis of the specifier/adjunct contrast proposed here, given that the contrast in (18) provides the most minimal of representations for each, but, of course, specifiers can be more complex than what amounts to a bare *n*P (using a traditional label). If more complex specifiers are as complex as the representation for adjunct in (18), then they too would have to be spelled

out as soon as they are introduced into the phase edge. This would seem to predict, incorrectly, that even minimally complex, *in-situ* specifiers would behave as adjuncts, for purposes of (sub-)extraction, for example.

The point is well taken, but I think that the basis of the distinction between specifiers and adjuncts proposed here can be maintained even if we expand the structure of specifiers in (18). To see how this is the case, I need to be more explicit about the structure of something like *a friend of Mary* – using labels for convenience: an [*n*P–[*p*P–*n*P]] structure. For us to be able to maintain the structural contrast in (18), such branching nominals cannot be of the form {Phase head, {Non-Phase head, Phase head, {Non-Phase, Phase head, {Non-Phase head}}}}, or, again, using traditional labels: {*n* head, {$\sqrt{N}$ head, *p* head, {$\sqrt{P}$ head, *n* head, {$\sqrt{N}$ head}}}} – this would be no different from the structure for adjuncts. Instead, I'd like to rely on the extensive literature (well summarized in Adger (2013b)) on nominals and their inability to take genuine complements and claim that a phrase like *of Mary* in *a friend of Mary* is actually part of the intransitive phase *edge* (and not its complement). That is, the relevant structure is as in:

(21) {{{Phase head, {Non-Phase, Phase head, {Non-Phase}}}} Phase head, {Non-Phase head}}
(linear order irrelevant)

Using traditional labels:

(22) {{{*p* head, {$\sqrt{P}$ head, *n* head, {$\sqrt{N}$ head}}}} *n* head, {$\sqrt{N}$ head}}
(linear order irrelevant)

What many would regard as the complement of *a friend* is then to be treated as an adjunct, but an adjunct that need not be spelled out as soon as it is inserted into the structure because it is a transitive phase buried inside an intransitive phase-head, for the same reason that an intransitive phase functioning as the specifier of a transitive phase need not be spelled out in the clausal domain. Given this representation, minimally complex clausal specifiers can be distinguished from minimally complex adjuncts, and need not undergo immediate Spell-Out. As a result, they need not lead to opacity in the case of subextraction.

To be sure, more complex nominals, such as definite DPs (*the (man's) book*), or strings like *a book of pictures of friends*, will have to undergo partial Spell-Out even prior to the clausal integration, because their expansion would contain at least one transitive phase inside another transitive phase (see our analysis of traditional DPs as a relative clause inside a double *n*P structure), which would

lead to a *[XX] violation. Partial Spell-Out being forced, these units are therefore expected to block subextraction from within the spelled-out domain. This is known to be the case for definite nominals (see already Chomsky (1973)), but it is also the case for non-definite nominals with expanded modifiers (see Bach and Horn (1976), and, for a recent reappraisal of the data, Gallego and Bosque (2011)).[59]

(23) a. Who did you see [a picture of *t*]?
(only one transitive *p*P phase inside an intransitive *n*P phase)
b. *Who did you see [the pictures of *t*]?
(two transitive phases: one corresponding to the *p*P, the other to the DP)
c. *Who did you see [John's picture of *t*]?
(three transitive phases:[60]one corresponding to the possessive phrase, certainly more complex than a mere *n*P; another corresponding to the *p*P, and a third, to the overall DP)

(24) *Who did you hear [stories about [a picture of *t*]]?
(at least two transitive phases, one for each *p*P layer)

To sum up, this section has shown that the difference between transitive and intransitive phases can not only be exploited to generate a category system, it can also be used to capture differences between units traditionally called adjuncts and specifiers. As a side effect, it leads to a rather natural account of certain opacity effects.

2.4.2 *The structure of chains*

Since the previous section touched on extraction, I'd like to devote some space to another asymmetry for which phases can serve as an achor. The asymmetry I have in mind is the asymmetric chain structure resulting from 'movement' (internal merge). As Chomsky (2004, 110) observed, barring an extra stipulation, "Merge is unconstrained, therefore either *external* or *internal*." Movement therefore comes for free, given the definition of Merge in (1), which does not say anything about the source of α and β. If their source is the pre-syntactic,

[59] Like Bach and Horn (1976), Gallego and Bosque (2011) put forth the idea that no subextraction can take place out of nominals, but I think the claim is too strong. The present model shows that subextraction is in principle possible from minimally modified nominals.

[60] This may account for why this example is judged worse than the (b) example; see Chomsky (1973).

narrow lexicon, we talk of external merge; if their source is the derivation, we talk of internal merge.

As Moro (2000) was the first to recognize, movement creates structures that can be used to break what would otherwise be fatal symmetries. In situations where not even traditional $\overline{X}$-distinctions suffice, as, e.g., in cases of merging two bare heads, Moro suggested we exploit movement to create an asymmetry. In traditional terms, the chain created by movement will turn the foot of the chain (the original 'copy') into a trace, an element sufficiently distinct from its merge partner to satisfy the *[XX]-constraint. In more minimalist terms (which avoid talk of traces), the original copy will be flagged as the member of a chain for purposes of the interfaces. Exactly what this flagging consists of is immaterial for Moro's purposes: *something* must instruct the external systems that the relevant element is to be interpreted as a variable, must be phonetically null, etc.[61]

Several authors have pointed out that structures created by movement ('chains') have a lot in common with traditional projections ('labeled phrases'): both are asymmetric; just like phrases decompose into maximal, intermediate, and minimal projection levels, so do chains (head, tail, and intermediate 'traces'); etc. (see Boeckx (2008b) and references therein). It is therefore tempting to use phase boundaries to recapture the asymmetry of chains in the same way we used them to recapture the asymmetry of labeling. Just like the element of the phase edge would act as the label for the element in the phase complement, the internally merged element in the phase edge could count as the 'label' (head of the chain) for the copy in the phase complement (tail of chain).[62] This immediately leads to the prediction that no proper chain can be formed in the absence of a phase boundary between the relevant copies. Put differently, the external systems interfacing with syntax won't be able to detect movement if it takes place internal to one of the two domains created by

[61] For an explicit mechanism (by no means the only possible one), see Nunes (2004).

[62] Chomsky (2012a, 3) states that "One goal of *Phase Theory* is to provide the mechanisms to distinguish copies from repetitions." But it seems to me that derivational history, rather than phases, does the trick. In the present model, phases only serve to distinguish between copies internal to a given chain (maximal vs. minimal). It should also be said that Chomsky's concern about copies vs. repetitions arises in a framework that is different from the present one in that it does not consider the scope of Richards's *[XX] ban.

Having said this, I think that if phases are to serve as useful instructions for the external systems, and avoid taxing the memory resources of these, it is safe to assume that for an element *x* to be considered part of a chain, a copy of it must be present in the immediately dominating phase edge. If it isn't, any higher occurrence of *x* is to count as unrelated (a repetition, not a copy). This is just another way of saying that movement must be 'successive cyclic.'

phases (the phase edge or the phase complement). The prediction appears to be correct, as it corresponds pretty closely to the so-called Anti-locality condition (Abels (2003); Boeckx (2008d); Bošković (1994); Grohmann (2003); Kayne (2005), and for the most comprehensive overview Grohmann (2011)), which bans movement that is "too local." Under most definitions, phrase internal movement counts as too local. Abels (2003) was the first to implicate phase theory in the context of Anti-locality. In his terms, phase complements (as entire units) are immobile because of the following conspiracy: Abels assumes, with Chomsky (2000a), that phase theory imposes a "Phase Impenetrability Condition" that renders the phase complement opaque upon completion of the phase. In order to avoid opacification, an element inside the phase complement must move to the phase edge. The problem for phase complements, the thing that renders them immobile, is that such a movement to the phase edge is too local (for Abels, it is phrase-internal, as it moves the complement of the phase-head to the specifier of the same head). Accordingly, only a portion of the phase complement can move to the phase edge. This is captured by the following representations.

(25) Given α a phase-head,

a. $^{*}[_{\alpha}\ t]$

b. $^{ok}[_{\alpha}\ [\ldots t \ldots]]$

In terms of the present framework, it is easy to see why entire phase complements are immobile. If phase boundaries are required to keep two merge partners distinct, bringing back these two merge partners into the same phase domain via internal merge would recreate the same configuration that required a phase boundary in the first place:

(26) a. $\{\alpha, \beta\}$

b. $\{\alpha, _{phaseboundary}\ \{\beta\}\}$

c. $!\{\beta\ \alpha, _{phaseboundary}\ \{\beta\}\}$

Another way to look at this anti-locality situation would be that given that labels are only useful if they are unique and exclusive, internally merging β in the configuration at issue would be like assigning two labels to the same object (the phase complement): the label provided by α and the label provided by the head of the chain. This is another way of saying that only phase-heads (labelers) are movable.

Readers familiar with the recent literature on phase theory are now likely to be thinking of the contrast between so-called A- and A-bar chains because a

common characterization of the difference in terms of phase-internal vs. phase-external movement appears to conflict with my contention that no movement can take place inside of either the phase edge or the phase complement (anti-locality condition). Several authors (including Norbert Hornstein and Shigeru Miyagawa, building on the parallel chain analysis in Chomsky (2008)) have suggested that A-chains land in a position below the relevant phase-head (e.g., A-movement from SpecυP to SpecIP, below phase-head C), whereas A-bar chains land in a position above the relevant phase-head (e.g., movement from SpecυP to SpecCP). Phrased in those terms, some instances of A-movement would be too local as they do not cross a phase boundary (e.g., movement from SpecυP to SpecIP). In the present model, all instances of movement must be 'phase-external.' This does not mean that we have no way of capturing the A- vs. A-bar distinction. As a matter of fact, the claim that A-movement is phase-internal whereas A-bar movement is phase-external is not tenable, given that elements in, say, SpecCP must eventually be transferred, hence must be part of some higher phase complement domain (i.e., phase-internal). Also, if all flavors of υ are taken to be phasal, passivization from within VP to SpecIP could be said to be phase-external, despite the fact that it uncontroversially counts as A-movement. The relevant movement would be phase-internal if C is taken as the reference point, but not if υ is. So, ultimately, the phase-internal/external characterization of the A/A-bar distinction is not as configurational/syntactic as it looks; rather, it counts as yet another lexicocentric description (as it crucially relies on specific heads, like C).

Still, the phase-edge internal movement corresponding to traditional instances of raising from SpecυP to SpecIP poses a problem for the present account, and deserves further consideration. We will see in the next section that the fact that this instance of movement enters into agreement ('feature valuation') goes a long way towards legitimizing it. This will also enable us to shed light on the so-called subject condition (the ban on subextraction from some – i.e., raised and agreeing – subjects). To get there, we need to be clear about what agreement, or more generally feature valuation, is in a feature-free syntax model like the present one. This is the topic of the next section.

2.4.3 Valuation

Let us start with how agreement (or more generally valuation) is handled in Chomsky's minimalist approach to grammar.

It appears to be a fact about natural languages that not all the features that appear on lexical items are semantically interpreted. Think of ϕ-features on finite verbs. These features are morphologically interpreted, but

semantically, they are inert. Chomsky (1995) calls these 'uninterpretable.' The very fact that they are realized morphologically suggests either that they are introduced post-syntactically, on the PF-side of the grammar, or else that they are present in the syntax (feeding morphology), but are marked in a special way in the course of the syntactic derivation, prior to interpretation, to make sure that the semantic component ignores them.[63] Chomsky assumes that the latter is the case. (Marantz (1991), Bobaljik (2008), Chung (2012), Landau (2013) and others have pursued the possibility that at least some uninterpretable features do not enter syntax at all.) Specifically, Chomsky assumes that uninterpretable features must be checked off in the course of the syntactic computation (via Agree, and sometimes also via Internal Merge). Chomsky (2000a) characterized (un)interpretability in terms of valuation, the reason being that barring undesirable look-ahead it is not clear why syntax should care about interpretability: why should it react (triggering movement, checking, etc.) to a notion that is only relevant post-syntactically? Chomsky proposed we understand (un)interpretability as the semantic counterpart of a lexical (pre-syntactic) property to which syntax could be sensitive: ± value (for a given feature F). According to Chomsky, some features on lexical items are lexically unvalued (they are stored lexically without a specific value), and unvalued features are intended to correspond one to one to uninterpretable features.[64] Syntax does not tolerate unvalued features because (by hypothesis) lack of value leads to illegitimate outputs at both SEM and PHON. (Think of a value as an instruction for the external system to perform a certain action. Lack of

[63] I say this because such features are clearly legible (i.e., interpretable in principle); see Brody (2003).

[64] Pesetsky and Torrego (2007) and Bošković (2011) question the link between uninterpretability and lack of value. They argue in favor of a double dissociation between (un)interpretability and (lack of) value so that some lexically valued features can lack an interpretation, and some lexically unvalued features can receive an interpretation. Clearly, this is the worst possible state of affairs in a minimalist context (especially one that argues against the proliferation of features), and so only the strongest kind of empirical argument should force us to go in that direction.

Two remarks are in order concerning proposals like Pesetsky and Torrego (2007) and Bošković (2011).

First, the claim that lexically valued features lack an interpretation is unspecific about which notion of the lexicon they mean. More precisely, if we adopt a distributed lexicon, as in Distributed Morphology, it is easy to obtain lexically valued features lacking an interpretation; all that has to be done is for the relevant features to be inserted post-syntactically, in the morphological component. These will not be interpreted, since they are not part of the computation feeding semantic interpretation.

Second, in an anti-lexicocentric model like the present one, where the pre-syntactic, narrow lexicon is so impoverished, lack of value can only mean lack of content, since there are no values assigned pre-syntactically.

value means lack of instruction, and assume that the external systems do not like points of indecision.) So, syntax must see to it that all features on lexical items come to be valued by Spell-Out. According to Chomsky, valuation arises as a result of special operation – AGREE – by which the feature values of a given lexical item (the valuer) are transferred to, or become shared with, the originally unvalued features of some other lexical item. The problem is that once valued internal to syntax all features will look alike at the point of Spell-Out: it will be impossible for the interpretive components to pick out those features that it should not interpret (the originally unvalued features). So, although we do not want syntax to have access to a post-syntactic notion like interpretability, it seems like we need the semantic component to have access to a pre-syntactic (lexical) notion like "lexically (un)valued." Chomsky (2004) proposed an ingenious way to have our cake and eat it too. Essentially,[65] Chomsky proposed that valuation (the point at which the relevant distinction for the semantic component disappears) takes place at the point of transfer, when the external systems interface with syntax. The idea is that by synchronizing valuation and transfer, the semantic component "sees" the valuation process (as opposed to just seeing its output), and is thereby able to tell which feature is getting valued (i.e., which feature was lexically unvalued). Once the semantic component sees this, Chomsky reasons, it is able to use this fact as an instruction not to interpret the relevant feature semantically.

Chomsky concluded from this that since uninterpretable/unvalued features are distributed on various elements across the syntactic computation (they are not all concentrated on the last stage of the derivation), the presence of these features forces cyclic Spell-Out. If Transfer only took place once (as in the earliest minimalist models like Chomsky (1995)), valuation and Transfer could not be synchronized, valuation requiring a very local domain.

Now let us look at the process of valuation more closely. Valuation is achieved by pairing the unvalued feature of the lexical item α with a matching (lexically valued) feature on another lexical item β that is connected to α. As already mentioned, the pairing/valuation process is called *Agree* in Chomsky (2000a). The bearer of the unvalued feature is called the *Probe*, and the bearer of the matching feature providing a value of? the feature on the other element is called the *Goal*. As Chomsky (2004) remarks, we do not want Agree to take place under Merge (i.e., we don't want the Probe and the Goal to merge directly) because Merge happens before Transfer does (not all instances of Merge lead to transfer), and we want to synchronize Agree (valuation) and

[65] For more extensive discussion, see Epstein and Seely (2002).

Transfer. To achieve this (i.e., to delay transfer, and thus dissociate it from Merge), Chomsky (2008) hypothesizes that the unvalued feature always originates on a phase-head (i.e., only phase-heads are probes), with the goal located in the complement domain of the phase-head. But now we face another problem. If the unvalued feature sits on the phase-head, it will not be transferred upon valuation, for only the phase complement (not the phase-head contained in the phase edge) is transferred. The problem is clear: we must find a way for the unvalued feature to be introduced after the goal is merged, but we must also make sure that it is transferred alongside the goal. The solution, proposed by Richards (2007) and adopted in Chomsky (2007), is to invoke a process of Feature Inheritance,[66] according to which in the course of Transfer (and valuation!), the unvalued feature on the phase-head (probe) is inherited by the non-phase-head sitting next to the goal in the complement domain of the phase (the domain that is being transferred).

Quite apart from the talk of features inherent to Chomsky's discussion, which, of course, begs the question of whether the Inheritance-based mechanism can be part of the present model, one immediate consequence of Richards's and Chomsky's reasoning about feature inheritance applied to the present model is that it restricts valuation to transitive phase domains.[67] With intransitive phases, only one element gets transferred, and this is not enough for proper valuation to take place, since the relevant configuration must contain (a copy) the valuer and the recipient of the unvalued features coming from the phase-head must end up on a Merge partner of the goal. To put it differently, only transitive phases are proper domains for phenomena like agreement. Arguably, this is why nouns (the outcome of intransitive phase Spell-Out in the present systems) are the only categories according to Baker (2003) that reliably support the interpretation of a referential index (which I take to be equivalent to interpreted ϕ-features). If one finds ϕ-features on nouns, these cannot have been the result of valuation; they must therefore be interpreted. (Note that I

66 Chomsky (2008) contemplates a generalized process of feature inheritance, but restricts it to unvalued features in Chomsky (2007).

67 In this sense, the present proposal converges with a key idea in Kayne (2011): nouns are not associated with unvalued features. (Recall that Kayne's proposal is similar in spirit to the present one, as it also takes Nouns to be singleton sets, though they differ in how these singletons emerge.)

Kayne actually pursues the stronger hypothesis that non-Nouns are necessarily associated with unvalued features, but I think this is too strong, as also I think is the idea that phase-heads necessarily bear unvalued features advocated by, e.g., Gallego (2011b). Gallego attributes the latter hypothesis to Chomsky (2008), but I think this is incorrect. Chomsky (2008) merely states that all probes must be phases, not that all phases must be probes (see also Chomsky (2012a)).

say Nouns, not nominal elements like Determiners, demonstratives and the like. These, as we saw above, are transitive phases, and thus support features that may not be interpreted semantically.)

The question looming large at this point is whether we can rescue the Inheritance account proposed by Richards and Chomsky in a model that has no room for "unvalued features" and processes like "Agree" and "valuation." (Notice, incidentally that the fact that I have found it impossible to avoid using terms like "unvalued features" and "Agree" when discussing agreement in a minimalist context is indicative of the centrality of featural transactions, and the lexicocentrism underlying it, in minimalist syntax.)

The first thing to note is that even if the actual valuation takes place outside of narrow syntax (as I will have to argue below, since there are no values in our syntax), the configuration leading to valuation must be something that the syntax must construct, in accordance with the idea that narrow syntax is the sole generative engine at the heart of this book (our heavy 'syntactocentric' stance). So, even if I slip into valuation talk in what follows, I ask the reader to bear in mind that I am really talking about configurations leading to post-syntactic valuation.[68]

The second thing to bear in mind is that the seemingly innocuous notation *u*F ('unvalued F') in minimalist syntax is an abbreviation for a feature lacking a value ([–F]), which is a property of a lexical item that (at the very minimum) bears at least one valued feature. So, when one talks about a lexical item X bearing *u*F, we should have the following (minimal) structure in mind:

(27) $X_{uF} = \{\{-F\}, \{+F\}\}$

But what is, really, [–F]? It's a feature lacking a value. Accordingly, I propose we represent this by means of an empty set: {{}F}. Lacking a value means (following the intuition in Chomsky (2004)) lacking (conceptual) content (hence its lack of interpretation at the semantic interface).

The intuition that I would like to work with is that the representation {{}F} is reminiscent of a functional item merging without a conceptual root in a Distributed Morphology framework (hence the lack of content). Now, in a model like the present one, the most plausible interpretation of this structure is for F (think of a functional element, such as Case, or ϕ) to correspond to a phase 'head,' taking as its complement a 'special' set: an empty set, special only

[68] In a certain sense, this proposal converges with the fractionation of Agree into two steps – a syntactic Agree-Link and a post-syntactic Agree-Copy – argued for in Arregi and Nevins (2012, 2013), Bhatt and Walkow (2013).

in the sense of being a set that fails to get associated with a conceptual root (hence lack of value leading to lack of content). Syntactically speaking, there is nothing exceptional: the empty set is a set like any other set. In fact as soon as we assume that Merge is set-formation, we have to assume that the empty set is something that Merge can operate with. Just as Chomsky (2004) said about internal merge, it would take a stipulation to bar the empty set from the pre-syntactic lexicon considered here. This effectively means that any model taking Merge to be set-formation anticipates the existence of elements without value/content. (Note, incidentally, that in order to entertain the option of merging with the empty set, it is necessary to view Merge as a free, un-triggered operation, as the empty set cannot satisfy any plausible vehicle requirement that one would impose on Merge in a crash-proof syntax model. In this sense, the present Merge α model provides a much better rationale for the existence of 'unvalued features.')

As pointed out above, unvalued features typically appear on lexical items bundled with at least one interpretable feature (perhaps as little as a category feature in some analyses). Indeed, Chomsky (1995) takes unvalued material to be added to a lexical item (a view already present in Chomsky (1965)). Chomsky takes this additional operation to take place upon selection, at the level of the numeration (i.e., in the lexicon). But consistent with our anti-lexicocetric stance, we could take the relevant bundle to be formed syntactically, via Merge. Specifically, we could take this addition of an unvalued feature as a case of syntactic adjunction, taking place not in the "lexicon," but in the course of the syntactic derivation. Probes, then, would receive the following representation, which matches Chomsky's claim that *u*F originate on phase-heads.

(28) {{Phase head, {}}, Phase head, {Non-Phase head . . . }}

Like any (high) adjunction configuration, this representation must lead to immediate spell-out of the adjunct. When discussing adjunction above, I mentioned that spelling out the adjunct can take the form of remerging it elsewhere (adjunct movement). One possibility not considered above but consistent with everything said here is to let the adjoined material {Phase head, {}} remerge with the non-phase-head inside the phase complement labeled by the phase-head to which *u*F originally was added. This is just another way of 'transferring' the adjunct, since the option under discussion amounts to making the adjunct part of the domain being transferred. Notice that this derivational step corresponds exactly to the process of Inheritance put forth in Chomsky. Contrary to what is sometimes being claimed in the literature, it is not necessary to view Feature Inheritance as a completely new process (nor is it correct

to view it as an instance of Feature sharing, contra Gallego (2011b)): feature inheritance is simply a way of discharging (transferring) material that overwhelms the workspace: it is simply a process of readjunction (reattachment), from the phase-head to an element that will be transferred.

Once added to the non-Phase head and transferred, *u*F will pick its 'value' from the closest c-commanding[69] intransitive phase (a copying operation that is the morphological equivalent of Chomsky's Agree),[70] which is the morphological component's way of handling empty sets (the semantic component simply appears to ignore it).

The addition of a phase inside the phase complement domain not only enables us to generate what one might call the Agreement/Valuation configuration, it also provides us with a way to characterize A-chains. Recall from the anti-locality discussion above that movement internal to a domain created by transfer, such as a phase edge or a phase complement, is problematic for the present framework. I hypothesized above that in order to be recognizable as such, chains must cross a phase boundary. The idea I'd like to put forth is that the addition of a phase-head inside the relevant phase complement makes A-chains definable. Given that the moving element provides the value for of *u*F, there is a sense in which, at the interfaces, it serves as the head of a chain from whose tail it is separated by a phase boundary.[71]

The upshot of the preceding discussion is that A-chains can only be defined in the context of a valuation configuration. This may explain why, for example, scholars have been led to implicate agreement in the characterization of the ban on subextraction out of A-chains (see Boeckx (2012b) for extensive

69 This c-command requirement is nothing other than the so-called "Extended Projection Principle." I am aware of instances of valuation in the absence of c-command (the so-called "long-distance agreement" phenomenon) (see Boeckx (2008a)). Pending a better understanding of what counts as the relevant morphological domain for the actual valuation process, I will leave a discussion of 'agreement-at-a-distance' for another occasion. Ideally, such domain should correspond to one defined by phases, but I know that arguments have been made in the literature for relaxing the locality domain defined by phases (the so-called Phase Impenetrability Condition) precisely to accommodate such instances of agreement. For relevant discussion of the latter point, see Boeckx (2009d), Richards (2012).

70 As pointed out by Roberts (2010a), depending on the features involved and the featural make-up of the Goal, the copying operation may even be the equivalent of a movement operation.

This statement is very similar to the morphological treatment of agreement offered in Bobaljik (2008).

71 Alternatively, we could pursue the suggestion in Epstein *et al.* (2012) and take Inheritance to amount to the formation of a doubly headed structure, with movement of the 'Goal' crossing a phase on a separate plane. In the present model, this would amount to saying that the adjunct structure corresponding to *uF* creates a structure on a parallel plane where the Goal also resides. The Goal would then be present on two structural planes separated by the phase boundary triggered by adjunction.

discussion of this issue). It may also explain why any attempt to derive the so-called Extended Projection Principle always seems to end up involving feature valuation of some sort (see Boeckx (2008a); Bošković (2002); Epstein and Seely (2006)). Finally, it may also explain why in the absence of agreement arguments for certain instances of A-movement (as opposed to, say, base-generation) are particularly hard to come up with (see, e.g., the literature on A-scrambling).

2.5 Additional remarks on the present model

Before concluding this chapter, I would like to make a few remarks concerning the present model. The first set of remarks pertains to my occasional use of terms like 'must' or phrases like 'this forces . . . ,' 'this requires . . . ,' 'this makes it necessary to . . . ,' in the preceding sections. What is the meaning of these expressions in a free-merge, anti-crash-proof model?

The second set of remarks will be about the differences between the present model and Chomsky's use of phases in recent publications.

2.5.1 Free operations in Merge α

A leitmotiv of the present work is that all syntactic operations are optional ("Merge α"). Syntax here is far less deterministic than in lexicocentric, crash-proof models. All syntax must do is provide options that the external systems interfacing with it can exploit, given their own resources. Syntax-internally, any instance of Merge or Spell-Out is essentially free (unconstrained). Nothing goes wrong, syntax-internally, if, say, a particular instance of merge fails to take place or if a particular point of transfer is delayed, or includes too much material. This perspective thus reduces the question of what it is about certain instances of displacement that makes these "obligatory" (e.g., satisfaction of the Extended Projection Principle) to a question about properties of the external systems. If these instances of movement don't take place in particular derivations, some post-syntactic condition will be violated. Likewise for transfer,[72] if too much (or too little) is transferred, or if some element is transferred

[72] Some of the empirical advantages of non-deterministic Transfer have already been provided in the literature. (By non-deterministic Transfer I mean that Transfer should be an operation as free as Merge is. That is, we should think of Transfer as Transfer α: bad transfer choices should be filtered out post-syntactically.) Although not all their details are reproducible in the present model, the spirit of these analyses is certainly the same.

For instance, Transfer α could lead to situations where what amounts in traditional terms to the phase-head is transferred alongside its complement, leaving but its specifier(s) in the (non-transferred) edge. This very possibility is suggested by Ott (2011) in his analysis of Free Relatives. Ott points out that transferring the phase-head would in many cases lead to

too late (or too early), the syntactic representation that serves as an instruction to the external systems to take a particular course of action will lead to a conflict somewhere down the road. There is no way to tell, syntax-internally. Accordingly, if I have used words like 'must' or 'necessary' above, I ask the reader to interpret these as unavoidable shortcuts for longer expressions such as "necessary to achieve a particular outcome post-syntactically," but without intending any look-ahead.

A consequence of this perspective is that constraints in movement (what Chomsky (1973) called "Conditions on transformations"), typically thought of as part of Narrow Syntax, cannot in fact be purely narrowly syntactic. Internal Merge ('movement') is just as free as any other type of Merge. To the extent that grammatical constraints on movement exist (and I take it that they do),[73] they must be construed as arising from problems at the external systems, or as emerging from the way syntactic information is passed on to these external systems, as I argued extensively, and independently in Boeckx (2012b).

That some classical cases of islands do not fall within the purview of narrow syntax is no longer an outlandish idea. For example, (some aspects of) the Coordinate Structure Constraint may reduce to a Parallelism condition – a post-syntactic, interpretive condition, maybe not even specific to the language faculty – on conjuncts, as argued by Fox (2000), Kato (2007), and others. Likewise, many 'Relativized Minimality'/intervention effects analyzed in the literature require a fairly large inventory of semantic or pragmatic features (witness Starke (2001)). It is quite possible, therefore, that at least some of these effects come about once the features to which Narrow Syntax is blind become available, i.e., post-syntactically (this is in fact one way of reading

a selectional failure (e.g., if C is transferred, it becomes impossible to state that a higher V takes the entire CP as complement, since C and V are not phase-mates), hence is 'prohibited' (again, I beg the reader to understand 'prohibited' as a descriptive term, not as a constraint on the operation Transfer: everything is possible; what is 'prohibited' is a post hoc evaluation based on what led to post-syntactic filtering). In some cases, though – for example, where selection is not an issue – transferring the phase-head as part of the phase complement is the desired result. Ott argues that this is exactly what is needed to understand the nature of free relatives.

Transfer α also forces us to leave the transfer-all option: we do not even have to stipulate that there must be a non-transferred portion (phase edge). If everything is transferred, the syntactic computation comes to an end. Arguably, this is a good solution to the persistent problem of what to do with the edge of the highest phase (root CP), in a framework like Chomsky (2000a): the absence of a phase edge serves as an indication that the sentence has been completed. This is exactly the conclusion reached by Obata (2010), who claims that root C is included in the last transferred portion. Arguably, this very indication could serve as an instruction to take the object thus transferred as the one to be judged as True or False (if the speaker intends to do so), if Hinzen (2007) is right about Root ForceP/CP being the point of judgment.

[73] See Boeckx (2012b, chap. 2) for arguments against so-called reductionist accounts of islands.

Szabolcsi (2006)). In a sense, the position advocated for here is a return to the (minimalist) logic of Miller and Chomsky (1963) (see also Bever (1970)): keep the theory of competence maximally simple, and find the source of unacceptability in the ways in which linguistic information is digested by the performance systems.

Quite independently of the set of assumptions made here, the recent literature on resumption and ellipsis (see Boeckx (2003), Merchant (2001) and references therein) suggests that it is simply false to hold onto the traditional slogan that movement cannot take place across islands. There is massive empirical evidence reviewed in Boeckx (2012b) that indicates that movement has taken place across nodes that many would regard as defining island domains, with acceptability determined only once a variety of other, post-syntactic factors has been taken into account.[74] In a phase-based system, it is actually quite hard (short of ad hoc moves)[75] to prevent movement from crossing phases. Quite apart from the fact that constraining internal merge would depart from Merge α, some 'communication' must be allowed to take place across phases. This is the role of phase edges. But as soon as some movement is allowed to target phase edges (as it must, to capture the phenomenon of long-distance movement dependencies), this escape hatch can be used to bypass any blocking effect one may be tempted to encode in phase-heads.

As I argued in Boeckx (2012b), to the extent one wants to involve phases in some aspects of locality (and I agree with everyone else that the kind of elements assumed to be phasal matches the traditional island typology so closely that it can't be an accident), the only natural option is to implicate phases indirectly. Dependencies across phases (via Internal Merge) can be formed, but perhaps something about some of these dependencies, *when spelled out cyclically*, causes the external systems to 'gag' (and in many cases, with no option for 'recovery'). I like to think of this as the logic of the garden path. As is well known, the status of sentences like *the horse raced past the barn fell* is degraded. The standard explanation for this fact is that the reduction of the relative clause ('raced past the barn') causes the language processor to parse

[74] Notice that it would be wrong to conclude from this literature that island effects have nothing whatsoever to do with syntax. The most one can conclude from this is that 'bounding' nodes do not block movement, but it may well be the case that (for some reason to be made precise) movement across bounding nodes leads to unacceptability unless something else (deletion, resumption, etc.) happens (see Lasnik's (2001, 2005) notion of 'repair by ellipsis').

[75] The clearest (but by no means the only) example of this is to be found in Müller (2010), where a very explicit way to capture CED-effects is proposed, which requires no less than all the following conditions: that (i) all instances of merge be feature-driven; (ii) all phrases be phases; (iii) features on lexical items be ordered (hierarchically organized); (iv) edge features be added in the course of the derivation.

'the horse raced past the barn' as a main clause, leaving 'fell' as a puzzling afterthought. Had the relative clause been spelled out in full ('the horse that was raced past the barn'), the full sentence would have been unproblematic. The hypothesis concerning (some) islands that I would like to entertain here is that the cyclic character of Spell-Out imposes a certain mode of processing syntactic chains (movement dependencies) that in some cases (i.e., when island effects obtain) leads to the perception, on the part of the external system, that the dependency has been completed, when in fact it hasn't. Put differently, cyclic Spell-Out can be misleading, and can lead the external system down the garden path, into an island.[76]

The minimalist literature on islands is vast, and I will not attempt to review it here (I have done so elsewhere, see Boeckx (2012b)). I will instead focus on what is I think a growing consensus in that literature: that traditional CED-effects, previously unified under the umbrella of proper government (Huang (1982)), should not in fact be brought together: subjects (or displaced arguments more generally) and adjuncts pattern differently. The robust impression[77] gathered from decades of cross-linguistic and cross-constructional investigation is that extraction from adjoined domains is consistently bad, except with low adjuncts (see the discussion above), whereas extraction from displaced arguments is subject to variation, and in some cases perfectly acceptable (an impression reinforced by more recent studies indicating different psycholinguistic profiles for the 'subject island' and 'adjunct island'; see Sprouse (2007) for valuable discussion).[78]

The general strategy for dealing with the adjunct island that the present model offers is to say that subextraction is blocked because the entire adjunct domain is spelled out and transferred as soon as it is introduced into the

[76] It is well known that repeated exposure to garden path sentences improves their status. Perhaps this is how one should understand the island-alleviating effects reported in the literature on islands (Hiramatsu (2000); Snyder (2000)); although see Sprouse (2007) for a dissenting view on satiation.

[77] See Stepanov (2001, 2007) for clear statements. See also Richards (2001, 187).

[78] To be fair, though, comparing the robustness of the subject island and of the adjunct island must, if it is to be compelling, ensure that the meaning of the term 'subject' in defining the 'subject island' is kept constant. After all, 'subject' is not a primitive notion in generative grammar: there are low (*in situ*) subjects, high subjects, agreeing subjects, non-agreeing subjects, and more. Perhaps part of the cross-linguistic variability of the subject island is the direct result of the many meanings of 'subjects.' It is also worth noting that the psycholinguistic evidence against the robustness of the subject island is still a matter of debate (see Jurka (2010); Phillips (2006)). Uriagereka (2012) makes use of such considerations to advocate for a uniform treatment of CED-effects, of the sort he originally put forth in Uriagereka (1999). For a related (uniform) treatment of CED-effects, see Narita (2010a).

derivation. It leaves no room (no edge) for extraction. As for subject islands, numerous authors (Boeckx (2003, 2008b, 2012b); Richards (2001); Rizzi and Shlonsky (2007)) have independently come to the conclusion that what freezes the content of subjects (or more generally, displaced arguments) is the valuation relation that I have argued here makes A-chains definable in the first place. Such valuation relation in effect defines the domain out of which one is trying to extract as a 'maximal' unit (head of a chain). And just like one cannot overwrite labels, one cannot overwrite chain-labels: if the elements that constitute the subject count as a maximal domain, no subpart of it can count as a maximal domain (say, head of an A-bar chain) on a subsequent cycle (this sort of chain interleaving would amount to a cyclicity/monotonicity violation). Hence the subject island.

In sum, unextractable domains[79] amount to pair-merged units and displaced valuers, which are islands because information delivered by syntax takes place in a cyclic fashion. That is to say, the very thing that makes it possible in the present system to define units like 'adjunct' or 'A-chain' in the first place (the cycle) is the thing that eventually makes these domains opaque.

2.5.2 *Syntactocentric vs. lexicocentric uses of phases*

Many different ways of understanding phases exist in the literature, as is to be expected from the open nature of research programs like minimalism. I will not attempt to compare my use of phases with all the other conceptions of phases that I am familiar with (readers can get a taste of the diversity of conceptions of phases from Gallego (2012), but I feel it is important to stress certain differences between the present model and the way in which phases have been put to work in Chomsky's writings, given the importance of the latter in the field.

A quick survey of the literature reveals that phases – much like bounding nodes in the 1970s – have mainly been used to capture two 'big' facts about human language syntax: successive cyclic movement and island/subjacency effects.[80] These are in fact the joint rationale for phases offered in textbooks on minimalist syntax (see, e.g., Adger (2003, chap. 10)).[81]

[79] The only way to circumvent these islands is to resort to resumption, which as I argued in Boeckx (2008b, 2012b) underlies all instances of 'island repair,' and amounts to duplicating the entity to be extracted, placing it in a separate workspace, outside the islands, and letting it 'move' from there.

[80] See, however, Samuels (2010) for other, more interface-based uses of phases, linking it to the notion of the "phonological cycle," the original notion of the cycle in generative grammar.

[81] Hornstein *et al.* (2006, chap. 10) also discuss Chomsky's original (2000a) argument for phases partitioning the numeration into subnumeration. But given that the concept of (sub)numeration

Phases, according to Chomsky (2000a), impose a "Phase Impenetrability Condition" on syntactic derivations, according to which at a given point (around which there is some debate; contrast Chomsky (2000a) with Chomsky (2001)) elements inside the complement of the phase-head become inaccessible for further computation. To remain active, elements from inside the phase-head have the option to move to the edge of the phase. At this point we are in familiar territory: the Phase Impenetrability Condition ensures that long-distance dependencies will have to be formed via successive cyclic movement ("Comp-to-Comp" or "phase-edge-to-phase-edge"). At the same time, the Phase Impenetrability Condition offers the possibility of viewing the trapping effects of islands as instances where the escape hatch, the edge of the phase, becomes – for some reason or other – inaccessible to a given element.

The trouble is – and, in fact, has always been – that it is far from easy to come up with a good reason why the edge of the phase becomes inaccessible. One can certainly code it featurally (witness Müller (2010)), but it should be obvious that imposing conditions on edge accessibility is simply a way of getting the facts,[82] not a way of going beyond explanatory adequacy. The move can hardly be characterized as 'minimalist.' The fact that this way of getting the data has been used for close to forty years indicates both how theoretically conservative linguists have been, and perhaps also how hard it is to come up with some alternative.

As for the idea that phases capture successive cyclic movement, let me briefly touch on an issue that was first brought up in Abels (2003), and has since been taken up in Boeckx (2008d) and Abels and Bentzen (2009). Abels contrasts two ways of conceiving of successive cyclic movement: a classic way, according to which an element moves only through some well-designated intermediate landing sites (forming "punctuated paths"), and an alternative way that takes successive cyclic movement to move through all the projections separating its projection of origin (forming "(quasi) uniform paths").[83] To the extent that not all phrases are phases, phases appear to favor the view that paths are punctuated, which Abels and Bentzen (2009) argue is descriptively more adequate. Unfortunately, even if the evidence is taken at face value, it does not tell us if movement proceeded in a punctuated, or (quasi-)uniform manner, since the evidence is only indirect, dealing as it does with interface effects

has fallen into disrepute since then, it is fair to say that this is no longer a major function of phases in current syntactic theorizing.

82 Here I should perhaps say '*at best* a way of getting the facts,' for I do not think that it even gets the facts, when the data base is expanded. For discussion, see Boeckx (2003, 2008b, 2012b).

83 Abels notes that the paths are not entirely uniform as movement does not target intermediate projections of the phrases it moves through, for reasons of chain uniformity.

such as reconstruction effects. Such evidence merely points to the fact that syntactic derivations interface with the external systems in a punctuated fashion. It does not, indeed cannot, indicate the path taken by a moving element.[84] The evidence in favor of the punctuated path hypothesis presented in Abels (2003), Abels and Bentzen (2009), Lahne (2008), based on the lack of reconstruction effects in some position, would be conclusive only if we assumed that every copy left by movement can be reconstructed, but we know independently that this assumption is untenable (see the discussion in Fox (2002), Takahashi (2010), Takahashi and Hulsey (2009) on the need for copies to be modified to feed reconstruction, and why some copies of some categories can't be so modified). In other words, such evidence is evidence for cyclic transfer (and concurrent interpretation), not for how chains are formed. Put another way, it is evidence that Spell-Out is cyclic/punctuated, not that paths are.[85] Accordingly, it is false to claim (as the textbooks do) that phases enforce successive cyclic movement. Phases, understood as cyclic (punctuated) Spell-Out, may provide a good way to capture the interface reflexes of successive cyclic movement, but they say nothing about the process of chain formation (or recursion, for that matter, contrary to Arsenijević and Hinzen (2012)).

The previous discussion makes it clear that the problem with relating phases and successive cyclic movement or islandhood boils down to lexicocentrism. The reason that phases are often said to motivate successive cyclic movement is because it has been assumed since Chomsky (1993) that movement only happens for a reason: be it a morpho-syntactic reason ("to check a feature"), or an interpretive reason ("to have an effect on outcome"; see Fox (2000), Reinhart (2006)), or to avoid the trapping effects of the Phase Impenetrability Condition. But if movement is conceived as internal merge (as proposed in Chomsky (2004)), and internal merge is really just merge ('internal' being devoid of theoretical import; simply a descriptive term), and merge is free, then the explanatory link between phases and, say, islands breaks down.[86]

[84] As already pointed out in Boeckx (2008d), Boeckx and Hornstein (2008).

[85] I thus still endorse my earlier conclusion (Boeckx, 2003, 2008d) that there is no reason to reject the idea that paths are quasi-uniform. In a framework such as the one pursued below, there cannot be any constraint on the way paths are formed.

[86] Although the term 'lexicocentrism' was not used in Boeckx and Grohmann (2007), Grohmann and I made it clear right from the beginning of the criticism of phases that "we could not agree more with the general vision and virtually all the arguments made by Chomsky over the years regarding the motivations behind the Minimalist Program" (p. 204). We even accepted the general intuition expressed by Chomsky over the years that phases play a role in reducing computational load.

Our concern stemmed from how this intuition was being cashed out in practice. To us, it seemed – and it still seems to me – that "virtually all the properties ascribed to phases in the

A major way in which the present system differs from Chomsky's phase-based approach is that it not only dispenses with predefined phasal categories (phase-heads) (thus avoiding altogether the so far intractable problem[87] of why C, not T, is a phase-head), but it also allows for more than two phase boundaries within a clause. Chomsky restricts himself to just two phasal nodes at the clausal level: C and υ (which begs the question of how question words in SpecCP$_{root}$ are ever interpreted since they are never transferred). The present system says that the more elements a clause contains, the more asymmetries will have to be anchored (the more conceptual addresses/roots will have to be categories), hence the more phase boundaries will be needed. The present take on phase is therefore more dynamic than one rigidly specified in the lexicon.

(At least) two conceptual objections could be (and have been)[88] formulated against the present conception of phase: (i) it could be said that adding phase boundaries in a derivation renders the computation more complex; and (ii) having more phases[89] than C and υ forces one to abandon the link Chomsky thought to establish between his two clausal phase heads and the "duality of semantics" (an idea going back to the distinction between Deep Structure semantics, which deals primarily with thematic properties, and Surface Structure semantics, which deals with discourse properties; cf. Jackendoff (1972)).

current literature have been recycled from the very first theoretical attempt to make sense of such phenomena as islands or successive cyclicity (Chomsky (1973))" (p. 205). We immediately pointed out that "[i]n and of itself, the fact that phases have theoretical antecedents is not a bad thing." The problem was at the theoretical level – at the level where attempts to go beyond explanatory adequacy are evaluated. Grohmann and I noted – what is obvious to everybody – that "phases are to minimalism what bounding nodes and barriers were to the Extended Standard Theory and Government-and-Binding Theory, respectively" (p. 205). Our overall assessment was that "Like bounding nodes and barriers, phases beg questions that lead to persistent problems. Accordingly, phases do not enhance our understanding of syntactic phenomena like locality; they simply recode insights from the past" (p. 205).

87 The difficulty in justifying the phasal status of certain heads is actually a reflex of the more general difficulty of providing a solid theory of categories in a lexicocentric model (see Baker (2003) and the concluding section of this chapter on this point).

88 Ángel Gallego (p.c.), referring to oral presentations of this material.

89 Let me stress that although a typical syntactic derivation is likely to have more phases in it under the present approach than under Chomsky's, it is important to bear in mind that we are talking here about phase tokens. When it comes to phase types, the present system only recognizes two phases: transitive and intransitive. In the absence of a theory of (functional) categories in Chomsky's system, it is unclear how many types of phases there are: for me, C and υ are of the same type (in line with what I just said, C is just a higher-order υ/p). I do not know how Chomsky would relate C, υ, and the other phase-heads he occasionally recognizes (see Chomsky (2007)), such as P, D, and n.

I do not think that these objections are strong enough to dismiss the present approach to phases. If anything, they are reasons to adopt it. For one thing, it is true that the present proposal says that as the size of the derivation grows, more phase boundaries will have to be posited. But for the syntactic component to self-organize in this way (one could speak of 'online scaling') is precisely what we expect if good design plays any role in constraining the language faculty. In fact, the increase of phase boundaries proportional to the size of the derivation appears to be a specific instance of what Chomsky would likely call a "third factor" principle; in this case, a specific instance of the Menzerath–Altmann law. The Menzerath–Altmann law is a general law, first formulated in the context of human language, but later on extended to non-linguistic domains, such as the organization of the genome (see Ferrer i Cancho and Forns (2009)), that says that the increase of a linguistic construct results in a decrease of its constituents, and vice versa. So for instance, the longer a word, the shorter its syllables (i.e., the larger the number of syllables). The Menzerath–Altmann law strikes me as very intuitive: the longer the sequence to memorize, the more one chunks it. It strikes me as wrong to believe that an increase in phase boundaries leads to greater computational complexity. Just the opposite, in fact. Phases reduce computational complexity (by reducing the size of the active workspace). We therefore expect more phases (to reduce the burden on computational memory) if the size of the derivation grows.

As for the second criticism, related to the duality of semantics, it too fails to cast doubt on the present proposal. It is true that CP and υP neatly map onto the domains of discourse structure and argument structure, respectively, but they do so only in the most basic of circumstances. As soon as one looks into the issue in more detail, the neat cut evaporates: for example, we find the same duality of semantics expressed in nominal phrases, and one can posit a similar C-υ articulation in nominals (indeed, many have done so), but note that the C-layer of nominals often enters into the υ-layer of the clause, disrupting the neat mapping. Similarly, clausal CPs function as arguments, again causing the two semantic layers to intersect. Moreover, it is now a well-established finding of the Cartography literature that the edge of the υ layer hosts elements that contribute to discourse articulations (topics, foci, etc.) (Belletti (2004)), which causes the two kinds of semantic relations to intersperse ($\overline{A}$–A–$\overline{A}$–A). Finally, it is becoming increasingly tempting to analyze Surface semantics as just another instance (a higher-order expression) of Deep semantics, that is to say, to view discourse structure as a special instantiation of argument/event structure, namely, an articulation of the speech event, with discourse functions corresponding to higher-order

theta-roles (discourse/speech-event participants). Evidence in favor of this view comes from analyses of Quantification (as reprojection/second-order monadicity) (Hornstein and Uriagereka (2002), Pietroski (2003a)), Focus articulation (Herburger (2000), Irurtzun (2007)), evidential markers (Speas and Tenny (2003)), and the nature of person-marking (Sigurdsson (2004b)). All these proposals cast doubt on the desirability of separating the C domain from the υ domain. In addition, the reason why phase-heads like C and υ are said to provide the right cut for semantics in Chomsky's system is because of the very limited functional inventory assumed. As soon as other nodes are considered,[90] shorn of lexicocentric properties, phases in the present model provide a seed asymmetry that all approaches to phase recognize (the transfer asymmetry). In terms of 'substance,' such an asymmetry is very much underspecified. It only served as a precursor for grammatically more contentful differences such as categories, chain members, extraction domains, and so on. While all of these preserve a syntactic aspect in the present model, none of them are purely syntactic. They arise as interaction effects involving syntax and external systems.

2.6 Conclusion: syntax as grammatical pre-pattern

Let me summarize briefly what I have tried to achieve in this chapter. Taking as my starting point the explosion of unexamined feature structures pre-syntactically, I have argued for a shift of perspective from lexicocentrism to a more radical version of syntactocentrism, which I view as a necessary first step in the direction of integration of theoretical linguistics within the interdisciplinary field of biolinguistics. The core idea of this book is to get grammatical order (in the sense of organization, hierarchy, etc.) not from the lexicon, but from the syntax (derivational dynamics).

One way of looking at the present proposal is to view it as a drastic reduction of the number of (morpho-)syntactic features, arguably the most drastic reduction there can be: for syntax to exist, only one syntactically relevant property is needed: the edge feature, which all syntactic units must have in

[90] In recent writings, Chomsky has gone back and forth regarding the phasal status of categories like P and D. He also has not been very specific about Distributed Morphology-style $\sqrt{Root}$-categorizing heads such as *n* or *a* (on a par with υ).

Needless to say, the exponential growth of the functional lexicon in the context of cartographic approaches also begs the question of whether other heads also qualify as phase-heads, a question rarely addressed explicitly in the relevant literature (but see van Craenenbroeck and van Koppen (2002)).

order to combine. (The edge feature is the least common denominator that can be found among lexical items.) I have proposed that in some sense this is all there is (specific) to human language syntax. All you need is the unrestricted merge property, which is of course not to say that all you have in syntax is the merge property. Once you have the merge property, several properties follow at once, for free, as 'third factor' effects (in the sense of Chomsky (2005)). First, the lack of selectional restrictions allows for Merge to be unrestricted (set-formation turns into full-blown merge, as it were: internal and external). Second, precisely because Merge is unconstrained, its power is to be curbed if it is to meet the most elementary conditions imposed by the external systems (such as the *[XX] ban).[91] This requires Merge to be cyclic, so as to allow for the asymmetry inherent in partial transfer to be exploited.[92]

[91] Although I have not insisted on this point in the text above, I take Richards's *[XX] ban to be a specific manifestation of a more general information-theoretic constraint. Consider what the syntactic component can produce on the basis of Merge alone: giant sets like {A, {B, {C, {D, {E, {F, . . . }}}}}}. Such sets don't look so inadequate until we remind ourselves that the As, and Bs, of the set just written down are indistinguishable from one another, that is: {A, {A, {A, {A, {A, {A, . . . }}}}}}. In turn, this giant set equals {A}, so all this Merge activity achieves very little. It is clear what the problem is. I assume that the external systems interacting with syntax have resources of their own, among which are distinctions ('features'), be they PF-distinctions or LF-distinctions. But such resources cannot be put to use (in a systematic fashion) in the absence of anchor points in the syntax or instructions as to what to distinguish where and when. If syntax produces structures like {A, {A, {A, {A, {A, {A, . . . }}}}}}, interpretively, it boils down to {A}. Think of this situation in information-theoretic terms: since Shannon's famous characterization of information, we know that if I say A, then I say A again, then I say A another time, very quickly (infant studies run by psychologists indicate that this is extremely quick), my utterances will become uninformative. They will have become too predictable. In terms of informational addition, their value is nil. To avoid this habituation effect, psychologists know that they have to introduce unexpected elements, which is to say, they have to introduce differences, or choice points to avoid the degradation of the signal (the irremediable effect of entropy). The same strategy must be used if we want to avoid the 'boring' effect of pure Merge (*[XXXXXXXX]). Phases are a good way to achieve this.

[92] The cyclic character of merge may be one way of making sense of why Chomsky called the merge property the *edge* feature. I suspect Chomsky called this property the edge feature because "merge is always at the edge" (Chomsky (2005, 13), Chomsky (2007, 11)), in accordance with the Extension condition (Chomsky (1993))/No Tampering Condition (Chomsky (2000a)), which forces Merge (be it external or internal) to always target the root. But the term edge is also used in the context of phases, to mark the growth zone of the derivation. In phase-based derivations, the edge is dynamically defined, traveling from unit to unit, as the derivation grows. Because the edge of the phase defines the portion of the derivation in which lexical items are accessible (the 'active workspace'), we could think of the edge property of lexical items as an oscillating (i.e., cyclically expressed) property – activated only while the element is in the phase edge.

Put differently, the present book argues against taking recursion to be the key factor of the faculty of language.[93] Rather, the crucial factor was the emergence of a universal, uniform, homogeneous pre-syntactic lexicon, which gives us the capacity to abstract away from selectional restrictions and the varying valency of concepts, and arrive at a Merge α system, which pushed the cognitive capacity of our species known as syntax into the realm of Turing patterns, dissipative structures, and self-organizing systems (what I have called Santa-Fe Style Syntax, alluding to the center where self-organization was taken to be the focus of inquiry across fields).

Indeed, the combination of unconstrained Merge and cyclic Spell-Out gives rise to a model that belongs to a family of systems that self-organize – avoiding the form-deteriorating effect of entropy by staying at the edge of chaos. To this family belong: the reaction–diffusion mechanism that Turing (1952) envisioned, Prigogine's dissipative structures, the Local Autoactivation and Lateral Inhibition (LALI) models of Gierer and Meinhardt (1972) and the Clock-Wavefront model of Cooke and Zeeman (1976), all of which now form part of the toolkit of theoretical biologists attempting to account for the generation of biological form and its robustness. The syntax that emerges from the present work has the characteristics of Bak's notion of 'self-organized criticality' (see Bak *et al.* (1988)): a mechanism that creates elements that will eventually lead to its destruction. Think of the sand pile: grains of sand accumulate to form a pile, and in so doing produce a structure that leads to its destruction (avalanche). If the system is connected to a constant source of energy/matter (constant sand addition), the accumulation will begin again after the avalanche, leading to another avalanche, and so on.

The Conjunction of Merge (local auto-activation) and Cyclic Transfer (lateral inhibition), connected to a constant source of "matter/energy" (the pre-syntactic lexicon), has the same characteristic as Bak's sand pile model: the more you merge, the faster you reach a phasal point, a point of transfer that destroys (a large part of) the structure achieved so far, only to serve as the

[93] In the present model, the outline of the recursive structures found in natural languages emerges from this combination of Merge and Cyclic Transfer. It is this very combination of Merge and Cyclic SpellOut, the alternation between Phase and NonPhase that gives rise to the patterns that Tom Roeper has called 'Indirect Recursion' (the impossibility of embedding a given category immediately inside another unit of the very same category). But it would be wrong to conclude, as Arsenijević and Hinzen (2010) do, from these surface patterns of category distribution that there is nothing like Merge. Merge is truly densely recursive, but the external systems processing the output of narrow syntax impose conditions that obscure this fact (yielding sparsely recursive structures), much like the system processing languages obscures the legitimacy of center-embedding in syntax, as Miller and Chomsky (1963) concluded over fifty years ago.

basis for the formation of a new structure, to be transferred/destroyed again, etc. The general properties of such a system can be related to processes in chemistry, biology, social structures, and so on (see Ball (2006)) because these systems have been viewed from a perspective that seeks to identify its universal, generic characteristics. It is precisely because all these processes are insensitive to the fine details of what they manipulate that they can be compared and related to one another. And it is precisely because I have reduced syntax's sensitivity to lexical information so much that the generic properties of the syntactic component can be made visible.

The conjunction of Merge and cyclic transfer vindicates the assertion I made in Boeckx (2009e) that Merge in in some sense 'not enough.' In Boeckx (2009e) I claimed that Merge had to be decomposed into Combine and Label (analyzed there as copying), that is as a combination of a symmetric and an asymmetric process. (For a related proposal, using a different set of primitives, see Hornstein (2009).) This idea is maintained in the present framework, but a more natural characterization of the asymmetric step is offered here in terms of cyclic Spell-Out. In so doing, we no longer need to say (as I did in Boeckx (2009e)) that Merge has to be decomposed. Rather, as Noam Chomsky noted in a remark reported in Boeckx (2009e, 52f.), a labeling algorithm has to be added on top of Merge. What I proposed here is a natural labeling algorithm (without actual syntactic labeling). The present approach thus agrees with Chomsky that the notion of labeling is indeed "a dispensable notion," "a convenient notational device," "playing only an expository role" (Chomsky (2007, 8, 23)). If I am correct, phasal transfer achieves dynamically much of what labels did representationally, or what features did lexically (since lexical features are micro-representations that project at the maximal node level as labels, labels and features can be used interchangeably. There is no gain using one or the other, contra Collins (2002)). The key difference is that the 'labels' of the present system do not encode the interpretive differences, that is, they do not redundantly express differences that are necessarily present post-syntactically. Instead, they construct points at which differences that independently exist get expressed linguistically. They render a Merge α model informative (that is, meaningful/fully legible) from the perspective of the (mind-internal) external systems.

As we saw, this way of looking at things has as a consequence that many properties once thought to be part of narrow syntax emerge dynamically, as interface (epi)phenomena. Take the case of labels. Instead of letting headedness inhere in the definition of Merge (as in Chomsky (1995, 243, 244, 246), or via feature-proxies (Collins (2002)), I claim that asymmetric labeling

should be defined outside of the product of Merge, in the context of phases, pretty much like lexical categories are defined contextually in Distributed and Exoskeletal Morphologies (Borer (2005); Marantz (2008)). The basic idea behind endocentricity is captured, but the label of a phrase (at least that of the phrase corresponding to the phase-complement) is defined from with-out, no longer from with-in – syntactically, not lexically.

The present model is obviously indebted to Borer's exoskeletal stance, and to the notion of a distributed lexicon in Halle and Marantz (1993). In a certain sense, it pushes these views to their limit. It also builds on Moro's (2000) insight that some asymmetries can arise syntactically, as broken symmetries. Finally, the present approach also benefited greatly from the configurational approach advocated by Hale and Keyser (1993, 2002) in the context of theta-roles. Indeed, the logic behind the category system proposed in this chapter is essentially an extension of Hale and Keyser's theory of argument structure.

Categories offer perhaps the clearest contrast between lexicocentrism and syntactocentrism. As Baker (2003) has made abundantly clear, standard treatments of categories in generative grammar leave much to be desired at the explanatory level.

In his book Baker focuses on lexical categories, but the situation is even worse in the functional domain, where positing new heads is as easy as positing new features. As is well known, the strength of theories is measured in large part by the sort of things they deem impossible. Weak theories don't exclude very much. In the context of categories, existing theories barely exclude anything. They rely on the raw data to tell us what seems to exist and what doesn't.[94] It should be clear to everyone that positing '*a*,' '*n*,' and '*υ*' merely begs the question of why we have these category-forming heads, and not others. In this domain, we have not advanced much since Chomsky's (1970) [±N; ±V]. Baker (2003) is perhaps the most comprehensive attempt in recent years to shed light on the nature of (lexical) categories, but there are reasons not to endorse his conclusions: Baker takes Nouns to be the things that bear a referential index, verbs to be the categories that project/license specifiers, and adjectives to be neither verbs nor nouns. In a minimalist context, it has become clear that specifiers do not have a special status, they are merely complements on the other side of the head as it were. Any condition that treats specifiers *sui generis* is a stipulation, rendering Baker's view on Verbs questionable. As for Nouns, is it because they bear a referential index that they are Nouns, or is it

[94] Witness, e.g., Fukui's (2006) Visibility Guideline, that relies on overt morphology alone to accept the existence of functional projections in a given language.

because they are Nouns that they bear a referential index? If the explanatory flow is from syntax to semantics, as Hinzen (2006, 2007) urges us to assume, we should be suspicious of Baker's stance. Finally, it is now clear that adjectives are composite categories, so treating them as default, as Baker does, does not strike me as promising. So, even the best attempt to offer a lexicocentric theory of categories quickly loses much of its explanatory hope.

Perhaps the most important lesson from this chapter is that free, unconstrained Merge need not be seen as giving rise to too unconstrained a theory of narrow syntax. As I have tried to show, using phases judiciously can go a long way towards curbing the power of unconstrained Merge. I say 'judiciously' because there exist proposals according to which "every phrase is a phase"[95] and the entire phrase is transferred (see Epstein and Seely (2002, 2006); Müller (2010); Bošković (2007)), which I think takes away most of the explanatory potential the notion of phase has to offer. That is, if everything that is merged gets transferred at once, my claim that Merge is regulated by phases cannot be correct. If there are too many phases, their role as anchor points is lost. Phases only become useful if there are non-phases too. Put differently, it is the delicate balance, the 'Goldilocks solution,' between phases and non-phases that takes the concept of phase beyond explanatory adequacy.

Although dramatically impoverished, the representations sent by syntax to the external systems can give rise to certain well-established grammatical phenomena. Moreover, as I pointed out in Boeckx (2009c), the present model leaves very few analytic options open to capture the range of facts that $\overline{X}$-constructs or labels were used for (it imposes certain limits on the range of analysis one can entertain for certain phenomena, e.g., ellipsis, and pied-piping, as briefly discussed in Boeckx (2009c)). Without features, there cannot be any talk of feature percolation, feature copying, feature checking, feature deletion, etc. in syntax. The range of analytical options is in fact dramatically reduced: all asymmetries deemed relevant must be related to points of transfer. From here on the task is clear: all lexical (pre-syntactic) and all syntactic

[95] This is not to deny that every application of Merge gives rise to an interpretive unit. It's just that this unit here must be viewed as corresponding not to the entire phase, but rather, to the phase complement domain. Incidentally, this is also true of Chomsky's phase model. As Epstein (2007) was the first to state explicitly, given that a part of the phase is not transferred in Chomsky's system, it is only the transferred part that is expected to correspond to an interpretive unit, not the phase as a whole. Reading Chomsky can be confusing in this context, as he often talks of phases as "having a natural characterization in terms of I[nterface]C[ondition]," as being "semantically and phonologically coherent and independent" (Chomsky (2004, 124)). For additional relevant quotes, see Gallego (2011b, 54–55).

asymmetries must be relegated to points of transfer or to post-syntactic interpretive processes that make reference to these transfer points.

True, the framework advocated here requires us to take a much more distributed view of grammatical computation than is currently assumed. Phenomena that used to fall within the purview of syntax proper, such as islands, must now be seen as interaction effects arising from properties of several grammatical components, a point that I will emphasize in the remaining chapters of this book.

The present model certainly rests more responsibility on the external systems to filter out the abundance produced by unrestricted Merge, but it would be wrong to claim that this is just a way of relegating phenomena elsewhere to keep syntax simple. As I will show in the next two chapters, putting more weight on the external systems does not mean ascribing new properties to them. As far as I can see, no new property needs to be created for SEM and PHON to accommodate the kind of underspecified syntax laid out in this chapter.

It is true that the present model puts more distance between syntax (UG) and phenomena in specific languages, so much so that I think it is wrong to ask typical questions like 'what is the syntax for *John left* in this model?' *John left* is an expression of a specific I-language, it's the result of many components interacting with one another, (universal) syntax being only one of them. The syntax of *John left* is about lexical precursor cells, transitive and intransitive phases, and nothing more. But the syntax of *John left* is then no different from that of *Mary kissed Bill* or *Who did you say Mary kissed?* Parametric choices specific to English (e.g., word order, but also specific lexical choices) fall outside of 'syntax,' but contribute significantly to the generation of *John left*.

Syntax here is very much like an embryonic process. It constructs a prepattern (hence my use of '(lexical) precursor cell') that carves derivational paths for subsequent processes (belonging to other grammatical components) that enrich syntactic representations (providing substance to form), much like embryonic processes lay down basic segmented patterns, giving rise to somites that subsequently become vertebrae, "converting embryonic time into spatial patterns," as Pourquié (2003) aptly puts it.

I see this development of our vision of syntax as very much in line with Turing's (1952) vision of biology (on this, see also Appendix 2). For Turing a true science of biology was about seeing how much could be accounted for by relying solely on what Chomsky (2005) would call 'third factor' effects, generic properties common to all organisms, perhaps even shared with the non-organic realm. What lies outside of the scope of these laws would, according

to Turing, belong to the history department; it would be accidents that many dignify by the name of natural selection. In the context of biolinguistics, what falls outside of third factors belongs to the history department too; it's the domain of culture, of specific (I-)languages, of philology, where indeed lexicocentrism reigns. But just as Turing pointed out that there can be a general biology beyond history, there can be a syntax beyond the expressions of specific I-languages. Put differently, there can be a natural language syntax for biolinguistics.

3 *Trusting in the external systems: descent with modification*

3.1 Opening remarks

Any non-lexicocentric model will have to rely on independent properties of the external systems that syntax interfaces with to recapture some of the grammatical properties that are coded in (pre-syntactic) lexical entries in more standard approaches. In this chapter I would like to examine more closely some of the properties that the present anti-lexicalist model must ascribe to the external systems if it is to move from elementary syntactic structures to more elaborate grammatical structures.

Let me state right away that I cannot hope to provide a complete list of everything one finds in the relevant external systems – such a task would require several volumes. I will therefore limit myself to some general properties, which I think should be sufficient for the reader to see how one could bridge the gap between a truly universal, invariant, minimally specified syntax of the sort advocated in Chapter 2 and the richer grammatical representations that are the concern of most linguistic analyses. It will quickly become obvious that the character of the mapping from syntax to 'meaning' ('SEM') is at present much clearer than the mapping from syntax to externalization ('PHON'), not because linguists have discovered more in the former area – if anything, far more attention has been devoted to PHON and morphosyntax – but because the mapping from syntax to meaning is likely to be much more transparent than the mapping from syntax to morpho-phonology, given that in the latter case variation has to be factored in, whereas in the former it does not. (It is generally assumed that there are no semantic parameters. The few cases that have occasionally been put forth in the literature have always received alternative treatments that dispense with the alleged semantic variation.[1] I return

[1] Chierchia (1998) is perhaps the best-known proposal concerning semantic parametrization. See, however, Longobardi (2006) and Jenks (2012) for alternatives that do not parametrize the syntax–semantics mapping.

to this issue in the next chapter.) This mapping asymmetry has been noted by Chomsky in recent writings, with passages like the following:

> It might be, then, that there is a basic asymmetry in the contribution to "language design" of the two interface systems: the primary contribution to the structure of FL may be optimization of mapping to the CI interface. (2008, 136)

> the relation of the generative procedure to the interfaces is asymmetrical, CI taking precedence: optimization is primarily to the CI interface. (2007, 12)

This is not to say that the mapping to morphophonology is completely chaotic, but it is likely to be less direct, and more articulated than what we find on the SEM side.

Before delving into some salient properties of these external systems it is important to stress that for syntax to mediate properly between 'sound/sign' and 'meaning,' avoiding mismatches that would be hard to repair without opening the door to baroque architectures for the language faculty, we must assume that the portion of syntactic structure transferred at each phase is the same for SEM and PHON (contra Marušič (2005), Richards (2008b)). This is the "best-case scenario" according to Chomsky (2004, 107):

> Assume that all three components [syntax, semantics, and phonology] are cyclic, a very natural optimality requirement and fairly conventional . . . In the best case, there is a single cycle only. [Phonology] is greatly simplified if it can 'forget about' what has been transferred to it at earlier phases; otherwise, the advantages of cyclic computation are lost.

As Bridget Samuels points out (p.c.), within the present approach, interface synchronization is not only "the best-case scenario," it is in fact the only viable option. Synchronization here is not just a question of optimizing computational efficiency, as in the passage by Chomsky just quoted, it is a matter of convergence, for only representational devices like labels, features, and the like could resolve mismatches arising from asynchronous transfer.

The present model must also assume that in addition to syntax mediating between sound/sign and meaning, there must be a direct point of contact between these two domains: given the generalized late (i.e., post-syntactic) insertion of specific lexical information, direct sound/sign-meaning pairs must be formed, for syntax never gets to see items like 'dog' or 'cat.' This architectural aspect was recognized early by Marantz in the context of Distributed Morphology. As he writes (1996, 17)

> Late insertion involves making a specific claim about the connection between LF and semantic interpretation. LF can't by itself be the input to semantic

> interpretation. If "cat" is inserted in the phonology at a node at which "dog" could just as well have been inserted – and if, as we assume, the difference between "cat" and "dog" makes a difference in semantic interpretation – then the phonological representation, specifically the choice of Vocabulary items, must also be input to semantic interpretation.

Such sound/sign-meaning pairs are not a new addition to the model forced by late insertion, such associations must exist prior to syntax, in the narrow lexicon, in more traditional models (there is, after all, nothing syntactic about *dog* meaning DOG). Moreover, evolutionarily speaking, such elementary sound/meaning associations are quite plausibly rooted in whatever system underlies vervet monkey calls or similar communicative strategies. What the emergence of syntax did is not so much create a new interface between sound/sign and meaning, but rather transform the shape of this interface. For this reason I think it is appropriate to talk of "descent with modification": syntax alone did not create grammar *de novo*. The object we call 'grammar' is grounded in cognitive properties likely to be shared in some way with other species (Hauser *et al.*'s (2002) "Faculty of Language in the Broad Sense") that in the context of an unrestricted Merge machine took on new flavors, which to us may sometimes look like radically new properties. In turn, as we saw in Chapter 2, this unrestricted Merge machine had to be supplemented by a cyclic transfer mechanism to meet the most minimal demand of the external systems, that of legibility. As Dennis Ott correctly points out

> A differently structured C-I system (one that "digests" different kinds of information) would yield different phases. (2008, 1)

> The particular phases we find in human syntax are thus not a matter of necessity; if the C-I system were structured differently, different structures would be 'picked out' … The structure of Expressions is thus [in part] not determined by narrow syntax, but by C-I properties. [footnote omitted] (2009a, 360)

Ott's remark echoes Uriagereka's (2008, chap. 1) well-taken observation that the existence of two interface levels (LF and PF) in the early minimalist model of Chomsky (1993) is a contingent fact, an accident of our species' cognitive make-up prior to the emergence of the language faculty. A similar point is made in Chomsky (2007, 15), where Hinzen's project of deflating semantics (Hinzen, 2006, 2007) is being discussed. Chomsky approves of the attempt to "reduce the primacy of CI [Hinzen's project, well captured in Hinzen (2011a, b, 2012)], [al]though satisfaction of CI conditions cannot be entirely eliminated. CI must have some range of resources that can exploit the

properties of the generated expressions" (a point that I think Hinzen does not always appreciate).

A good example to bear in mind here is the perspective on argument structure advocated by Hale and Keyser (1993, 2002). Recall that for Hale and Keyser, two things must hold for argument structure to emerge: there must be a well-defined structural template mapping onto the external systems (for Hale and Keyser, this was the *X*-bar template), produced by the syntax,[2] but *at the same time* there must be available at the interface, coming from the SEM-component, a suite of thematic notions (agent, patient, goal, etc.) to graft onto the structural template. To paraphrase Kant, structural templates without conceptual notions are empty; conceptual notions without structural templates are blind.

These remarks apply with equal force on the PHON side of the grammar: syntax being massively underspecified, everything that is late inserted must be shown to be independently available.

One final remark before delving into the properties of the external systems: one of the goals of this chapter is to show that virtually nothing new has to be added to what has already been defended in the relevant literature for an anti-lexicalist model like the present one to give rise to elaborate grammatical structures. This strikes me as a very interesting point, as it suggests that the claims put forth in the relevant literature need not depend on the lexicocentric assumptions they rely on. Eliminating these in fact yields a system with a lesser degree of redundancy.

3.2 Adding content, or making sense of syntax

I'd like to begin these reflections on the syntax–semantics interface with the interesting remarks on "naturalizing meaning" made by Searle (2006). Searle begins by pointing out that he believes that "the greatest achievements in philosophy over the past hundred or one hundred and twenty five years have been in the philosophy of language." But "[h]aving said that, however, I have to record a serious misgiving I have about the subject. The problem is that its practitioners in general do not treat language as a natural phenomenon." Searle's main objection echoes that of Pietroski (forthcoming) or Hinzen (2006) (two philosophers close to minimalist circles): "few contemporary and recent philosophers of language attempt to treat language as a

[2] Strictly speaking, for Hale and Keyser, the *X*-bar template was produced by lexical phrase structure rules. They did not in any way attempt to derive/construct it. Their perspective was still very much projectionist.

natural extension of non-linguistic biological capacities. Language is not seen as continuous with, nor as an extension of, the rest of our specifically human biological inheritance." Right here Searle is raising an important theme for what can be (and has been, see Hinzen (2008)) called "bio-semantics," thoroughly addressed in Hurford (2007).

Searle observes, rightly I think, that "there is a deep reason, both historically and intellectually, why language has not been treated naturalistically. It is because the philosophy of language went hand in hand with the development of mathematical logic." The grip of mathematical logic on the study of language (and specifically semantics) is still very strong, and I think that the sustained influence of Montague and Davidson (both students of Tarski's) on the study of natural language semantics has been a major obstacle in this project of naturalizing meaning (on this point, see especially Pietroski (forthcoming)).

Following this statement, Searle turns to the central question of "what it would mean to try to treat language naturalistically." According to him, "[t]he first step would be one that many philosophers have resisted and that is to see linguistic meaning, the meaning of sentences and speech acts, as an extension of the more biologically fundamental forms of intentionality that we have in belief, desire, memory and intention, and to see those in turn as developments of even more fundamental forms of intentionality, especially, perception and intentional action." That is to say, Searle believes that "we should see the biological foundations of language in prelinguistic intentionality." More specifically

> Our initial question should be, What are the similarities and differences between the prelinguistic forms of consciousness and intentionality and the linguistic forms? We do not know how in fact language evolved, and in the absence of fossil evidence we may never know exactly how it evolved, but we do know that it did evolve, and we ought at least to be able to answer the question, What are the logical, conceptual relations between prelinguistic forms of consciousness and intentionality and the evolved linguistic forms?

More concisely, Searle writes that "[m]any species of animals have perceptions, perform actions and are capable of acquiring beliefs, desires and intentions, though they have no language. Furthermore, several species are capable of prelinguistic thought processes. I suggest that we think of human language as an extension of these prelinguistic capacities." Searle's program echoes Dobzhansky's famous pronouncement that "nothing in biology makes sense except in the light of evolution," where evolution is meant to be understood as 'descent with modification.' Searle makes his program very explicit when he asks: "Subtract language from a species like us: What do you have? Now add language: What are you adding?"

Searle here departs from the long tradition (almost invariably traced back to Descartes, although I think this is a misrepresentation of Descartes's view; see Cottingham (1978)) that views pre-linguistic thought as so extremely impoverished as to be inexistent, and converges with much recent work on animal minds that has uncovered considerable semantic richness in species other than ours (see Carruthers (2006); Cheney and Seyfarth (1990, 2007); Hauser (2001); Hurford (2007); Jackendoff (2011) and references therein). The remarks I will make below certainly belong to this growing tradition.

It seems to me just right to stress the importance of animal intelligence, and the implications of this for our distinct conceptual structures, as Darwin had already recognized in *The Descent of Man*. After all, if our semantic component is as rich as we tend to think, and if minimalists are right that the faculty of language in the narrow sense (the uniquely human part of the language faculty) is radically minimal, then much of the richness we find in humans ought to be traced back to the conceptual richness in other species.

This is not to say, of course, that everything about natural language semantics can already be found in some (embryonic) form in other species. As Carruthers, Spelke, and others have observed in recent works (see especially Carruthers (2006), Spelke (2003)), human language seems to play an essential role in cognition. Everyone would agree that the emergence of natural language gave *Homo sapiens* a new conceptual repertoire. The question, then, becomes: how best to characterize this 'natural language semantics in the narrow sense'? Like me, Searle points at syntax as the locus of semantic innovation. Specifically, he seems to suggest that the emergence of syntax allowed for a new kind of logical form (where form is to be understood here as 'structure,' or 'representation'). As he states, "syntax organizes semantics." It organizes semantics in a way that transcends what other species seem to be able to achieve, precisely because they lack the sort of syntax that is unique to us. According to Searle, syntax takes the separate, independently available representations of reference and predication (for Searle, NP and VP, respectively) and yields a unified representation called the proposition. Furthermore, syntax provides "devices to break up the propositional content into components" (it "segments" the proposition).

Searle's characterization seems to me to go in the right direction, although it misses an important point of discontinuity between our species and the rest: the systematicity of cross-modular thinking, made possible by lexicalization (endowment of an edge feature). Following Pietroski, I take the edge feature to amount to an instruction to "fetch a concept C." But whereas concepts C and C′ may not on their own be readily combinable, owing to their belonging to different cognitive modules (by definition, the combination, or merger, of

modular concepts is restricted, or 'encapsulated'), the merging of {C} and {C′} results in cross-modular union. As I have argued elsewhere (Boeckx (2011b, c)), the presence of an edge feature on concepts is what gives Merge its unbounded character. It also captures the properties that define human uniqueness, or 'humaniqueness,' as Hauser (2009) calls it:[3] the ability

1. to combine and recombine different types of information and knowledge in order to gain new understanding;
2. to apply the same "rule" or solution to one problem to a different and new situation;
3. to create and easily understand symbolic representations of computation and sensory input; and
4. to detach modes of thought from raw sensory and perceptual input.

I take it that concepts in other species are essentially perception-based (for this reason, many philosophers resist the use of the term 'concept' in the context of species other than us, preferring terms like 'percept' or 'proto-concept'; see Hurford (2007) on this point). The adicity of these concepts is plausibly related to the perceptual contexts that root them. But once endowed with edge features, such concepts suspend these selectional restrictions, and thus detach concepts from their raw sensory and perceptual ties. In other words, thanks to the edge feature, human thought becomes truly algebraic and stimulus-free. With language, creativity emerged, understood (as did Arthur Koestler) as "the sudden, interlocking of two previously unrelated skills or matrices of thought," an almost limitless capacity for imagination, metaphorical extension, etc.

[3] Details of formulation aside, Hauser's hypothesis is a very familiar one. The essence of Hauser's claim really goes back to Descartes and his fascination with human cognitive flexibility, its fluidity, its detachment from perception, and its unbounded character – in short, its creative character. This is what led the Cartesians to claim that Man has no instinct, by which they meant that Man's cognitive faculties rise above the *hic and nunc*. This too was clear to Konrad Lorenz, who said that "man is a specialist in not being specialized" (1959). Tattersall (1998, 197) calls it "the human noncondition." and writes:

> [O]ver millennia now, philosophers and theologians have made something of an industry of debating the human condition. Even if inevitable, it is rather ironic that the very species that apparently so much enjoys agonizing over its own condition is, in fact, the only species that doesn't have one—or at any rate, whose condition, if any, is most difficult to define. Whatever condition it is, it is surely a lot easier to specify it in the case of an amoeba, or a lizard, or a shrew, or even a chimpanzee, than it is in our own.

Elsewhere (p. 207), Tattersall notes that in our case, "natural selection has gone for 'flexibility' instead of specificity in behavior."

But as I have been at pains to show in the previous chapter, in addition to unbounded Merge, one must recognize the importance of cyclic (phasal) Spell-Out in characterizing our species-specific format of thought. It is for this reason that I think that Searle is right in stressing that syntax provides "devices to break up the propositional content into components," "segment[ing]" the proposition.

Specifically, I'd like to argue that thanks to phases, syntax created constituents at the level of logical form which, once created, could be recombined in novel ways. I'd like to claim that this enables humans to move from a Davidsonian representation of events (which Hurford (2007) argues is available to other species) to a more articulated neo-Davidsonian representation of events. As I will argue, it is this shift that gave rise to some of the key properties of (human) logical form such as the 'duality of semantics' and the articulation of discourse structure, but also that of aspect, tense, modality, and quantification. (In a certain sense, if I am right, philosophy recapitulated phylogeny.)

As is well known, Davidson (1967) argued that a sentence like

(1) We bought this wine in Barcelona

is to be represented as in

(2) ∃e[bought(this wine)(we)(e) & In(Barcelona)(e)]

Such a logical form allowed Davidson to immediately account for inferences like 'If we bought this wine in Barcelona, we bought this wine.' But right after the talk where Davidson famously introduced such event representations, Castañeda (1967) argued that one could go one step further and radically separate the event argument from the verb's traditional arguments, along the following:

(3) ∃e[bought(e) & Agent(we)(e) & Theme(this wine)(e) & In(Barcelona)(e)]

Such a representation, which came to be called Neo-Davidsonian, now provides the basis of much work in natural language semantics. Compelling arguments in its favor have been given in Parsons (1990), Schein (1993, 2002), Pietroski (2005). One appealing consequence of adopting neo-Davidsonian representations is that they offer a near-perfect match between phase-based syntactic representations and semantic representations (see Boeckx (2008b); Hornstein and Pietroski (2009); Lohndal (2012)) for various ways of articulating this consequence, whose origin, I think, goes back to Pietroski (2003b)). It is indeed obvious that the conjuncts of neo-Davidsonian representations

correspond to syntactic constituents headed by phases:[4] C corresponds to the point of existential closure, *υ* to the point where the internal or external status of an argument's theta role is determined, *n*, to the type-lifting that turns a predicate into an argument, and *p*, to the point of introduction of adjuncts and arguably to the establishment of a figure–ground asymmetry. The match between syntax and semantics is even made more apparent once one adopts the independently motivated proposal of 'immediate existential closure' in Pietroski (2011). In that work Pietroski suggests that the points of limited departures from monadicity (e.g., the relation 'Agent (we)(e)') be eliminated as soon as they arise, via a local application of closure. Accordingly, 'Agent (we)(e)' is to be represented as ∃·[AGENT(e, X), WE(X)] (where '·' is the symbol chosen by Pietroski to represent the conjunction of monadic predicates). Pietroski's move is essentially the same that led Uriagereka (1999) to propose the first cyclic spell-out model (Uriagereka was concerned with keeping the derivations monotonic, adhering to Kayne's (1994) Antisymmetry approach). Accordingly, our neo-Davidsonian representation above can be rewritten as[5]

(4) ∃e[bought(e) & ∃·(Agent(e,x), We(x)) & ∃·(Theme(e,x), This-Wine(x)) & ∃·(In(e,x), Barcelona(x))]

From this perspective it can be said that points of Spell-Out in syntax correspond to points of closure in semantics.

Having made this clear I now want to turn to the issue of what happens cognitively speaking when we move from a Davidsonian to a neo-Davidsonian universe. Notice one thing that does not change from one representation to the next: nothing 'new' is needed; both representations make use of existential closure and conjunction. The only difference is that these tools cut the conjuncts more thinly in the case of neo-Davidsonian representations. But I believe that this difference is very significant, for it allows for two main 'innovations': the first one is the appearance (or articulation) of thematic predicates (Agent, Theme, etc.) on all arguments (not only on locatives), and second, it stretches logical forms and in so doing allows for the multiple occurrence of certain relations. Let me try to clarify this by means of a syntactic example: if there were

[4] Note that the conjunctions in neo-Davidsonian representations are not to be confused with the conjunction markers found in syntax. The latter likely correspond to complex bi-phasal representations of the sort argued for in Mitrović (2013).

[5] 'This wine' to be further decomposed, but this detail is ignored here. The same is true of 'bought' which should be further decomposed to make the Tense relation explicit.

only one phrase in a given syntactic representation, there would be no possibility for re-merge (internal merge), not because of the absence of the operation itself, but because there would not be enough room for reordering constituents (Anti-locality). I believe that the same is true at the level of logical forms: once stretched to its neo-Davidsonian dimension, an event representation allows for certain relations to recur.

In some sense the point I am trying to make is very obvious, but I believe that it can shed light on the fact that human thought provides a format – indeed *the same* format – for argument structure and discourse structure. Much like syntacticians have come to distinguish (using familiar categories for now) between VP and vP (e.g., Chomsky (1995)) to articulate the traditional Verb Phrase, they have also come to distinguish between ForceP and FinitenessP (e.g., Rizzi (1997)) to articulate the traditional Complementizer Phrase. Also, both can be further 'augmented' by making use of Applicative and Focus/Topic Phrases. And much like Verb phrases make room for a subject, direct object, and indirect object (and other relations to be expressed by means of prepositions), Complementizer phrases make room for evidential relations that can be exhausted by the following four categories: personal experience, direct (e.g. sensory) evidence, indirect evidence, and others (hearsay) (see Speas and Tenny (2003)).

A similar point was made by Herburger (2000), who took focus to be the result of a higher argument structure relation (in the CP domain). In a similar vein, Pietroski (2005) took quantificational relations to be a higher-order instance of an argument structure (involving an internal and external argument). The same intuition was captured by Hornstein and Uriagereka (2002) and Irurtzun (2007) in terms of movement of the relevant element (focused phrase, quantifier phrase) followed by reprojection of the moved element; essentially, a rearticulation of the event relations expressed in the VP area. (Significantly, all these authors make their proposal within a neo-Davidsonian context. As a matter of fact, I do not know how this could be expressed in another formalism.)

It could in fact be said that once event representations are expanded to allow for a VP domain and a CP domain (a double argument structure as it were), the second, higher-order argument structure can arise via external merge or internal merge, much like wh-phrases can be inserted externally (e.g., the adjunct wh-phrase corresponding to English *why*), or internally (via movement). If the external merge option is chosen, we get an articulation of existentials (a topic space, articulating the source and nature of known information); if the internal merge option is chosen, we get an articulation of the focus/quantificational

field (the field expressing 'alternatives'). (Both external and internal merge options can, of course, be combined, yielding very rich Topic and Focus fields.)

The shift from Davidsonian to neo-Davidsonian representations, then, has expanded the amount of information that can be represented in terms of elementary tools like Conjunction and Closure. It has allowed for the same event to be represented from different perspectives (new modes of presentation, one could say). By segmenting event representations, it allows for the articulation of aspect, tense, modality and quantification in terms of the same primitive predicates. The clearest expression of this was suggested to me by Jaume Solà (p.c.):[6]

(5) Segmenting events:

a. imperfective: a subevent that precedes the final subevent
b. perfective: a subevent that is the final subevent
c. (plu)perfect: a state that follows the final subevent

(6) Segmenting tensed events:

a. past: an event in an interval that precedes the 'final interval' (=present)
b. present: an event in an interval that is the 'final interval' (=present)
c. future: an event in an interval that follows the 'final interval' (=present)

(7) Segmenting worlds:

a. 'may': an event in an interval in a set-of-worlds that 'precedes' (=is-included-in) the 'total-set-of-worlds'
b. 'must': an event in an interval in a set-of-worlds that is the 'total-set-of-worlds'
c. 'should': an event in an interval in a set-of-worlds that 'follows' (=includes) the 'total-set-of-worlds'

(8) Segmenting attitude-reports:

a. 'assert': an event in a time in a set-of-worlds in an 'attitude' that precedes (=is-included-in) 'what-I-know-on-this'
b. 'know': an event in a time in a set-of-worlds in an 'attitude' that is 'what-I-know-on-this'
c. 'ask': an event in a time in a set-of-worlds in an 'attitude' that 'follows' (=includes) 'what-I-know-on-this'

[6] On the argument structure of Tense and Aspect, see also Demirdache and Uribe-Etxebarria (2000).

It is clear that this way of thinking provides us with the beginning of a reconstruction of the Cinque hierarchy (the functional articulation of the clausal spine) that does not require a very rich Universal Grammar.[7] As I argued in Boeckx (2008b), the functional sequences of the cartographers are but 'fractal' representations of elementary (argument) structures, which I here claimed become possible once we move from Davidsonian to neo-Davidsonian representations.

The fractalization in question will necessarily be bounded by the size of Spell-Out and the amount of cognitive content one can graft onto the structures generated by Merge.[8] As I indicated already, even for the most basic argument structure, thematic roles must come from some system outside syntax. Likewise for the interpretive relations 'precede,' 'is,' 'follow,' etc. The substance at issue is necessarily finite, and, I believe, reflects the (limited, modular) resources of our primate inheritance.

In many ways, the perspective on meaning that emerges in these pages converges with the deflationist/minimalist project developed in Hinzen (2006, 2007); Hinzen and Sheehan (2013). According to the latter, "[a]s grammar kicks in . . . new perspectives on the world become possible" (Hinzen (2011b)). I agree with Hinzen that "even if some forms of human thought are shared with non-human animals, a residue remains that characterizes a unique way in which human thought is organized as a system," and that is, "a grammatical way of structuring semantic information" (2013). But my position is distinct from Hinzen's in that it takes syntax to be feature-free. This ultimately means that, unlike for Hinzen, semantic categories cannot ultimately be reduced to syntactic categories. Whereas Hinzen's approach relies on an informationally rich set of functional (phase-)heads (D, C, υ), to which he reduces semantic categories, I take syntax to provide substance-free configurations. Such configurations are critical for semantic purposes, but in and of themselves, they are meaningless. For example, there is no syntactic reason to map the phase-edge–phase-complement distinction onto thematic notions like Agent and Theme. These notions must be imposed from outside of syntax. Likewise for 'Force' at the C-level: the fact that root CPs are evaluated for truth (as opposed to, say, for color) is not a fact about grammar. It is true that one can ensure this syntax–semantics transparency by assigning a [Force] feature onto the category C (equivalently, treating the root transitive phase-head as 'C'), that is, by

[7] It is even poorer than recent alternative attempts to derive Cinque's sequence such as Ramchand and Svenonius (2013).

[8] Let me be clear (to avoid misrepresentations of my position such as that made by Speas (2010)): being unbounded Merge itself does not impose the relevant constraints.

semanticizing syntax (and not, 'syntacticizing semantics,' as the cartographers often claim, cf. Cinque and Rizzi (2010)), but in so doing one has not explained why root C encodes Force. One has merely represented this fact.

Thus, when Hinzen (2013) writes that "semantics does not come with independent ontological distinctions: it uses the ones that grammar pre-configures," he must assume a lexicocentric model (with all the problems this implies, as discussed in chapter 1). Although Hinzen and Reichard (2011) write "[o]nce structure building becomes bi-phasal, a relation of predication arises, as do ontological distinctions such as 'object,' 'event,' and 'proposition' as well as relations of entailment between them," I do not see how this can be derived from syntax alone. Why precisely the categories of object, event, and proposition, as opposed to say, colors, smells, and other dangerous things?

To put the matter differently, why take "υPs as expressions that establish reference in terms of ordering relations, DPs as expressions referring in terms of distance relations and CPs as expressions referring in terms of aboutness, or of determining a spacetime," if a uniform general templatic structure is assigned to all three phases" (as Arsenijević (2013) does, in line with Hinzen's approach)? If the template is uniform, why do they use different referential coordinates?

In a feature-free model, syntax (or grammar, for Hinzen, since "grammar" is taken to be "distinct from morpho-phonology" (Hinzen and Reichard (2011))) merely provides a structural scaffolding that a primate mind endowed with (proto-)concepts can exploit in news ways, allowing it to move from a Davidsonian to a neo-Davidsonian mode of representation. As we saw earlier in this section, such a move clearly offers more referential options (it increases the number of modes of presentation), but only once some substance can be grafted onto the configurations syntax generates.

It is one thing to recognize (as Hinzen does, see especially Hinzen (2007), Hinzen and Sheehan (2013)) that "the ontology and the grammar of reference … co-vary" (2013), and that "truth and reference, as much as the ontologies of (different kinds of) objects and (different kinds of) events, are mediated by specific forms of syntactic complexity,"[9] but mediation does not imply possible reduction. The specific mapping from semantic distinction onto syntactic configurations must be explained, and this is what a lexicocentric

[9] Hinzen finds that the richer the syntactic information (at the phase-level) is, the more specific semantic reference is. This strikes me as a reflex of Panini's Elsewhere Principle at the semantic level of the grammar.

model hides. As I tried to make clear in Chapter 1, lexicocentrism presupposes the answer of the question it claims to be asking.

We still "expect non-human animal thought and reference to pattern in ways distinct from us," but our approach (I believe consistent with the evidence reviewed in Hurford (2007)) requires the pre-syntactic primate mind to be cognitively richer than what Hinzen's approach takes it to be. It also suggests that the rationale of the mapping between syntactic configurations and referential categories lies in the way language is used (see also Pietroski (forthcoming)). But in saying this, I have not explained why language is used the way it is. I have merely pointed out that syntax alone can't explain it. It is still mysterious how words like *book*, *river*, and *London* acquire the sort of semantic richness that Chomsky likes to write about (see, e.g., (2000b)). As Lenneberg (1963) already pointed out some fifty years ago, "a lexicon does not bear a direct one-to-one relationship to the physical world. Labeling the world by means of words is essentially dependent upon a peculiarly human conceptualization of reality." How we achieve this peculiar conceptualization is unclear. It is likely that many factors come into play. It may well be that some of the properties of human concepts can be derived from the fact that they have been lexicalized (i.e., endowed with an edge property). Thanks to the edge property, concepts become detached from their conceptual modules, and new, cross-modular concepts become available. By the same token, once endowed with an edge property, concepts become detached from their raw perceptual (or proto-conceptual) anchors, which may account for the very indirect way in which our (lexicalized) concepts connect to the outside world. The promiscuousness of lexicalized concepts may also account for why human concepts seem to be embedded into such rich semantic networks (fields). And surely, syntactic structures, elementary as they are, expand the referential range of 'roots.' But all these semantic effects of syntax alone don't seem to me to suffice to explain what books, rivers, and cities mean to us.

3.3 Making content public, or how to put syntax into words

When one shifts one's attention to how syntax is externalized, one immediately encounters a salient difference between how (elementary) syntactic structures are interpreted at SEM and at PHON. It is not the case that such elementary structures don't require any semantic adjustment (being feature free, they do: concepts like (proto-) theta-roles must be grafted onto them), but it is the case that there does not happen to be any room for cross-linguistic variation at that level. By contrast, at PHON, variation is rampant. Taking points of variation to

reflect areas of grammar where "UG does not mind" (to borrow Biberauer and Richards's (2006) apt phrase), one could say that UG appears to mind much less at PHON than at SEM.

With Chomsky, one could conclude from this that the design of language is more thought-friendly than communication-friendly. To me, the situation is very reminiscent of a general evolutionary logic that Deacon (2010) calls the "Lazy Gene" effect, which suggests that genes will tend to offload control of morphogenetic processes to epigenetic mechanisms in evolution whenever reliable extragenomic constraints or influences can induce the same effect. Deacon points out that "this is because such extragenomic factors will produce relaxed selection with respect to these traits and their differentiation. But this reduction in the genetic constraints increases the probability that dynamically generated combinatorial effects of epigenetic interactions will emerge, increasing the probability of generating synergistic interactions among these epigenetic processes, and – if this proves to be adaptive – a more complex phenotype." In other words, reliance on the environment tends to produce layers of complexity that a rigidly programmed organism would not achieve on its own. Deacon illustrates this logic by means of the song of the domesticated Bengalese Finch, studied by Okanoya (2012), Kagawa *et al.* (2012), which shows an increased complexity and variability of structure, increased complexity of its neural production, and an increased role for social learning, compared to its wild conspecific, the white-rumped Munia.

In the case of language, it is clear that reliance on the environment is most obvious in the process of externalization. By Deacon's reasoning, it is where we should expect – and where one finds – relaxed selection on genes (where "UG doesn't mind"), and where layers of complexity accrue.[10] Externalization goes hand in hand with grammaticalization or grammar formation: the emergence of paradigms, next to syntagms; the emergence of elaborate structures.

In this section, I wish to examine more closely the various steps required to produce this elaboration. Much like I did in the previous section, I will limit myself to pointing out that virtually everything we need can already be found in the literature. No new mechanisms are needed; it's just a matter of putting everything together, and staying away from lexicocentric dogmas.

In what follows I will draw heavily on previous work, in particular Distributed Morphology (DM; see Embick and Noyer (2007); Halle and Marantz

[10] Deacon indeed reaches this conclusion, taking the extensive offloading of language maintenance onto social transmission processes as "evidence that the human language capacity is emergent from epigenetic interactions with this extrinsic inheritance, and not a product of genetically encoded language-specific information." For reasons that should be clear, I think Deacon is right about grammar, but wrong about syntax.

(1993, 1994); Harley and Noyer (1999) for overviews) and its phase-based extensions (Marantz (2008); Marvin (2002); Newell (2008); Samuels (2011b)), and also on ongoing research being conducted in Tromsø under the banner of "Nanosyntax" (Caha (2009); Pantcheva (2011); Starke (2010); Svenonius *et al.* (2009)). These models have the advantage of seeking to reduce morphology "entirely to the function that spells out the syntactic tree by choosing and inserting phonologically contentful lexical items" (Bye and Svenonius (2012)), which is what a framework like the present one naturally leads to. But the models just cited retain a lexicocentric character by assuming that there exists a level of representation prior to syntactic derivations that guides such derivations. For DM advocates, this level corresponds to "List A" (the set of abstract morpho-syntactic feature bundles), and for Nanosyntacticians, it's the functional sequence. Because both DM and Nanosyntax assume that these objects (feature bundles and functional sequences) are syntactic trees, I'd like to construct them by syntactic means, by having the syntax precede them. In DM terms, I'd like to relocate List A to the PF-wing of the grammar, alongside List B (the set of language-particular vocabulary items). I don't mean to collapse List A and List B; rather, I wish to claim that one should recognize the existence of two lists at PF.

As a matter of fact, something like the present solution has already been reached in the literature. I have in mind here the model of "lexical insertion in two stages" put forth in Svenonius (2012a) and Bye and Svenonius (2012) (as we will see below, the model converges with certain treatments of linearization in DM). In Svenonius (and Bye)'s new variant on the model of late lexical insertion, the insertion of phonological exponents occurs in cycles, from the bottom of the syntactic tree upward, as in DM. Within each cycle, individual exponents may associate with 'spans' of functional material larger than a single head, as argued in Nanosyntax, but unlike what is allowed in DM (where lexical insertion is confined to terminals). According to Bye and Svenonius (2012), lexical insertion takes place in two stages, a syntactic one, called L-match, which makes no reference to phonology (this will be where I will locate List A), followed by a phonological one, called Insert, which makes no reference to to syntax (this will be where I will locate List B).

The two-stage process of lexical insertion is motivated by the fact that enlarging the domain of lexical insertion operations (from terminal to phase) without any other changes, would have the effect of relaxing the restrictiveness of the model, compared with standard DM. However, a high degree of restrictiveness can be achieved by taking lexical insertion to be a two-step process, with strict modular encapsulation of the two stages. As Svenonius (2012a) notes, this means that "'normal' phonology takes over an increased role in

allomorph selection (compared to standard DM, and in particular the model proposed in Embick (2010)), reducing the role of special morphophonological rules or constraints."

Here is how Bye and Svenonius (2012) construe L-match and Insert. As in phase-based DM, they assume that syntax builds tree structures from an inventory of syntactic atoms (categories/feature-bundles), and that syntax designates certain domains as cyclic (phases). Cyclic domains are processed by Spell-Out, which associates lexical entries (Vocabulary Items, or VIs) with the syntactic trees (late insertion) and linearizes them. "Spell-Out has a syntactic side, Match, and a phonological side, Insert":

(9)
a. L-match matches lexical entries to tree structures, making no reference to phonology
   (i) Targets categories
   (ii) Sees syntactic features, including both the projecting, semantically interpretable category features and the non-projecting, uninterpretable agreement features
   (iii) Sees syntactic structure, including dominance relations among features, syntactic words (the output of head-movement), and phase boundaries
   (iv) May not uniquely determine a specific allomorph for insertion: the output of Match may include alternatives
b. Insert selects exponents for realization, from the output of L-match, making no reference to syntax
   (i) Operates on exponents associated by L-match
   (ii) Sees phonological features, including segmental and autosegmental features
   (iii) Sees class features in lexical entries (declension class, conjugation class, which have no syntactic content)
   (iv) Sees phonological structure, including prosodic structure, which has been constructed in previous cycles
   (v) Sees place attributes in lexical entries (for infixation and special clitics)
   (vi) The output of Insert may be phonologically underspecified and is the input to Phonology

Following this description, Bye and Svenonius (2012) observe that "when exponents are associated with such nodes by Spell-Out, then they will have to be linearized, subject to at least three factors: universal principles of linearization, language-specific parametric settings, and morpheme-specific

Table 3.1. *Linearization suboperations, following Idsardi and Raimy (2013)*

Process	Characteristics [Module]
Merge + Phasal Spell-Out	Hierarchy, no linear order, no phonological content [Narrow Syntax]
Immobilization	Hierarchy, adjacency, no phonological content [Morphosyntax]
Spell-Out	No hierarchy, directed graph, phonological content [Morphophonology]
Serialization	No hierarchy, linear order, phonological string [Phonology]

constraints." We will return below to the principles of linearization that may be necessary ("language-specific parametric settings and morpheme-specific constraints" will be dealt with in the next chapter). But let me point out that the model under discussion converges with the 'distributed' (or 'decomposed') view of the linearization operation made explicit in Idsardi and Raimy (2013) in the context of DM. Idsardi and Raimy point out that in a model where the tasks assigned to the component called the Lexicon in earlier theories are distributed through various other components, there cannot be a direct association between hierarchical phrase structure and linear order of the sort famously argued for in Kayne (1994). According to Idsardi and Raimy, at least three steps must be distinguished to convert hierarchical structures into strings. First, the syntactic structures must be 'immobilized' (no further syntactic manipulations – reorderings, 'movements,' etc. – must be allowed, and unique representations must be selected). Second, specific vocabulary items must compete for insertion into the abstract slots that syntax provides. Finally, a total linear order must be imposed, as summarized in Table 3.1.

Returning to the topic of lexical insertion, Bye and Svenonius (2012) ascribe the following properties to L-match and Insert:

(10) L-match

a. Associate lexical items
For each category in a phase cycle, associate lexical items with matching features

b. Minimize unmatched Features
The associated lexicalization (the Association) leaves the smallest possible number of unmatched features for each category

c. Preserve equally good options
In case of a tie (i.e. same number of unmatched features), multiple competing lexicalizations are kept (to be compared by Insert)

(11) Insert: Realize, Linearize, and Phonologize

a. Bottom-up
Insertion applies from the bottom of the cyclic domain upward

b. Realize Context-Sensitive Allomorph
Where a lexical item has disjunctive exponence, then properties of the alternative allomorphs themselves determine which is realized (subject to the well-known Elsewhere Condition)

c. Maximize Links
All else being equal, an exponent with more links to the structure is preferred over one with fewer (intended to derive Minimize Exponence)

d. Linearize

(i) A non-affix is realized to the left of previously realized material in the same phase cycle (along the lines of Kayne (1994))

(ii) An affix is linearized to the right of previously realized material in the same phase cycle (following Brody (2000))

(iii) Linearization is subject to morpheme-specific position specifications of Vocabulary Items

e. Phonologize a preliminary phonological representation (the input to the phonological derivation)

I am going at such length to reproduce the details of the two sides of Spell-Out proposed in Bye and Svenonius (2012) to make very clear that the road from syntax to phonological representations is complex, and can only be made to look direct and minimal in lexicocentric models that conceal layers of derivational complexity into feature-bundles.

I am sure the reader will have noted that in a heavily syntactocentric model like the present one, Bye and Svenonius's Insert operation can be adopted without any fundamental change.[11] By contrast, L-match requires

[11] In DM, the insertion mechanism is set up in such a way that lexical vocabulary items are inserted into root terminal nodes and functional vocabulary items are inserted into functional terminal nodes. But, as De Belder and van Craenenbroeck (2013) show, this lexical–functional dichotomy is not adequate: there exist cases where functional elements are inserted in root

some modification because for Bye and Svenonius, the operation not only sees phase-boundaries, but also "targets categories, sees syntactic features (including both the projecting, semantically interpretable category features and the non-projecting, uninterpretable agreement features), and dominance relations among features."

While this type of information is standardly available in lexicocentric models, it must be reconstructed in the present model. Fortunately, again, much of the work has already been done independently. The work I would like to rely on here is Svenonius (2012b), where "a model for the generation of feature combinations which does not rely on Merge" is proposed. Svenonius starts his study by pointing out that "in addition to syntactically complex structures, natural language has morphologically complex words and featurally complex morphemes and heads." Svenonius correctly points out that "[it] is sometimes entertained that heads are not featurally complex. That is, syntactic features are syntactic atoms, so that each head bears only one feature (Kayne 2005). However, there are various indications that heads can be featurally complex" in lexicocentric models (see Chapter 1). In a model like the present one, featurally complex heads are not eliminated from the grammar, rather they arise post-syntactically.

Departing from standard accounts (but agreeing with Adger (2013b)), Svenonius (2012b) proposes that the bundling of features into heads and the formation of extended projections – precisely the type of information that L-match requires and that 'bare' Merge cannot give you – does not require Merge, and can in fact be modeled as a finite state network. Svenonius makes clear that "such a system cannot replace Merge, which is necessary for embedding one extended projection inside another." But by freeing Merge from constructing feature bundles, "Merge can be restricted to those cases where it is truly needed, and can be studied in a sharper light." According to Svenonius, "bundling" – producing a string of features – arises by general cognitive mechanisms ("third factor principles," to use Chomsky's (2005) terminology). Some of those have to do with cognitive substance ("modules or faculties, like the one distinguishing discourse participants or the one distinguishing cumulative from noncumulative instances of substance or experience"). Other constraints have to do with the computational system, and here I think Merge is key: it provides the engine to interpret the grafted substance compositionally,

positions, and likewise instances of lexical elements in functional nodes. (De Belder and van Craenenbroeck are skeptical about the latter, but I am not, given the widely attested cases of grammaticalization, which essentially amounts to the insertion of a lexical element into a functional node.)

yielding 'compositionally interpreted categories' such as 'Aspect,' 'Tense,' and the like. Yet other constraints on 'Bundling' arise from acquisition. Indeed, Svenonius (2012b) appeals to the work of Feldman *et al.* (2000), Briscoe and Feldman (2006), and collaborators such as Kemp *et al.* (2007), Stuhlmueller *et al.* (2010), who provide an algebraic theory of concept learning which makes it possible to distinguish more and less highly valued regularities in data patterns.[12] (In these works, certain kinds of regularities, such as straightforward implications, are valued over other kinds, such as exclusive disjunctions, arriving at the sort of implicational relations among features that morphologists know well.) It is these valued regularities in data patterns that the child learning her language will seek to match onto Merge. That is, she will seek to match feature-strings/extended projection lines (finite-state representations) onto the structures formed by Merge and delivered phase by phase. This is what the first step of L-match boils down to, once we leave the realm of lexicocentrism: a head/category formation algorithm, joining syntactic structures (treelets), conceptual substance, and compressed, 'optimized' finite-state representations of regularities in data patterns.

One immediate consequence of this matching between phase-sized Merge-sets and bundles (equivalently, feature-strings/extended-projection-lines) is that the morphemes/heads that will emerge will not correspond to syntactic terminals (contrary to DM). Rather, they will be what Nanosyntacticians have called "spans," i.e., units spanning a phase-complement-sized treelet. As Svenonius (2012c) observes, spans may be trivial (corresponding to traditional terminals, or 'heads'), but they may be non-trivial (strings of heads that are in a complement relation with each other, i.e., extended projection lines).

As Svenonius (2012a) further notes:

> linguistic theory of the twentieth century was based on a small number of syntactic categories, each of which bore a large number of features. These feature-laden heads were the locus of many operations. In the twenty-first century, features have been increasingly parceled out into separate heads, increasing the height of the trees but simplifying their leaves. This has allowed for a more explicit and well-grounded theory of featural interactions, but has lost an account of some of the phenomena that were previously associated with the domain of the syntactic category. The phase reintroduces a unit which groups together a number of distinct features, and so in a way is

[12] To this line of work one could add other factors entering into the acquisition of morphemes, such as those reviewed in Fasanella-Seligrat and Fortuny (2011), who build on works like Gervain and Mehler (2010), Endress *et al.* (2007, 2009), and Endress and Mehler (2009). Such factors may not be specific to humans, as Endress *et al.* (2009a, 2010) show. For further discussion, see Chapter 4.

> like the head of the twentieth-century. So what was a head is now, roughly, a phase, and what was a feature is now, roughly, a head.

This take on heads and features is a necessary consequence of abandoning lexicocentrism.

It may be said that having to match syntactic treelets and feature-strings is very onerous (why not have just one kind of representation?), but let me stress that this is only an apparent redundancy: as Hale and Keyser (1993) noted, one needs both an l-syntax and an s-syntax. One can't reduce one to the other. The best one can do is achieve an efficient mapping (matching) between the two. Contrary to recent claims that frameworks like DM reduce morphology to syntax, Hale and Keyser were well aware that "l-syntax" vs. "s-syntax" should not be conflated. Here is what they wrote:

> We have proposed that argument structure is a syntax, but we have also separated it from s-syntax ... probably an onerous distinction, perhaps nothing more than a temporary terminological convenience.
>
> We must nevertheless assume that there is *something* lexical about any verbal/lexical entry ... What is it that is lexical about the entry corresponding to *shelve*? Clearly, it is a lexical fact that *shelve* exists as a simple transitive verb in English ... in reality all verbs are to some extent phrasal idioms, that is, syntactic structures that must be learned as the conventional "names" for various dynamic events.

In effect, Hale and Keyser are pointing out that their "l-syntax" is a syntax in the representational sense (a post-syntax, a morphology, in my terminology), whereas "s-syntax" is a syntax in the dynamic, derivational sense (narrow syntax, for me). As I will repeat in the next chapter (and Appendix 3), you really need both.

Apart from the modification of L-match required by the model of syntax presented in the preceding chapter, I'd like to also make clear that the linearization algorithm that the present framework must assume (at PF) must differ from the standard one (essentially, that of Kayne (1994)). Not only does it have to take place in several steps, as we saw above (Idsardi and Raimy (2013)), it must also be able to process multi-dominance structures – a necessary consequence of Internal Merge. Fortunately again, several scholars have proposed linearization algorithms that can deal with such structures, such as De Vries (2009); Johnson (2010, 2012); Toyoshima (2011). The details do not need to concern us here, but as Johnson (2010) makes clear, the algorithm is likely to be "much freer" than Kayne's original (1994) Linear Correspondence Axiom, which required an elaborate syntax. Johnson's system, for example, puts forth a linearization algorithm that simply generates every conceivable ordering of

lexical items and lets a series of constraints (some violable) weed out those ill-formed products. (Toyoshima (2011), building on Johnson's proposals, offers a system that allows for language-wide parametric options as to how to traverse the tree, incorporating ideas from Kural (2005). Johnson too recognizes a "language particular component" in his model.)

In addition to L-Match, Insert, and several constraints on linearization, the PF-component terminates with phonology proper. Here I will simply assume that something along the lines of Samuels (2010, 2011a, b) is correct. Building on DM work, especially in its phase-based version, Samuels shows that syntactic derivations by phase, coupled with the many sensori-motor capacities that are not specific to language and are shared with many other species (see Yip (2006), Samuels (2012a, b), Samuels *et al.* (in press)), allow for elementary phonological forms to emerge (see also Boeckx and Samuels (2009)). More elaborate phonological forms result from 'external' factors, pertaining to the pressures exerted on language communication and transmission (see Blevins (2004)).

3.4 Conclusion

Summing up this chapter, I have argued that the elementary syntactic structures explored in Chapter 2 are legible at the interfaces, containing the seed asymmetries required for interpretation (both at SEM and PHON), but lacking substance needs to be supplemented in various ways to make sense, and to be made public. Building on proposals that have been made independently in the literature, I have shown how this could be achieved. Once the richness of the external systems (much of it, neither language-specific nor human-specific) is brought into light, it seems to me that Merge can be studied in a purer fashion.

4 *Elaborate grammatical structures: how (and where) to deal with variation*

Perhaps the most obvious consequence of the system elaborated so far concerns the treatment of cross-linguistic variation. The absence of relevant diversity in the pre-syntactic, narrow lexicon, coupled with the impoverished syntactic apparatus we have dealt with (unrestricted Merge and phasal Spell-Out as a regulator), necessarily entails what I have called 'Strong Uniformity Thesis' (SUT) in Boeckx (2011a):

(1) Strong Uniformity Thesis
Principles of narrow syntax are not subject to parametrization; nor are they affected by lexical parameters

That is to say, all of cross-linguistic variation reduces to realizational options available in the externalization component ('PF').

Such a consequence of the system clashes with much of the work that has taken place over thirty years under the umbrella of Principles-and-Parameters. In particular, it leaves no room for the classical notion of 'Parameter' (I insist on the adjective 'classical' here, an issue I return to below). The present chapter examines whether such a consequence is a good thing (and concludes that it is). (See also Appendix 3 for additional discussion and relevant examples.)

4.1 Parameters: caught between GB and minimalism

It is important to remember that the now standard treatment of cross-linguistic variation in terms of Parameters played a significant role in the development of minimalism. It is indeed often said that it is because the Principles-and-Parameters approach solved Plato's problem (the logical problem of language acquisition) that other concerns, lying 'beyond explanatory adequacy,' could be addressed. I continue to agree with this assertion that Principles-and-Parameters led to minimalist questions. But not because it solved Plato's

problem, but because it succeeded in showing that Principles-and-Parameters – what's universal and what's specific to particular languages – could be dissociated.

Making this thought clear will require a good deal of detailed discussion and will occupy most of this chapter. But let me try to give the gist of what I have in mind.

A detailed investigation of the nature of cross-linguistic variation revealed something quite unexpected: virtually all the points of variation could be found at the periphery of the language system. The core of language turned out to be invariant. As such, it became possible to ask whether deeper 'laws' were responsible for these principles. Recall that in the classical P&P approach, Parameters were principles (principles with options in them, to be precise). To build these options in, principles necessarily had to be fairly domain-specific. They could not be third-factor principles, which by definition are domain general. As a result, they could not follow from anything very generic. They had to be *sui generis* principles. If, however, variation turns out to be dissociable – an empirical question (which P&P resolved) – then principles can be freed from their domain specificity, and minimalist questions can be asked about them.

It took many years for this to become clear (indeed, this conclusion was not explicitly stated in the first ten years of minimalism), and many still resist it today (specific examples will be discussed below). Part of the reason for this resistance is the intrinsic interest in variation for many linguists. No one likes to see their favorite object of inquiry relegated to the margins of the system. But another, more interesting, reason for this is that the very architecture of grammar assumed in minimalism does not make this conclusion necessary. In fact, the standard model continues to leave the door open for Parameters at the heart of the theory, therefore inviting parametric treatments that run into serious problems, of the sort we will discuss momentarily. The main culprit for this state of affairs is, once again, lexicocentrism. Consider the fact that even after casting out variation from narrow syntax to the lexicon (the "Borer–Chomsky conjecture"), in his first explicitly minimalist paper, Chomsky (1993, 44) states as an axiom that "derivations are driven by morphological [i.e., featural] properties to which syntactic variation of languages is restricted." By letting syntax be driven by lexical properties (lexicocentrism), syntatic variation is ruled in, even if all the evidence points to the absence of actual syntactic variation. Here I agree with Roberts (2010b, 2011). While recognizing the existence of realizational 'Parameters,' Roberts thinks that it would be wrong to limit variation to the PF-component of the grammar, as he

sees no non-stipulative way to exclude syntactic Parameters in current minimalist models, hence, he claims, such syntactic options for variation should be exploited as well. Roberts is certainly right in claiming that in current minimalist models of syntax there is no "inherent ban on narrow syntactic variation." As the following quote from Chomsky (2001, 2) makes clear, syntactic variation can arise in current minimalist models via the influence of pre-syntactic lexical Parameters:

> Parametric variation is restricted to the lexicon, and insofar as syntactic computation is concerned, to a narrow category of morphological properties, primarily inflectional.

But, to repeat, this is true only because current minimalist models of narrow syntax are "lexiconcentric": all properties of syntax are supported by a (hyper)active and far from minimal pre-syntactic lexicon. As soon as such a view of the pre-syntactic lexicon is abandoned, the ban on narrow-syntactic variation is not a stipulation, it automatically follows from the architecture of the grammar, and to the extent that realizational strategies can account for the observed variation, such an architecture is empirically supported.

Because the overall point just made is rarely made clearly in the literature, and because the minimalist literature on variation is itself unclear (and misleading) on many points, I want to spend the time to go through the argument in detail. (Readers familiar with my publications on this issue (Boeckx (2011a, 2014c, forthcoming)), on which the following subsections are based, may want to skip ahead.)

4.1.1 Uniformity and other issues

To understand the current uneasiness existing between minimalism and the standard P&P model it is instructive to go back to an important document of the GB era: Chomsky's introduction to Lectures on Government and Binding (Chomsky (1981, 1–16)). There Chomsky outlines the P&P approach that has been pursued ever since and that Mark Baker articulated in a very accessible way in his *Atoms of Language* (2001). Chomsky makes clear that the appeal of the P&P model is that it provides a compact way of capturing a wide range of differences. As he notes (p. 6), "[i]deally, we hope to find that complexes of properties ... are reducible to a single parameter, fixed in one or another way." This is clearly the ideal of Parametric Syntax. Elsewhere, Chomsky makes clear that this ideal depends on the richness of UG: "If these

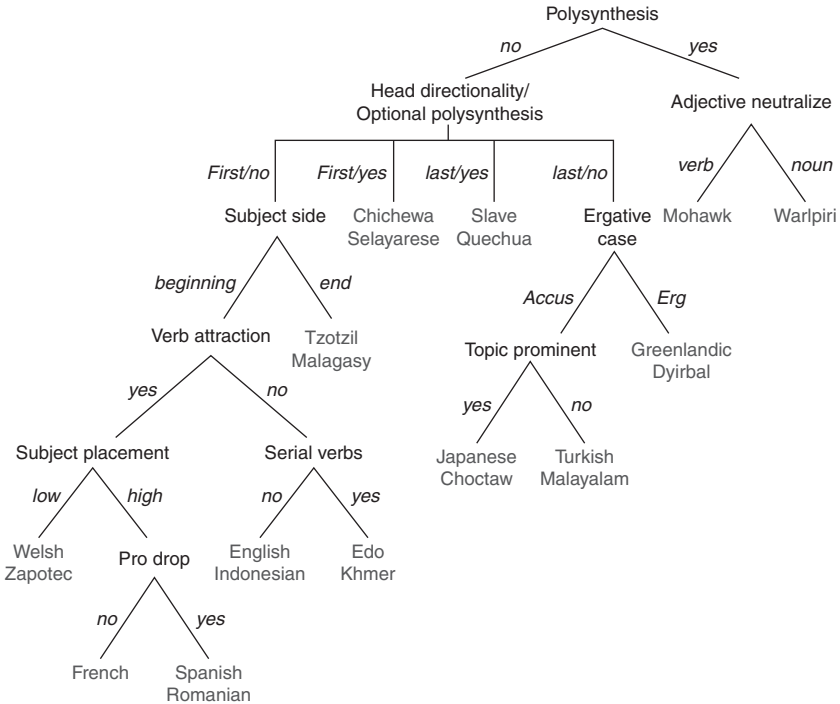


Figure 4.1. *Classical parameter hierarchy*

parameters are embedded in a theory of UG that is sufficiently rich in structure, then the languages that are determined by fixing their values one way or another will appear to be quite diverse" (p. 4). The starting assumption of Government-and-Binding was this: "What we expect to find, then, is a highly structured theory of UG" (p. 3).

In a recent paper, Chomsky (2007, 2) again makes this very clear: "At the time of the 1974 discussions, it seemed that FL must be rich, highly structured, and substantially unique." As Baker (2005) insightfully observes, the traditional P&P model takes UG to be "overspecified." This is perhaps clearest in Yang's (2002) model, where the acquisition task is reduced to choosing one among all the fully formed languages that UG makes available. In other words, the traditional Principles and Parameters model is ultra-selectionist, guided by the slogan that learning (a little) is forgetting (a lot).

Such an approach, relying on a richly structured UG, culminates in Baker's (2001) Parameter hierarchy (Figure 4.1), a (partial) characterization of the dependencies among Parameters (i.e., parametrized principles).

The most obvious question that arises in a minimalist context, where one seeks to go beyond explanatory adequacy, is: Where does the hierarchy come from? That is: What are the design principles that would make this specific hierarchy emerge?

I do not know of many works addressing this issue. I suspect that this is due in part to the fact that Baker's hierarchy makes use of concepts (such as 'topic prominence') that have never been rigorously defined in a generative framework. The hierarchy also conceals layers of complexity, well known to practitioners in the field, in the formulation of 'serial verbs' or 'pro-drop' that would undoubtedly render the hierarchy more intricate and elaborate. Indeed, once a sufficient number of detailed Parameters are taken into account, and their dependencies formalized to a degree that one cannot expect in a popular book like Baker (2001), parametric hierarchies acquire a rather different topology. Thus, consider the network coming out of the work of Longobardi and colleagues (Longobardi (2004); Guardiano and Longobardi (2005); Longobardi and Guardiano (2011)).

There is no doubt that Figures 4.1 and 4.2 are very different objects. Deriving Figure 4.1 from deeper principles may be of little help if Figure 4.2 is closer to the truth. But the lack of explicit discussion of Baker's hierarchy is also due to the fact that most syntacticians working within the minimalist program have shifted their attention away from rich, complex, parametrized principles, and toward the formulation of more basic operations (such as Merge, Agree, etc.). This is part of the shift that Chomsky (2007, 4) characterizes thus:

> Throughout the modern history of generative grammar, the problem of determining the character of FL has been approached "from top down": How much must be attributed to UG to account for language acquisition? The M[inimalist] P[rogram] seeks to approach the problem "from bottom up": How little can be attributed to UG while still accounting for the variety of I-languages attained.

Such research has (implicitly) abstracted away from the fact that most principles in the GB era were parametrized, and has assumed that things pertaining to linguistic variation will fall into place once we understand the nature of principles. In a certain sense, it can be said that such work has tacitly assumed that the core components of language can be treated independently of variation. This is, I take it, the spirit of Chomsky's (2001) Uniformity Hypothesis:

(2) Uniformity Hypothesis
In the absence of compelling evidence to the contrary, assume languages to be uniform, with variety restricted to easily detectable properties of utterances

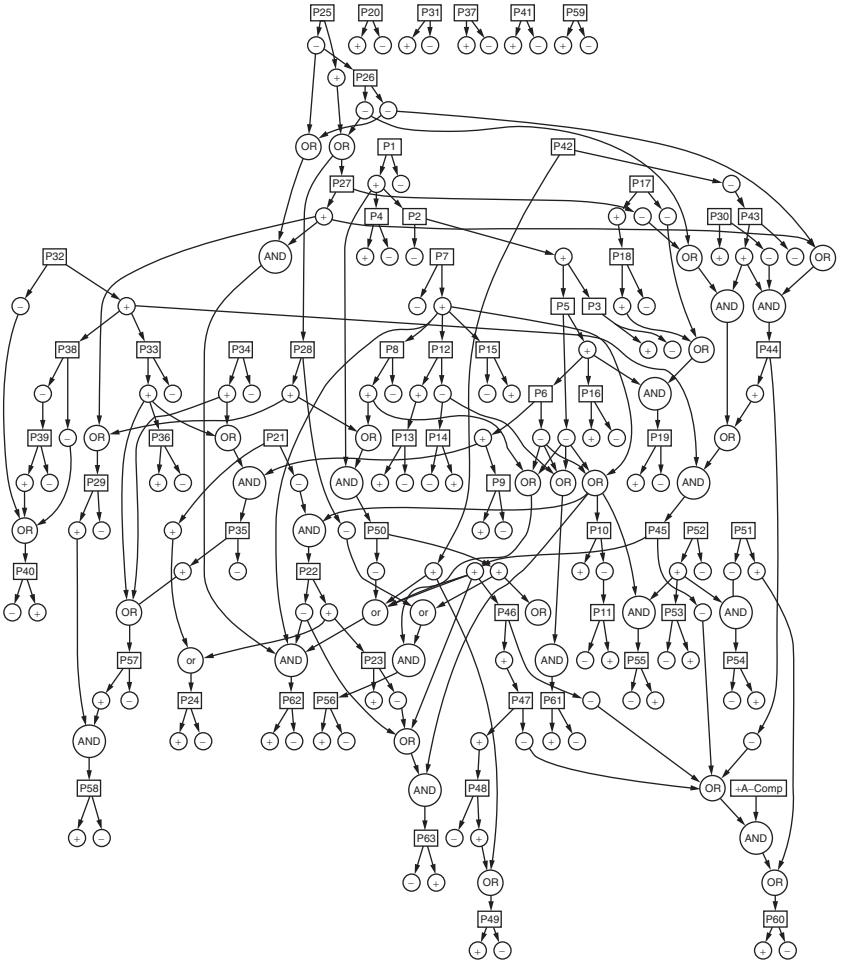


Figure 4.2. *Parametric web (entangled hierarchy)*

I believe that several factors played a role in leading Chomsky to this conclusion. The first one is the gradual decomposition of macro-parameters into micro-parameters. This certainly led to the impression of an abundance of low-level variation at the periphery. The second factor is the conjecture that variation is confined to lexical properties (the Borer–Chomsky conjecture). By distributing the lexicon and concentrating most of it to the post-syntactic component, as in Distributed Morphology, syntacticians got the first glimpses of an invariant syntax. Third, the absence of robust semantic Parameters, coupled

with the very close correspondence between syntax and semantics (interpretive transparency), also led to the belief that syntax may be invariant (if semantics doesn't vary, and it faithfully reflects syntax, then syntax may not vary either). None of these factors ensure the validity of the Uniformity Hypothesis, but they certainly led to such a hypothesis in a natural and straightforward way.

To these considerations, I think one may add more interdisciplinary concerns. The first one of these has to do with that general feeling that the exponential growth of Parameters in the literature renders the model biologically implausible. Newmeyer (2005, 84) expresses this particularly well in the following passage:

> If the number of parameters needed to handle the different grammars of the world's languages, dialects, and (possibly) idiolects is in the thousands (or, worse, millions), then ascribing them to an innate UG to my mind loses all semblance of plausibility. True, we are not yet at a point of being able to prove that the child is not innately equipped with 7846 (or 7,846,938) parameters, each of whose settings is fixed by some relevant triggering experience. I would put my money, however, on the fact that evolution has not endowed human beings in such an exuberant fashion.

Some advocates of parametric approaches have sought to downplay this problem. Thus, Kayne (2005) remarks that a big Parameter space (as big as 2^{100}, or even 2^{400}; cf. Cinque and Rizzi (2010)) "is no problem (except, perhaps, for those who think that linguists must study every possible language)." But as Roberts (2011) points out, the microparametric space is perhaps too big, for if Parameters are indeed as independent as a strictly microparametric approach would have it, languages "should appear to vary unpredictably and without assignable limits." Roberts is right in noting that this is not what we seem to find. Roberts (2001, 90), in fact, calculates that under certain reasonable assumptions concerning the number of speakers and human generations, and given a grammatical space of 2^{30}, it would take 18,000 centuries for each language type to be realized once! (bear in mind that *Homo sapiens* has only been around for 2,000 centuries).

More recently, Sheehan (2013) has argued that some degree of dependency among Parameters dramatically reduces the number of possible systems for the child to consider. Sheehan observes that if we assume the existence of dependent Parameters (in the sense of one-way implications holding between Parameter settings so that a positive setting for Parameter B depends on a positive setting for Parameter A, but not vice versa), and also assume that 30 binary dependent Parameters stand in a single hierarchy, then these Parameters yield only n+1 linguistic systems (=31). Even if they are divided into five

independent hierarchies each comprising 6 dependent Parameters, this yields 'only' $(6+1)^5$ linguistic systems (=16807).

Though mathematically correct, Sheehan's argument still relies on a simplifying assumption that we have reasons to doubt: her argument goes through if there aren't too many independent parametric hierarchies. If there are many independent hierarchies, we face the same problem as the one about independent Parameters. Sheehan seems to assume an empirical state of affairs like the one in the background of Figure 4.1, but if things are closer to Figure 4.2, not only do we expect indeed many independent parametric hierarchies (Figure 4.2 is only about the nominal domain), but we don't even have neatly nested dependencies among Parameters, but intertwined networks (webs).

Another tack taken to get around the problem at hand is offered by Smith and Law (2009). Responding to proposals that take parametric choices such as head-final/head-initial order to be reducible to physical requirements imposed on externalization, and as such, not encoded in UG (see, e.g., Boeckx (2011a)), Smith and Law write that "the physical necessity for linearization may be the ultimate cause of the parameter but the skew distribution of the world's languages and the consistency of head direction within a language suggest that the parameter does exist: The physical constraint has led to grammaticalization [genetic assimilation] of the parameter." Because genetic assimilation may be very fast (we know of examples in early human history, such as lactose tolerance), it may have been a mechanism to accommodate many Parameters into UG. But although I take genetic assimilation (or accommodation) to be real, I seriously doubt that genes accommodate epigenetic effects of this specificity. For the non-specificity of genetic coding, especially in the context of complex cognitive traits like language, see Benítez-Burraco (2009); Lorenzo and Longa (2003). To the extent that one can speak of assimilation in the case of linguistic variation, it takes place at the phenotypic level, where assimilation is simply another term for learning (see West-Eberhard (2003)).

In fact, Deacon (2010) gives ample reason to expect parametric properties not to become genetically assimilated. Drawing on attested examples from biology where properties that were once genetically coded for ceased to be part of an animal's genome as soon as the source for this property was reliably found in the animal's environment (learnable, in the context of cognition) – what he calls "relaxed selection" – Deacon suggests that the relaxation of selection at the organism level may have been a source of many complex synergistic features of the human language capacity, and may help explain why so much language information is 'inherited' socially. In other words,

languages may have such complex and varying properties precisely because UG does not code for them (underspecification, as opposed to overspecification). (And learning biases, not specific to language, may account for the typological skewing that Smith and Law took as the basis for their claim. For relevant discussion, see Culbertson (2010).)

A second interdisciplinary consideration against parametric models may be the growing sense that after a first decade of close collaboration between theoretical linguists and specialists on language acquisition, parametric proposals have lost their traction in the eyes of many experimentalists. Of course, this could be due to experimentalists favoring alternative (more empiricist) approaches (Luigi Rizzi, p.c.), but my feeling is that it also reflects the inability of parametric proposals to generate clear, testable predictions and/or their failure to solve well-articulated problems concerning possible acquisition trajectories. Part of the problem lies in the interdependency of Parameters, as was identified relatively early in the history of P&P (see Niyogi and Berwick (1996)) (a view that has resurfaced in various guises recently (Sakas and Fodor (2012); Boeckx and Leivada (2013)). There are basically too many paths the child could take, and no plausible UG strategy guarantees success.

In addition, it's proven difficult to find actual instances of Parameter setting in process. This unfortunate state of affairs (unfortunate because it leaves acquisition experts with nothing to test) was in fact elevated to the status of a hypothesis by Wexler (1998), who formulated a Very Early Parameter Setting hypothesis stating that "basic parameters are set correctly at the earliest observable stages, that is, at least from the time that the child enters the two-word stage, around 18 months of age." What about non-basic Parameters? These may be set later, but parametric studies offer no intrinsic criteria predicting which Parameter will be basic (i.e., set early) and which won't. Tellingly, P&P proponents like Yang (2002) or Rizzi (2006) appeal to UG-external factors (such as frequency or performance effects) to draw the distinction.

Interestingly, this very state of affairs was anticipated back in the early 1980s, when the notion of Parameter setting and its attendant switchboard metaphor (Chomsky (1986)) emerged. Lasnik (2002) pointed out the following:

> in the very very early 1980s, maybe 1980 or '81, … Noam in his class was laying out the theory in relation to the question of language acquisition and there was a lot of discussion in the class about how the big problem was why

> language acquisition is so rapid, given that language is such a complicated thing – but as the theory was laid out it occurred to me: Jeez, we've almost reached the point where the question should be turned around. So I raised my hand and said: "Don't we have a new question, now – Why is language acquisition so slow?" . . . Why doesn't it take six minutes? Interestingly, at that same era when the theory of parameters began to be very popular, there was a lot of work in theory and acquisition and learnability. Parameters was just the breakthrough we had been waiting for. It's been observed all around the world that kids go through discrete stages independent of the language, etc. That's an interesting fact we have to explain and the theory of parameters is designed to explain that. But I never completely believed that at the time and I still don't completely believe it. If the theory of parameters explains stages, those stages shouldn't last more than a couple of minutes each. There's gotta be something else that explains stages.

What Lasnik is stressing is that the burden of the acquisition problem falls on something other than Parameters. At the very least, it shows that Parameters are not sufficient to "solve" Plato's problem. You need to appeal to (non-grammar-based) "strategies."

This much more recent passage from Yang (2010, 1161) reaches the same conclusion:

> There was a time when parameters featured in child language as prominently as in comparative studies. Nina Hyams's (1986) ground-breaking work was the first major effort to directly apply the parameter theory of variation to the problem of acquisition. In recent years, however, parameters have been relegated to the background. The retreat is predictable when broad claims are made that children and adults share the identical grammatical system (Pinker 1984) or that linguistic parameters are set very early (Wexler 1998). Even if we accepted these broad assertions, a responsible account of acquisition would still require the articulation of a learning process: a child born in Beijing will acquire a different grammatical system or parameter setting from a child born in New York City, and it would be nice to know how that happens. Unfortunately, influential models of parameter setting (e.g. Gibson and Wexler 1994, but see Sakas and Fodor 2001) have failed to deliver formal results (Berwick and Niyogi 1996), and it has been difficult to bridge the empirical gap between child language and specific parameter settings in the UG space (Bloom 1993; Valian 1991; Wang et al. 1992; Yang 2002). The explanation of child language, which does differ from adult language, falls upon either performance limitations or discontinuities in the grammatical system, both of which presumably mature with age and general cognitive development – not thanks to parameters.

The root of the problem was well identified and illustrated in Longa and Lorenzo (2008), Lorenzo and Longa (2009), Longa and Lorenzo (2012),

Lorenzo (2013), who observed that by assuming an overspecified UG, generative grammarians ended up blackboxing development (much like the genocentric Modern Synthesis did in the context of growth of organisms), effectively rendering the role of acquisition studies irrelevant. The following passage, from Chomsky (2000a, 140–141, n. 11), makes this consequence clear: "[o]ne simplifying assumption is that L [a particular language] is literally deducible from a choice of parametric values and lexicon, so acquisition is as if instantaneous."

Let me mention one final argument against Parameters in the classical sense. The success of at least one detailed parametric proposal to reproduce results in historical language changes concerning the formation of language families (Longobardi (2004); Guardiano and Longobardi (2005); Longobardi and Guardiano (2011)) could be taken to suggest that parametric clustering is a reflection of cultural evolution rather than a reflection of the initial state of the language faculty. Although I am not aware of any explicit discussion of this point in generative studies, it has not been lost on linguists that have traditionally viewed (a rich) UG with skepticism. Thus, Dediu and Levinson (2013), correctly in my opinion, point out (p. 10, n. 10) that:

> this perspective ["the picture just sketched inverts the usual suppositions, which assume a genetically coded, fixed linguistic structure, with variable cultural uses—far more plausible is a slow accumulation of the genetically influenced motivations and contexts for language usage, making it possible to 'outsource' much of language structure to the newly evolved capacity for culture"] is entirely compatible, pace a reviewer, with recent developments in generative theory where variation is relegated to peripheral processes.

(Incidentally, several generative grammarians (e.g., David Lightfoot, p.c.) have expressed "surprise" when faced with Longobardi's results, since the traditional parametric logic does not naturally lead one to expect such a correspondence between parametric values and language change/cultural history.)

4.1.2 Artificial Parameters

In light of all these problems, one may wonder why, much like the Broca–Wernicke model in neurolinguistics, the P&P model continues to dominate approaches seeking to understand cross-linguistic variation. Though valid, this question takes on a different dimension when we examine the current literature appealing to the notion of 'parameter' more closely. Up until now, when using the term 'parameter,' I have been referring to the substantive notion of Parameter (hence the upper-case P), a notion introduced into linguistic theory

by Noam Chomsky (see Chomsky (1980, 1981)).[1] But a close reading of the literature reveals that linguists have been using a much watered-down notion of parameter (lower-case p) for quite a while, (I suspect) to provide artificial life support to the P&P model (fearing that by losing the model, we would be left with no constrained theory of variation, a point I return to in the next section, to show that it is incorrect). This notion of parameter is not what I am focusing on here, as it is clearly devoid of any theoretical teeth, hence for me does not even begin to exist in a theoretical context. But, of course, if one wishes to use the term 'parameter' as a synonym for 'difference,' then no one can deny that there are 'parameters' between John's English and Koji's Japanese. But when one does use 'parameter' thus, one should explicitly recognize that 'parameter' is "nothing but jargon for language-particular rule" (Newmeyer (2005, 53)). I take it that advocates of the P&P model are trying (or at any rate, should try) to advocate something stronger (and more interesting).

In this context, it is worth thinking about the following statement from Jan Koster:

> As for parameters, things are perhaps even worse. I cannot get into this topic here, but I believe that the notion "parameter" has hardly been developed beyond the traditional observation that there are "differences" among languages, like with respect to pro-drop or the order of head and complement. In short, the interesting principles were mostly discovered before Minimalism and the notion "parameter" has always remained underdeveloped from a theoretical point of view. (2010)

It is certainly revealing that after observing that very few linguists have taken the time to lay down a few guidelines for what counts as a Parameter, and after trying to offer such guidelines ("definitions"), Smith and Law (2009) conclude on a grim note (confirming the suspicion of Newmeyer (2005), but also of others (Culicover (1999)): "The preceding discussion implies that many of the

[1] Rizzi (1978), often given as the source of the notion 'parameter,' in fact credits Chomsky for the suggestion. For what is perhaps the earliest mention of the term 'parameter' in the generative literature, see Chomsky (1977a, 175).

> Even if conditions are language- or rule-particular, there are limits to the possible diversity of grammar. Thus, such conditions can be regarded as parameters that have to be fixed (for the language, or for particular rules, in the worst case), in language learning. . . . It has often been supposed that conditions on application of rules must be quite general, even universal, to be significant, but that need not be the case if establishing a "parametric" condition permits us to reduce substantially the class of possible rules.

It is interesting to observe, in the context of what follows in the text, that Chomsky talks about rules.

parameters postulated in the literature are, by our criteria, accidents rather than reflecting genuine, but not exceptionless, generalizations." Smith and Law are far from explicit about which of the parameters postulated in the literature remain as genuine parameters by their standard. They only mention *pro*-drop and head-directionality, but those are precisely the 'parameters' that began their theoretical lives as bona fide (macro-)parameters, only to see their scope diminish to the level of micro-parameters and possibly item-specific rules. If these are the standing parameters Smith and Law have in mind, it is conceivable that there are even fewer parameters than they think – perhaps as few as zero.

I cannot stress enough the fact that work on Parameters suffers from the disappearance of principles caused by the advent of linguistic minimalism. The dramatic reduction of principles has been pushed to the limit in recent years, with the recognition that movement is just another instance of Merge. This leaves virtually no room for Parameters, in the classical sense of the term. Recall that Parameters in Chomsky (1981) were not independent of Principles. Contrary to what the name 'Principles-and-Parameters' may suggest, it is not the case that some condition can be a Principle or a Parameter in that model: Parameters are principles (more precisely, principles with a choice point to be fixed embedded in them). If Principles disappear, Parameters can't be maintained.

The clash between Parameters and the minimalist drive is well captured in the following quote from van Riemsdijk (2008, 243f.):

> One of the main problems that we now face is the question of how the actual repercussions of such highly general principles of physical/biological organization in the grammar of specific languages can be insightfully represented . . . It would be absurd to propose that the constraint[s] [them]sel[ves] [are] parametrized.

Willy-nilly, with the classical notion of Parameter (with upper case P – the only notion worth its theoretical salt, in my opinion) gone, some linguists have developed an artificial, substance-less version of it. Holmberg and Roberts seem to concede this much when they say, as they did at a 2010 Barcelona meeting (see also Roberts and Holmberg (2009, Introduction)), that "P&P theory [i.e., the notion of Parameter] is compatible with current minimalist theorizing, once parameters are seen as effects of the absence of UG specification, but where the range of variation allowed is nevertheless constrained (often by extralinguistic factors)." Whereas they continue to defend the notion of Parameter after endorsing this view, I reject it. The reason I do so is that as Holmberg and Robert themselves acknowledged at the 2010 meeting, once this

underspecification view is adopted, "the notion of parameter is almost empty; it really doesn't have much content." Well, if it doesn't have much content, if it's almost empty, why do we hold onto it?

Elsewhere, Holmberg (2010) has again defended the notion of 'parameter' in terms of underspecification, saying that "[a] parameter is not a principle plus something, it's a principle minus something." The problem is that "a principle minus something" is just a *façon de parler*, not a Parameter (*qua* parametrized principle), at least in a minimalist context, where principles are (in the best-case scenario) generic processes or laws. Minimalist principles are completely divorced from differences, they do not contain 'minuses.' The minuses arise at the meta level, when linguists look at how these principles interact with the rest of the mind. Not being language specific, their formulation cannot contain language-specific vocabulary by means of which the 'minuses' could be defined. The correct conclusion to draw from the statement that parameters are not principles plus something is that parameters aren't, period. Their fate is that of the passive and other constructions in Chomsky (1981): taxonomic devices that are not genuine properties of the initial state of the language organ.

In addition to endorsing an underspecification view, Holmberg (2010) defends the notion of Parameter by pointing out that, contrary to claims in Newmeyer (2005), one can find empirical effects of parameters of the sort that motivated the whole parametric approach: 'octopus' or cascade effects that were intended to show how parameters facilitates the acquisition task ("macro-parameter"). Remember the following passages from Chomsky (1981):

> If these parameters are embedded in a theory of UG that is sufficiently rich in structure, then the languages that are determined by fixing their values one way or another will appear to be quite diverse . . . yet at the same time, limited evidence, just sufficient to fix the parameters of UG, will determine a grammar that may be very intricate and will in general lack grounding in experience in the sense of an inductive basis. (p. 4)
>
> . . . there are certain complexes of properties typical of particular types of language; such collections of properties should be explained in terms of the choice of parameters in one or another subsystem. In a tightly integrated theory with fairly rich internal structure, change in a single parameter may have complex effects . . . Ideally, we hope to find that complexes of properties . . . are reducible to a single parameter, fixed in one or another way. For analogous considerations concerning language change, see Lightfoot 1979. (p. 6)

Holmberg (2010) discusses contrasting data from Mainland Scandinavian and Insular Scandinavian (as well as parallel data from Finnish) to show how differences in properties like Stylistic Fronting, Quirky subjects, and the like

can be made to follow from agreement properties (along the lines originally argued for in Holmberg and Platzack (1995); see also Ott (2009b)). However, upon closer scrutiny, this kind of empirical evidence does not militate in favor of maintaining 'parameters' as interesting theoretical constructs. The reason for this is that for a given grammatical property to have collateral effects does not speak directly to Plato's problem.[2] The reason cascade effects were seen as evidence in favor of the Principles-and-Parameters in the early days of the model (when such effects seemed much more numerous than they turned out to be; cf. Newmeyer (2005)) is that they were effects for which it was hard to imagine what kind of evidence the child could use to learn them from the available data. If these effects could be made to follow automatically from other properties of the grammar for which the child could use the available data as evidence, the acquisition task was dramatically simplified. The lack of *that-t*-effects in *pro*-drop languages discussed in Rizzi (1982) was just such an effect. (Unfortunately, this particular prediction, as so many others with the same profile, turned out to be empirically incorrect; see Newmeyer (2005); Nicolis (2008); Rizzi and Shlonsky (2007).) The mere fact of finding cascade 'effects' is not an argument for the existence of a parameter if these effects could be learned by the child acquiring the language using primary linguistic data.[3] I do not have space to go through Holmberg's evidence here, but I suspect most of the effects he discussed are reasonably salient in the data available to the child, and as such could be learned even in the absence of a parametric structure.

The discussion of Holmberg's proposal illustrates well that part of the reason why parameters get used is due to the hope that they may capture typological clusters. But I think this is a dangerous path to take, for we run the risk of confusing Plato's problem (for which the classical notion of parameter was invented) and what Fasanella-Seligrat (2011) called Greenberg's problem (the

[2] I remember discussing this point with Juan Uriagereka many years ago. I am glad he voiced concerns similar to mine in Lohndal and Uriagereka (2010).

[3] I note here in passing that I am not at all convinced that the specific effects discussed by Holmberg, which were repeated at the 2010 Barcelona meeting, really are that different from one another (a very real possibility, as Holmberg himself acknowledged during the meeting). It is always possible to make numerous cascade effects emerge if one cuts the theoretical vocabulary of constructions very thinly: e.g., instead of Quirky (i.e., non-nominative) subjects, one could speak of Genitive Subjects, Dative Subjects, and Accusative Subjects, thereby making three effects emerge where there is only one. Kayne (2005) seems to make the same point when he writes "It has occasionally been thought that the term 'parameter' itself should only be used when there is such a notable or 'dramatic' range of effects. I will not, however, pursue that way of thinking here. In part that is because what seems 'dramatic' depends on expectations that may themselves be somewhat arbitrary."

issue of typological clusters). It seems to me that there are several reasons to think that typological concerns should take a backseat in a generative context.

As an illustration of this tendency to 'Greenbergize Plato,' consider the existence of "fundamental *syntactic* (and semantic) difference between English and Serbo-Croatian" that led Bošković (2008) to postulate a parameter according to which language may or may not make syntactic use of a D-layer in nominal structures (if they don't, nominal structures are NPs). Bošković shows that assuming this difference leads to significant generalizations of the following sort (Bošković (2010) lists many more):

- Only languages without articles may allow left-branch extraction of the sort illustrated here by means of Serbo-Croatian: *lijepe je on vidio djevojke* ("beautiful he saw [*t* girls]").
- Only languages without articles may allow adjunct extraction from NPs.
- Only languages without articles may allow scrambling.
- Only languages with articles may allow clitic doubling.

Notice the important modal auxiliary "may" in all of these statements. As Bošković discusses, some languages lacking an overt definite article (hence, prima facie qualifying for an NP-, as opposed to a DP-status) do not necessarily allow for the options opened to them. What this means is that these options are sub-'parameters' to be set by the child. But now notice that the DP-/NP-parameter does not work for this particular instance of Plato's problem: if the child does not know whether her language will allow left branch extraction even once it has set the DP-NP-parameter appropriately, she will have to look for evidence in the primary linguistic data to find out. Given the paucity of examples of adjunct extraction from NPs, for example, one can see that the NP/DP macro-parameter is of little help.

As this example shows, Plato's problem got confused with Greenberg's problem. Bošković (2008) explicitly commits what I like to call the typological fallacy when he writes in the context of the generalizations he has identified: "My main argument for a fundamental difference in the structure of [NPs] in languages with and those without articles concerns a number of generalizations where articles play a crucial role . . . The generalizations could turn out to be strong tendencies, which would still call for an explanation." It is true that an explanation is called for, but why should it be an explanation *in terms of parameters*?[4]

[4] I agree with Smith and Law (2009) that assuming that all differences must be treated in parametric terms – as Kayne (2005) does when he writes "I will consequently freely use the term

'Greenbergizing' Plato's problem inevitably leads one to entertain incoherent notions such as the existence of a "High Analyticity" parameter (Huang (2005)) (a continuous notion disguised as a discrete state),[5] or, when not all expected consequences of a parameter hold in a particular grammatical system, one is led to untenable conclusions such as "th[is] language is in flux" (Bošković (2008)). Grammatical systems may be highly analytic or in flux, but only in the E-language sense, not in the I-language sense with which Parameters must necessarily be associated.

It is no surprise that such incoherent notions are entertained, due to the fact that what typologists describe are not proper objects of biolinguistic inquiry (the distribution of grammatical systems will necessarily be full of historical residues and arbitrary properties that cannot be attributed to human biology, but to culture). In the words of Chomsky (1995, Introduction, n. 11)

> Thus, what we call "English" or "French" or "Spanish" and so on, even under idealizations to idiolects in homogeneous speech communities, reflect the Norman conquest, proximity to Germanic areas, a Basque substratum, and other factors that cannot be regarded as properties of the language faculty. Pursuing the obvious reasoning, it is hard to imagine that the properties of the language faculty – a real object of the natural world – are instantiated in any observed system. Similar assumptions are taken for granted in the study of organisms generally.

As a result, Parameter-based typological inquiry (especially those of the macro-parameter type) fall into the same problems that plagued most claims about holistic types from the nineteenth century and the pre-Greenbergian twentieth century: "they have not been substantiated and have fallen into oblivion" (Haspelmath (2008)). As Otero (1976) pointed out almost forty years ago, "[i]t hardly needs to be added that these archetypes are nowhere to be found." So, why look for them through Parameter-lenses?

Not surprisingly, advocates of parametric treatments of typological generalizations are often led to 'Platonize Greenberg,' and focus on idealized typological tendencies from which actual grammatical systems deviate in various ways. The clearest statement of intent in this direction comes from Cinque (2013):[6]

'parameter' to characterize all cross-linguistic syntactic differences" – renders the notion of parameter completely vacuous.

[5] If Japanese is less analytic than Chinese, but more so than French, is it more or less analytic?

[6] Cinque here is expressing ideas shared by Baker, Kayne, and others. In this context, I have not been surprised to hear from Richard Kayne or William Snyder (p.c.) that we should keep looking

> We may wonder whether something would change if we reversed this perspective; not by asking what the predominant correlates of OV and VO orders in actual languages are, but by asking what precisely the harmonic word order types are that we can theoretically reconstruct, and to what extent each language (or subset of languages) departs from them . . .
> This change of perspective entails viewing the "harmonic" orders as abstract and exceptionless, and independent of actual languages, though no less real.

I hope that the above discussion suffices to reveal the artificial nature of current minimalist notions of 'parameter.' This actually may be a good place to reflect some more on the nature of parameters, and how people have come to think of them over the years.

In the generative literature, several hypotheses have been entertained regarding the locus of variation. Perhaps the most well known is the so-called Chomsky–Borer conjecture, based on the following two statements:

(3) a. Variation is restricted to possibilities that the inflectional component makes available. (Borer, 1984, 3)
b. Variation is restricted to the lexicon; to a narrow category of (primarily inflectional) morphological properties. (Chomsky, 2001, 2)

But there are at least two other visions or 'conjectures' that one can discern in the literature, though they have not been named as such. I will do so here. The first one could be called the "Chomsky–Baker conjecture." This conjecture takes the view that there are "parameters within the statements of the general principles that shape natural language syntax" (Baker (2008a)), a view that was arguably the one in Chomsky's original (1981) formulation of Principles-and-Parameters.

The second other conjecture is of more recent vintage. It's one that takes variation to be confined to morphonological variants. It's a view endorsed in Berwick and Chomsky (2011), so let's call it the "Chomsky–Berwick conjecture." It corresponds to my Strong Uniformity Thesis ("Principles of narrow syntax are not subject to parametrization; nor are they affected by lexical parameters"). All variation is, accordingly, post-syntactic.

To be sustainable, the Chomsky–Berwick conjecture must rely on another conjecture, implicitly assumed for many years in the field, though rarely defended in a systematic fashion (but see Ramchand and Svenonius (2008)):

> for more 'abstract' kinds of parameters, even if so far we have not had any success in identifying ones that stood the test of time.

there cannot be any semantic parameter. Given the transparency of the syntax–semantics mapping (so transparent that it has led some to entertain an Identity thesis; e.g., Hinzen (2006)), if there are semantic parameters, there must be syntactic parameters. But if "Principles of narrow syntax are not subject to parametrization; nor are they affected by lexical parameters," and syntax "carves the paths that semantics must blindly follow" (Uriagereka (2008)), then, it follows that in the absence of syntactic parameters, semantics will be invariant as well. In other words, the "No semantic parameter Conjecture" directly conflicts with the Chomsky–Baker conjecture. Depending on one's take on lexicocentrism, the No semantic parameter Conjecture also conflicts with at least one reading of the Chomsky–Borer conjecture. Depending on whether by "inflectional component" (or "a narrow category of (primarily inflectional) morphological properties") one refers to a pre-syntactic (narrow) lexicon or a DM-like post-syntactic component, one may anticipate syntactic variation, and therefore semantic variation.

There are good reasons to believe that Hagit Borer herself had something like the "Chomsky–Berwick" conjecture in mind when she wrote that "the availability of variation [is restricted] to the possibilities which are offered by one single component: the inflectional component." In particular, I don't think she meant this in the way that was explored subsequently in Ouhalla (1991), Webelhuth (1992), Fukui (2006), which Chomsky made standard ("Parametric variation is restricted to the lexicon, and insofar as syntactic computation is concerned, to a narrow category of morphological properties, primarily inflectional"). As the following passage (much richer than the portion of it that is usually quoted in the literature: "Associating parameter values with lexical entries reduces them to the one part of a language which clearly must be learned anyway: the lexicon") reveals, Borer was talking about learning (constructing) rules.

> The inventory of inflectional rules and of grammatical formatives is idiosyncratic and learned on the basis of input data. If all interlanguage variation is attributable to that system, the burden of learning is placed exactly on that component of grammar for which there is strong evidence of learning: the vocabulary and its idiosyncratic properties. We no longer have to assume that the data to which the child is exposed bear directly on universal principles, nor do we have to assume that the child actively selects between competing grammatical systems. (1984, 29)

By saying that "We no longer have to assume that the data to which the child is exposed bear directly on universal principles, nor do we have to assume that the

child actively selects between competing grammatical systems," I think Borer was essentially saying that by divorcing variation from syntactic principles, we no longer need a parametric theory to support language acquisition. This is the view I advocate in this chapter.

The clarification just made allows me to return to a point I stressed at the beginning of this book. Although the "Chomsky–Borer" conjecture is often touted as progress, I do not think that this is so, largely because linguists appealing to the conjecture have never been clear about the term 'lexicon.' Indeed it seems to me that the lexical parameter hypothesis was motivated more by empirical reasons (failure of Baker-style syntactic parameters to display all of their consequences) than by explanatory reasons, despite the repeated appeal in the literature to notions like 'simplicity,' 'restrictiveness,' and, in fact, 'learnability considerations' ("since children have to learn the lexicon anyway"). How else could we account for the fact that what we wanted to understand (the nature of variation, the character of parameter) has been relegated to the part of the language organ that we understand the least: the lexicon, the part for which we have no theory?[7]

Consider the fact that many syntacticians would agree with me that parameters like Ken Hale's "Configurationality Parameter" – according to which the phrase structure in non-configurational languages is not projected from the lexicon (i.e., non-configurational languages are not subject to the what was then called the Projection Principle) – no longer fall within the realm of options they are willing to entertain, the sort of lexical parameters that they favor turn out to be far less lexical, and much more syntactic than they appear at first. Thus, many syntacticians in more recent years (see, among many others, Bobaljik and Thráinsson (1998); Fortuny (2008); Giorgi and Pianesi (1997); Pylkkänen (2008); Savescu Ciucivara and Wood (2010); Sigurdsson (2004a); Zanuttini (2010)) have argued for parameters that take the following form:

(4) "Bundling" Parameter
Given two lexical features f_1 and f_2, drawn from a universal repertoire (UG), does a given language L project f_1 and f_2 as a bundle or do f_1 and f_2 function as distinct heads in syntax?

[7] It is very curious indeed to see that proponents of the notion of 'parameter' have argued against Newmeyer's (2004, 2005) suggestion to replace parameters by rules by stressing that we have no idea of what counts as a possible rule, when the notion of parameter they propose makes use of a component of the grammar that is equally underdeveloped theoretically speaking.

Such "bundling" parameters account for a large number of parameters put forth in the literature under the Chomsky–Borer Conjecture.[8]

But what is the nature of this lexical bundling operation? As far as I can see, for the majority of the works cited, bundling is nothing but Merge operating 'in the (pre-syntactic) lexicon.' Like Merge, "Bundle" operates on lexical features, it groups them into a syntactically combinable (mergeable) set; in fact, "Bundle" builds syntactic structures (feature trees; cf. Harley and Ritter (2002a,b)). So why do we take bundling parameters to be lexical when they pertain to the availability of particular *syntactic* constructs in a way that is no less radical than some of Baker's "parameters within the statements of the general principles that shape natural language syntax"?

If "Bundle" is not Merge, then, we have a division of labor between the syntax and the lexicon as generative engines along the lines of Reinhart and Siloni (2005). If "Bundle" is Merge, Bundling parameters are syntactic parameters. Either way, although they may appear innocuous, bundling parameters fall within the "Chomsky–Baker" conjecture, or what Snyder (2011) calls "constructive parameters," since they either add new 'building blocks' (bundles) or new structure-building operations ("Bundle," or, as Gallego (2011a)

[8] As a matter of fact, the number of bundling parameters proposed in the literature increases exponentially if we take into account those lexical parameters that are in effect implicit bundling parameters. Indeed, many of the lexical parameters that focus on whether a given language *L* makes use of a given feature *f* boil down to a bundling issue, for in the vast majority of cases, it is only the presence of *the unvalued version* of *f* that is being parametrized. But "the unvalued version of *f*" is nothing but a feature bundle (see Chapter 2): {{ },*f*}.

My claim is well illustrated in the following passage from Longobardi (2005b):

> Though I agree with Benvéniste (1971) that one cannot imagine a natural language where the meaning of person (i.e. the role of individuals talked about with respect to the speech act) is really ineffable, it is the case that some languages have been argued to be deprived of syntactic effects of ϕ-features altogether, including person (e.g. person agreement on predicates or anaphors), a case in point being e.g. Japanese.

Likewise, Rizzi's take on lexical parameters (see Rizzi (2009)) – "A parameter is an instruction for a certain syntactic action expressed as a feature on a lexical item and made operative when the lexical item enters syntax as a head" – also reduces to a bundling parameter, given that instructions for syntactic actions, such as EPP features, Agree features, even Merge features, are nothing but features of features (i.e., feature bundles).

In a similar vein, Kayne's view on parameters articulated in Kayne (2005) boils down to "bundling": Kayne writes (p. 15) that "UG imposes a maximum of one interpretable syntactic feature per lexical or functional element," and since elsewhere (p. 11) he states that "every functional element made available by UG is associated with some syntactic parameter," such a parameter must pertain to the 'uninterpretable feature(s)' (whose cardinality, interestingly, Kayne does not address) on the relevant lexical/functional item; in effect, it pertains to a feature of a feature and the way it bundles with the interpretable feature.

suggests, an extra Spell-Out operation). In fact, bundling parameters are no less radical than Hale's Configurationality parameters or other recent proposals (e.g., Huang's claim that "highly analytic" languages like Modern Chinese lack l(exical)-syntax, or Reinhart and Siloni's (2005) "Lexicon–Syntax parameter" according to which UG allows thematic arity operations to apply in the lexicon or in the syntax). *All* of these proposals presuppose an active (or hyper-active) lexicon, which is more than a mere list of (vocabulary) items; one that allows the application of derivational operations. In fact, bundling parameters turn the lexicon into a syntactic component, and by doing so, they effectively turn all lexical parameters into syntactic parameters, thereby nullifying the alleged explanatory superiority of Borer's insight of locating all parameters in the lexicon, and exposing such parameters to the same critique of syntactic parameters touched on above.

4.2 If not Parameters, then what?

I am aware that many colleagues that share my 'Chomskyan' persuasion think that by claiming that there are no parameters, as I have done in this chapter, I am throwing out the baby with the bathwater, that without Parameters we are going back to the days of Skinner, or Joos, that without Parameters, we face the dreadful prospect of infinite variation, that I forget that Principles-and-Parameters is an open and flexible program, that it is so much superior to the rule-based approaches that preceded it, that no one is *that* kind of parameter-advocate any more (referring to the classical notion of Parameter articulated in Chomsky (1981)), and so on.

I don't share these fears. True, I will put more emphasis on environmental factors when I sketch my alternative in this section, suggesting indeed that we ignore insights from the learning literature at our own peril, but from there to say that I am giving up on Chomsky and buying into Skinner is too much of a stretch. I will argue in favor of a very lean (and invariant) Universal Grammar. But I will not reject UG completely. Finally, to those who think that by discarding the notion of Parameter, I am reviving the specter of infinite variation (the notorious Joos's statement that "languages can differ without limit as to either extent or direction" so often cited by Chomsky), let me point out a few things:

1. It is not at all clear that the idea of actual infinite variation was ever entertained even by scholars of Joos's persuasion (see Biberauer (2008, Introduction) for relevant discussion, accompanied by citations).

2. Even a minimal amount of syntactic invariance suffices to avoid infinite variation.
3. It is not at all clear that the exponential growth of parameters that syntacticians are willing to entertain is so much better a situation for the learner than a model without parameters at all.

I don't seem to be the only one thinking this. As Chomsky (2012b) writes, "most of the parameters, maybe all, have to do with the mappings [to the sensory–motor interface]. It might even turn out that there isn't a finite number of parameters, if there are lots of ways of solving this mapping problem" (pp. 54–55). Joos's ghost (if it ever existed) has been exorcised. The success of P&P is to have made it possible to dissociate worries about invariance from worries about variation. The existence of invariance is enough to move beyond the fear of unconstrained variation.

It is also worth bearing in mind that the return to externalization rules is not a return to the dark ages of unconstrained variation. After all, the parameter format proposed by Roberts and Holmberg (2009), like the parameter schemata of Longobardi (2005a), are much like rule formats in Chomsky and Halle (1968). In fact, the format for parameters put forth by Holmberg and Roberts at a meeting in Barcelona in 2010 – Q(ff ∈ C) [P(f)] (for some quantification Q over a set of features FF included in the set of categories C, some predicate P defined by the theory of grammar like "is a label of," "agrees," "attracts" holds of this set) – does not contain any explicit choice point, unlike the parameters of old. It's really a rule/construction format; an idiomatic template à la Hale and Keyser (1993). It is in fact the very same schema argued for by Reiss (2003) and Samuels (2009) for the formulation of phonological rules. Learning a language, then, boils down to learning its morphophonological mapping, with all its rules and their exceptions.

Fortunately, there already exists a vast literature to draw from in beginning to understand how learning this mapping takes place. The following algorithm, originally developed by Boeckx and Leivada (2014), draws on this literature in an attempt to provide an alternative to the traditional parameter-setting scenario. (What follows reproduces a section of Boeckx and Leivada (2014).) It is based on four preliminary requirements:

(5) It must

a. account for the productivity of the hypothesized rules (see Yang (2005))
b. integrate a parsing component that distinguishes between ambiguous data and unambiguous cues (following Fodor (1998))

c. tolerate exceptions also by taking into account computing time of rule-application vs. exceptions list-parsing (see Legate and Yang (2012))
d. determine which biases aid the learning process in its initial stages without assuming that the learner is already able to understand heads from non-heads or other syntactic notions

It seems clear that the acquisition process relies on a variety of factors, some of which are informed by processes also relevant in other modules of human cognition, hence fall within the third factor domain. The first principle is the ability for 'reasoning under uncertainty.' Bayesian Networks are considered as one of the most prominent frameworks for this. A key characteristic of this reasoning is the ability to entertain overhypotheses and constraints on hypotheses at the same time. As Kemp *et al.* (2007) observe, inductive learning is not possible without both overhypotheses and constraints on them entertained by the learner. Kemp *et al.* (2007) accept the innate nature of some of these hypotheses, however they argue that Hierarchical Bayesian models can help to explain how the rest of the hypotheses are acquired.

Establishing the parallelism with acquisition, the efficient learner should be able to integrate in the process of learning some conflicting tendencies, such as the need to formulate generalizations over input, without however making the acquisition task more burdensome via forming assumptions that may later be hard to retract from. More specifically, the efficient learner internalizes linguistic knowledge by making use of biases that simultaneously allow for both overgeneralizing hypotheses (Boeckx's (2011a) Superset Bias), but also for adequately constraining overgeneralizations, in line with Briscoe and Feldman's (2006) Bias/Variance Trade-off, according to which learners adopt an intermediate point on the bias/variance continuum in order to refrain from overfitting, backtracking, and reanalyzing data.

Another property of the efficient learner is the ability to pay attention to statistical properties. Many studies point out that humans are powerful statistical learners (e.g., Saffran *et al.* (1996)). Yang (2005) suggests that productivity of hypothesized rules is subject to the Tolerance Principle, which seeks to define how many exceptions to a hypothesized rule can be tolerated without the learner deciding to abandon the rule as unproductive. One of the more recent formal representations of the Tolerance Principle holds that Rule *R* is productive if $T(ime)(N,M) < T(N,N)$, with (N-M) being the rule-following items and M the exceptions. If $T(ime)(N,N) < T(N,M)$, then *R* is not productive and all

items are listed as exceptions (see Legate and Yang (2012)). This principle accurately predicts rule productivity and inference in the course of acquisition in terms of overhypotheses formulated by the learner. Paying attention to morphophonological cues is the third characteristic of the acquisition algorithm we proposed here. Prosody, for example, defines constituents and aids identifying position/edges of syntactic representation, and is plausibly rooted in our evolutionary history (Endress *et al.* (2010)). Boeckx and Leivada (2014) hypothesize that subtle points of variation (i.e. what would be referred to as microparameters in parametric approaches) should be set on the basis of explicit, saliently accessible morphophonological cues.

There are at least two more types of third factor principles that aid learning: first, the Elsewhere Condition, going back to Panini, according to which the learner applies the most specific rule when multiple candidates are possible. Second, perception and memory constraints of the sort described in Endress *et al.* (2009) and Gervain and Mehler (2010) also carry an important role. Endress *et al.* juxtapose the prevalence of prefixing and suffixing across languages with the rarity of infixing in terms of a memory constraint according to which sequence edges are particularly salient positions, facilitating learning and giving rise to either word-initial or word-final processes much more often than otherwise.

Following Yang (2002, 2010) in assuming that the child upon receiving datum s selects a grammar G_i with the probability p_i and depending on being successful in analyzing s with G_i, punishes or rewards G_i by decreasing and increasing p_i respectively, the acquisition process corresponds to a learning algorithm that integrates the principles summarized in Table 4.1

Table 4.1 presents the list of relevant biases in an unordered fashion. Put differently, the relevant factors are identified but they are not ordered, related to each other and eventually integrated in the form of a learning schema, which is what happens in Figure 4.1. Ordering them gives rise to the following algorithm that seeks to approach the acquisition process from the very beginning – as represented in Figure 4.3.

Although the algorithm presented in Figure 4.3 needs to be developed further, and importantly, put to the test, it shows that in principle it is possible to provide acquisition scenarios that do not rely on triggering innately specified values of UG-encoded parameters. Instead, the learner integrates a variety of different factors in the process of learning. The algorithm has the added appeal of keeping the assumptions about the initial state of the language faculty at a minimum.

Table 4.1. *List of biases and factors that aid acquisition*

Name	Description
A. Reasoning under uncertainty (based on Bayesian models of learning; e.g. Kemp *et al.* (2007))	Integrate conflicting tendencies in the process of learning through simultaneous entertaining of both overhypotheses as well as constraints on hypotheses
B. Superset Bias (Boeckx (2011a))	Strive for value consistency
C. Bias/Variance Trade-off (2006)	Adopt an intermediate point on the bias–variance continuum. Do so by keeping (B) a bias, not a principle, in order to avoid backtracking
D. Statistical computation (Yang (2002, 2010))	Analyze datum *s* through a hypothesized grammar G_i with the probability p_i. Depending on being successful, punish or reward G_i by decreasing and increasing p_i
E. Tolerance Principle (Yang (2005); Legate and Yang (2012))	Based on (D), turn G_i into a rule. Assume a Rule *R* is productive if $T(ime)(N,M) < T(N,N)$
F. Elsewhere Condition	Following (E), once multiple candidates are available, apply the most specific rule
G. PF-Cues Sensitivity	Fix points of variation on the basis of explicit, saliently accessible morphophonological cues. Make use of prosodic cues to define constituents
H. Perception and Memory Constraints (e.g., Endress *et al.* (2009), Gervain and Mehler (2010))	Keep track of sequence edges which are particularly salient positions in facilitating learning, giving rise to either word-initial or word-final processes much more often than otherwise

4.3 Conclusion

Like the modern synthesis did, the classical Principles-and-Parameters model blackboxed development, and dreamt of a single-level, reductionist theory to capture the generation of variation.[9] Much like what happened in linguistics,

[9] In fairness, the dream of a single-level, reductionist theory also exists on the 'other' side of the learning divide: The rediscovery of the importance of learning by experience (Saffran *et al.* (1996) and much subsequent work) brought about a revolution, opening up the radical possibility that nothing more than a sophisticated statistical learner, attentive only to surface properties of sounds in the environment, would suffice to learn any aspect of language. This too won't do.

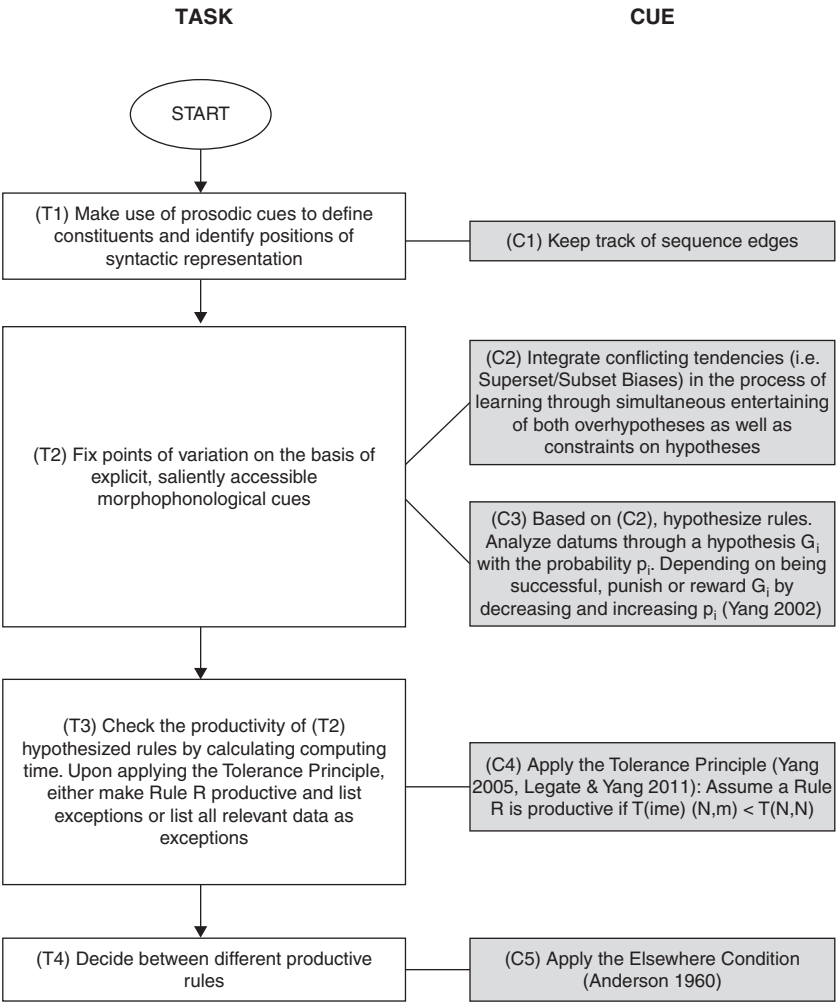


Figure 4.3. *Acquisition algorithm*

biologists were in part attempting to exorcize the ghosts of Lamarck. Linguists were attempting to minimize if not Skinnerian, at least Piagetian tendencies. But biology (and, I contend, linguistics) is now mature enough to accommodate some of the insights of alternative visions without any existentialist dilemma.

Much like modern biology, modern linguistics will have to soften its stance on various issues,[10] especially those touching on specificity and innateness (Massimo Piattelli-Palmarini would talk about this in terms of leaving behind the (necessary) age of specificity; see Piattelli-Palmarini (2010)). The range of processes explored are likely to be more abstract (less-task-dependent) and generic, nothing like the Parameters of old.

A feature-free syntax results in a very underspecified UG, with lots of blanks to be filled in. The kind of minimal syntax envisaged here provides the raw material for variation, acting as the generator of diversity. It continues to constrain variation by generating the coordinates, the boundaries of the morpho-space, but it no longer accounts (on its own) for the clusterings one finds within this morpho-space. To capture these, we need to recruit lots of factors, most of which are not specifically linguistic.

Instead of solving Plato's problem, P&P made us aware that we are just starting to scratch the surface of this marvelous yet complex construction called an I-language. I certainly don't want to minimize the impact of work within the Principles-and-Parameters tradition. Its great virtue is to have been proven wrong. It's been an inspiring error that led to the truly astonishing hypothesis that syntax is invariant, symmetric under morphophonological variation. It's made clear to all of us that learning a language is not as easy as setting a few switches correctly. It would have been nice if it had been that easy. But as the late historian Tony Judt once put it, an accurate mess is far truer to life than elegant untruths.

[10] If they don't, they are doomed to face what we may call 'Piattelli-Palmarini's dilemma.' As I pointed out in Boeckx (2006, 2010a), when Piattelli-Palmarini (1989) (rightly) cast doubt on adaptationist scenarios in the context of language evolution, there was no alternative, given the rich-UG model he assumed. The non-adaptationist alternative invoking laws of form didn't look too promising either. How could very general laws of form yield the degree of specificity that the UG model he assumed was made of? It took minimalism (and the extended synthesis in biology) to help us bridge this gap, and make the laws of form conjecture plausible.

5 *Interdisciplinary prospects*

Surveying *The Oxford Handbook of Linguistic Analysis* (Heine and Narrog (2010)), a 1,000-plus page volume aiming at "compar[ing] the main analytic frameworks and methods of contemporary linguistics," Arbib (n.d) asks himself the following question (for each chapter): "Does the theoretical stance of this chapter map onto what we can learn about the brain?" He concludes (and I concur) that the answer in virtually every case is 'no.'

Because one definitely wants to reach a more positive conclusion (the brain is the object that evolved to give rise to behaviors that came to define our species, the brain is the object that matures in the individual, making the formation of an I-language possible), one is compelled to seek alternative formulations of theoretical insights that lend themselves more readily to interdisciplinary confrontation (in the positive sense of the term). It is in the mutual interest of both linguists and scholars from adjacent disciplines to bridge the gap between mind and brain, and make substantial progress in "connecting abstract models of linguistic competence and performance, and the study of the neural implementation of the computing mechanisms" (Hinzen and Poeppel (2011)). Members of other disciplines ought to "demand from theoretical linguistics a generator of hypotheses on the different kinds of mental computations for language" (Hinzen and Poeppel (2011)), and linguists ought to be open to the consequence that "evidence from the neural implementation may put decisive empirical constraints to narrow down the vast class of formal models compatible with linguistic data" (Hinzen and Poeppel (2011)). To anyone familiar with the literature, the almost complete absence of influence of linguistic theory on neuroscience is reason to make us pause.

This is what I have tried to remedy in this book. Although 'syntactocentrism' has been criticized by numerous authors (most forcefully, perhaps, by Ray Jackendoff), it seems to me that it has not been given a fair chance, because it's never been explored in a pure incarnation. It's always been subordinated to a strong lexicocentric current, a remnant of philological thinking. Once divorced from lexicocentric tendencies – by no means an easy task,

given that these are rampant – basic operations like Merge of Spell-Out can be studied in a sharper light, and, one hopes, be brought closer to what one can learn about the brain. This has invariably been my experience in work I have carried out in parallel with the writing of this book (see Boeckx (2013b,c); Boeckx and Benítez-Burraco (2014); Boeckx and Martínez-Álvarez (2013)). Given that this is already a long book, I do not think it is the right place to go into the results already obtained on the basis of the alternative approach advocated here. Let time tell if the consequences are as positively one-sided as I think they are. After all, lexicocentrism has had over five decades to show its interdisciplinary worth. Like its cousins, genocentrism in biology and neo-phrenology in neuroscience, lexicocentrism seeks to reduce explanation to atomic components (genes, cortical maps, features). But biology has moved beyond genocentrism, and there are signs that neuroscience is on the same path to progress. It's time linguists (*qua* cognitive scientists) follow suit. It's high time we realize that if genes like FOXP2 can be a linguist's nightmare (to borrow from the title of Piattelli-Palmarini and Uriagereka (2011)), features are a biologist's nightmare, and, by the same token, become a biolinguist's nightmare.

Let's not forget what Lenneberg (1964, 76) told us fifty years ago: "[n]othing is gained by labeling the propensity for language as biological unless we can use this insight for new research directions – unless more specific correlates can be uncovered."

Appendix 1
Déjà vu all over again?

The architecture of grammar taking shape in the current work bears some resemblance to models pursued in the wake of Chomsky (1965). It is interesting to note that current trends in generative grammar are in fact reviving ambitions first expressed in light of the prospects discussed in very clear fashion in the first chapter of *Aspects of the Theory of Syntax*. Thus, we are witnessing a revival of the term Biolinguistics, a term that emerged in the early 1970s (for discussion, see Di Sciullo and Boeckx (2011)). We are also witnessing a revival of the Derivational Theory of Complexity (Marantz (2005); Boeckx (2009b)), as well as a renewed interest in Cartesian antecedent (Boeckx (2009a, 2011c)). Following early suggestions of Chomsky's *LSLT*, we are currently entertaining seriously the possibility of uniting statistical learning and the (innate) priors (Boeckx (2009b); Dillon and Idsardi (2009); Pearl (2007); Yang (2004)). And last, but not least, the distance between syntactic and semantic representations has been so reduced as to give the impression it may have vanished (see Boeckx (2009f); Hinzen (2006); Uriagereka (2008)), much like it did in the days of Abstract Syntax and Generative Semantics. Of course, current generative works are not merely restating theses first formulated in the years immediately following *Aspects*, but there is a sense in which current work vindicates Mark Twain's statement that "history does not repeat itself, but it rhymes." Some of us feel that we are once again living in exciting times, in which the prospects of interdisciplinarity are bright and the importance of theoretical linguistics, paramount.[1]

[1] Those of us that share in this Zeitgeist disagree strongly with Jackendoff's (2002) bleak prospects for Chomskyan generative grammar, which is said to have "alienated biology and

I am by no means the first to see the resemblance between some current works in syntax and works in, say, Generative Semantics. Hale and Keyser (1993) acknowledge it, and both Pullum (1996) and Culicover and Jackendoff (2005) insist on the parallelism. It is true that one can see hints of Generative Semantics in (cartographic/nanosyntactic) attempts to semanticize syntax, (antisymmetric) attempts to establish a universal base, (minimalist) attempts to eliminate D(eep)-Structure, as well as the extensive use of late lexical insertion, and constant arguments for the simplest theory, all of which were part and parcel of the Generative Semantics movement. But one can also see similarities between pre-Generative Semantics syntax and minimalism, such as Last Resort and *Syntactic Structures*-style obligatory transformations, the heavy reliance of features, the cross-linguistic component of many analyses, the lexical irregularities governing Transformations (as in Lakoff (1970)), and so on. Although such parallelisms are by no means devoid of interest, I will not pursue them here, and instead will concentrate on one of the most detailed (albeit unpublished) discussions of the nature of the lexicon in those days, namely Otero (1976).

To close this brief excursus on the history of the field, it may be worth mentioning that other frameworks also have a few things in common with Generative Semantics. As Searle (1972, 148) points out, "Those who call themselves generative semanticists believe that the generative component of a linguistic theory is not the syntax . . . but the semantics, that the grammar starts with a description of the meaning of a sentence and then generates the syntactical structures through the introduction of syntactic rules and lexical rules. The syntax then becomes just a collection of rules for expressing meaning." This is a description that fits Cartographic approaches as well as Culicover and Jackendoff's (2005) "Simpler Syntax" vision – a description that goes in the opposite direction of what I am advocating here. As a matter of fact, I think that recent attempts to deflate the syntactic component of the grammar reflect the functionalism of Generative Semantics,[2] and its fear of pure form (see Koster (1987, chap. 7)). As Chomsky notes (p.c. to M. Brame, August 13, 1975, cited in Brame (1976, 26)):

> I think the point is that they [generative semanticists] are unwilling to stray far, in abstractness, from the "given" – i.e., phonetic fact and semantic fact.

the rest of cognitive science." For reasons to be highly skeptical of Jackendoff's characterization, see Boeckx and Piattelli-Palmarini (2005, 2007), Boeckx (2013a); see also Boeckx (2006, 2005, 2009b), Marantz (2005).

[2] For illustration of this point in the specific realm of control, see Boeckx *et al.* (2010, chap. 7).

> Structures that do not directly mirror one or the other are unacceptable. Hence, the virtual demise of syntax.

What is clear from looking back at the Generative Semantics period is that, much as I have done at the beginning of this study, the issue of the nature of the lexicon occupied pride of place then. At the time of the so-called "linguistic wars," the lexicon was the most dangerous minefield. Hence my interest in Otero's work.

A1.1 Otero (1976)

At the heart of Otero (1976) is the need to recognize two distinct grammatical components:[3]

> The grammatical system of a natural language L consists of two autonomous, but closely interrelated, subsystems:
>
> i. A set of *syntagmatic* operations that pairs the pronunciation of each of the constructions it generates directly or derivatively, the set of the constructions in L being *infinite*, and
> ii. A set of *paradigmatic* relations and operations the output of which is the set of all and only the phonologically independent words, the set of words in L being *finite*.

As Otero immediately observes, this so-called "two sub-systems hypothesis" has "immediate consequences." First, "No level of the (syntagmatic) derivation of a *sentence* is represented as a string of 'morphemes'" (an idea that has been revived, without acknowledgement, in the context of nanosyntax; see Starke (2010), Caha (2009)). Second, "All processes involved in the determination of '*word* forms' and their corresponding 'word contents' belong in the paradigmatic subsystem." And finally, "'Word units' are *inserted* into an abstract phrase marker only *after* all the purely syntactic rules (nontransformational and transformational) have applied."

Otero goes on to note that

> [i]t seems clear that the [Two Subsystems Hypothesis], if basically correct, yields a much improved theory of generative grammar – one with fully differentiated but internally homogeneous components. The syntagmatic subsystem consists of a central component (the syntax) and two interpretive

[3] Because Otero (1976) remains unpublished, I decided to use extensive quotes in this subsection to help the reader appreciate Otero's argument.

> components (the phonetics and the semantics). The syntactic component consists of a recursive set of context-free phrase-structure rules and a transformational subcomponent with root transformations, one nonlocal transformation ("move C") and a set of local transformations in the sense of Emonds (to a great extent language particular), which together generate what might be called "construction forms" (cf. *LSLT*, §33.1), that is, abstract phrase markers including only syntactic category and subcategory feature specifications ... The "construction forms" will presumably be enough to derive a "logical form" ... a full interpretation can only be derived after the insertion of phonological matrices of words (in the extended sense) from the paradigmatic subsystem.

Elsewhere (Otero (1983)), Otero clarifies his proposal and notes that his "Dual hypothesis," which distinguishes between a syntagmatic grammar and a paradigmatic grammar,[4] results in an overall grammatical architecture that is "conceptually simpler."

Already then, it was clear to Otero that the "lexicon" is to be understood as a family of components, where one ought to distinguish between a "dictionary," a "lexicon in the narrow sense," and an "encyclopedia" – a three-way distinction echoed in Marantz (1996). Otero (1983) observes that "this [paradigmatic system] is the subpart that exerts a special fascination over the minds of some students of language." But it is only "the syntagmatic grammar [that] can be assumed to be a fairly direct reflection of the language faculty of the mind/brain." Citing Chomsky (who echoes Jespersen), Otero notes that "no student of human language ever dreamed of a universal dictionary." For, if

> [a] syntagmatic grammar is essentially universal (biologically given in essence), a paradigmatic grammar is, to a considerable extent, a historically evolving subsystem, burdened with the weight of the past, like other cultural systems. Only a paradigmatic grammar can be fossiliferous. This brings to mind the distinction between "core grammar" and a "periphery" of "borrowings, historical residues, inventions, and so on, which we can hardly expect to – and indeed would not want to – incorporate within a principled theory of UG." (Chomsky (1981, chap. 1))

Otero is quick to add that even if "[e]very paradigmatic grammar is, to a considerable extent, language particular, and to some extent fossilized ... this is not to say that everything in the paradigmatic grammar is language-particular,

[4] A distinction that harks back to Saussure's view that, next to *Langue*/*Parole*, and synchrony/diachrony, a theory of syntagms vs. a theory of associations is "the most basic rational division" for the linguist.

or that the paradigmatic grammar doesn't have any internal structure. The model of paradigmatic grammar that is emerging in the most recent investigations is itself highly modular."

Interestingly, Otero (1983) observes that under his model, "there are no processes such as 'affix hopping' in the syntagmatic grammar, and plausibly there is no 'copying' of agreement features in the course of the syntagmatic derivation," something that is also true of the syntactic component in the present approach.

A1.2 Radicalizing Otero's proposal

I have gone at some length to quote passages from Otero (1976, 1983) because, like Otero, I am in favor of a sharp divide between a pure syntactic component, freed from imperfections, and another system which (as Sapir would have said) necessarily "leaks," containing as it does a fair share of accidental properties. It seems to me that this separation is necessary if we are to entertain the strong minimalist thesis seriously. Why this separation has not been made standard yet is something of a puzzle. Already back then, Otero (1976) points out that "[g]iven the theoretical framework Chomsky had developed in [Aspects], it is somewhat surprising that he did not go on to draw what, from a generative perspective, appears to be a very natural, if not inescapable, conclusion, namely that morphemic representations play no role in the (syntagmatic) derivation of a sentence."

I have found two references to Otero's work on the lexicon in Chomsky's writings. Both are buried in footnotes. The first is in Chomsky (1980, 277 n. 10), where Chomsky writes that "[o]ne might nevertheless argue that full lexical insertion, including phonological and morphological properties of words, takes place at the level of S-structure, along lines that have been suggested by Carlos Otero ... and Hans den Besten ..."[5] The second reference is in Chomsky and Lasnik (1977, 432, n. 18), where they state that "[i]n fact, there is little reason to suppose that lexical items are inserted in base structures, in this theory ... We will continue to accept this assumption here for ease of exposition, but everything we say can be translated into an alternative theory in which lexical insertion takes place in surface structure and only abstract features are

[5] Chomsky here refers to unpublished work by Hans den Besten to which I have not been able to gain access.

generated in the base (which is now limited to the categorial component) in positions to be filled by lexical items."

Contrary to what Chomsky and Lasnik claim, I think that there is more at issue than just "ease of exposition." First, keeping the syntagmatic component in the same batch as the paradigmatic component invariably leads, one way or another, to lexicocentrism, and all the problems discussed in previous chapters in this book. Segregating the two components, as Otero suggests, leads to a level of representation of the syntactic component that is not only far more explanatory, but also truly invariant, as I have shown. As Otero (1996, 321) already noted, if one adopts the Two Subsystems Hypothesis, one is led to claim that "there is only one language, as the evolutionary biologist would expect." That is to say, for purposes of linguistic description, Chomsky and Lasnik's stance may well be a matter of "ease of exposition." But when it comes to the biolinguistic program, there is far more at stake. Failing to segregate the two subsystems boils down to claiming that syntax is at every level a parametric syntax at best. It is never a truly principled syntax, as it is always relativized to a lexicon tied to a particular language. As a result, the possibility of a truly principled explanation, of the sort demanded by minimalism,[6] is out of reach. It is always a construction-specific syntax. A syntax that is indeed shaped by the lexicon, as Caponigro and Polinsky (2011) assume; a syntax about which it can be said (as Baker does in the metaphor of the Swiss watch discussed in Appendix 2) that "small changes in the properties of a single word or class of words, or the addition of a single grammatical principle can have large scale repercussions on the entire language." Incidentally, it is also the reason why every characterization of syntax (as an independent generative system) by Jackendoff (and others) deals with issues that are confined to parameters (case, word order, etc.). As the following quote (taken from Culicover and Jackendoff (2005, 22)) reveals (see also Jackendoff (2010)), the rationale for syntax is frustratingly highly language-specific (i.e., tied to specific languages):

> Should *all* syntactic structure be slashed away? Our goal, a theory of syntax with the minimal structure necessary to map between phonology and meaning, leaves open the possibility that there is *no* syntax at all: that it is possible to map directly from phonological structure (including prosody) to meaning. Although some people might rejoice at such an outcome, we think it

[6] On "principled explanation," see Chomsky (1972, 173, 182, 183, 185), to be contrasted with "engineering solution[s]," "useful and enlightening, but . . . of roughly the order of complexity as the original problem[s]" (Chomsky, 2000a, 93).

> is unlikely. Perhaps this represents a certain conservatism on our part, and someone more daring will be able to bring it off. But at a minimum, we believe that *syntactic categories* such as noun and verb are not definable in purely semantic terms – and that fundamental syntactic phenomena such as agreement and case-marking are based on these categories. And we believe that there are *syntactic constituents* whose categories are determined (for the most part) by the categories of their heads, i.e., that there is something like X-bar phrase structure. We think it is not a matter of phonology or semantics that English verbs go after the subject, Japanese verbs go at the end of the clause, and German inflected verbs go in second position in main clauses but at the end in subordinate clauses. We think it is not a matter of phonology or semantics that English sentences require an overt subject but Italian sentences do not; that English has ditransitive verb phrases but Italian does not; that English has *do*-support but Italian does not (but see Beninca' and Poletto 2004 for a northern Italian dialect that does have Do-support); that Italian has object clitics before the verb but English does not. That is, we are going to take it for granted that there is some substantial body of phenomena that require an account in terms of syntactic structure. It is just that we think this body is not as substantial as mainstream generative grammar has come to assume. This is why we call our hypothesis "Simpler Syntax" rather than just plain "Simple Syntax."

The architecture I am advocating is in some sense (and with the benefit of hindsight) a radicalization of Otero's proposal (much like my proposal can be seen as a radicalization (and generalization) of Hale and Keyser's (1993, 2002) proposal concerning argument structure, and also a radicalization of Distributed Morphology's late insertion mechanism). Otero still maintains "syntactic category and subcategory feature specifications" as part of the syntagmatic component. I am doing away with even this much.

There is another reason why I think that there is more to Otero's proposal than just a matter of "ease of exposition." By keeping separate the syntagmatic and paradigmatic systems, Otero makes it possible to view the lexicon as non-distinct from the (paradigmatic) grammar. (By lexicon here, I mean the post-syntactic lexicon. As Piera (1985, 311, n. 2) already noted, "[t]he term 'lexicon' is clearly inappropriate for the complex component envisaged by these theories [see also Otero 1976].") Over the years, Jackendoff has argued that Chomsky (following Bloomfield) has been wrong in "regarding the lexicon as altogether separate from the grammar" (Jackendoff (2005)) (see also Jackendoff (1997, 2010)) – for reasons that are compelling (roughly speaking, they are the reasons why so much work in morphology assumes a realizational, post-syntactic component that views "words" as phrasal). Otero's separation thesis could be recast (in light of the present proposal) as a separation between

syntax and grammar, leaving the door open for the grammar and the lexicon to be one.[7] It can also be seen as a separation between principles and parameters. A separation that is necessary if we want to move beyond explanatory adequacy.

[7] A position that can make much better sense of the claims found in the "grammaticalization" literature, as I argue in Boeckx *et al.* (in press).

Appendix 2
Switching metaphors: from clocks to sand piles

By any measure the mental organ called the language faculty is very complex. Moreover, as Hauser, Chomsky, and Fitch (2002) emphasized, our capacity to develop (internalize/cognize) a language is far from monolithic, consisting as it almost surely does,[1] of many ingredients, many of them of distinct – and non-linguistic – origin. It is because of this complexity that Chomsky, beginning with his *Logical Structure of Linguistic Theory* in 1955, insisted on distinguishing various levels of 'representation' (equivalently, 'structure') to be studied (as a first pass) in relative isolation. These levels correspond to traditional subdisciplines within linguistics: phonology, syntax, semantics, morphology, etc. To this day, the approach guides theoretical linguistic practice, as well as the practice in neurolinguistics, where researchers attempt to locate these various levels in brain wet-ware. By focusing on the concept of interface, minimalism has turned what everyone passively acknowledged (that these various levels must function in tandem) into an active area of research, to the point that the autonomous status of some levels is being questioned (witness Marantz's (1995) "end of syntax" pronouncement). Certainly, the distributed character that some levels have acquired (as is for example the case with morphology) is indicative of this shift in perspective, which promises to be more fruitful in the context of mind/brain unification, given the generally distributed character of higher-level brain processes. What needs to happen now is linking this distributed character to the Hauser–Chomsky–Fitch question of which aspect

[1] I view this recognition of a-monolithicity as a biological imperative. Denying this would make the language organ unlike everything else we know about how biological structures come about (see F. Jacob's notion of *bricolage*).

(if any) is specifically linguistic. Marcus's (2006) descent-with-modification approach to modules is a step in this direction. As a matter of fact, if, as I have argued in this book, features are nothing more but nano-molules, a descent for modification is called for in this domain as well. This will inevitably entail a softening of the 'specificity' stance that is generally adopted in generative circles. Almost certainly, what are now considered highly specific linguistic properties will reduce to a conjunction of various non-linguistic, generic factors – ultimately, the recognition of a vastly underspecified linguistic code, and of the importance of emergence – the perspective that I find missing in generative/I-linguistics circles;[2] the very perspective that defines what Brian Goodwin called 'generative biology.'

The Gestalt-shift just mentioned will be felt most clearly in the context of syntax, traditionally the level of representation taken to be most specific and unique (for good reasons, see Anderson (2004) for a good overview based on comparative ethology). It is in the domain of syntax that an 'approach *rigorously* from below' is most urgently needed, one that genuinely begins the analysis from the ground up.

As Dick Lewontin has often pointed out (the following is taken from Lewontin (2000, 3))

> It is not possible to do the work of science without using a language that is filled with metaphors . . . While we cannot dispense with metaphors in thinking about nature, there is a great risk of confusing the metaphor with the thing of real interest . . . The result is that the properties we ascribe to our object of interest and the questions we ask about it reinforce the original metaphorical image and we miss the aspects of the system that do not fit the metaphorical approximation.

In this appendix I would like to cast doubt on a metaphorical image that continues to dominate syntactic studies (and, in my view, vitiates them), and suggest that the metaphor that it is usually contrasted with may, in fact, be more conducive to progress (in the direction of 'beyond explanatory adequacy').

Consider now the following passage, quoted in full, from the beginning of Baker (1999):[3]

[2] For a very rare exception, see Uriagereka (1998). Emergence has been exploited in connectionist and other approaches that are all too often contrasted with the Chomskyan/I-linguistics perspective. One of the lessons of the present book is that we ought to reconsider that literature in a more positive light.

[3] Baker's discussion ought to be considered in the light of Bybee (2010), who talks about sand piles in the context of language unity and diversity.

Why are there so many languages in the world—on the order of 5000, not counting many others that have become extinct in historical times? This question would be fairly easy to answer if languages differed only in relatively superficial ways, such as which set of sounds is conventionally used to refer to a particular kind of concrete objects, or the details of how that set of sounds is produced phonetically. The word for any given object is clearly arbitrary in most cases, so one choice is as good as another, and speakers in different parts of the world can be expected to make different choices. Similarly, there are different ways of accommodating sounds to one another in the phonology and phonetics that decrease effort while maintaining contrasts: again it is not surprising that languages make different choices in these matters. However, linguistic diversity goes far beyond these local and superficial matters. This is harder to explain in a priori terms, and it places an important boundary condition on the task of constructing an adequate linguistic theory.

In this paper, I present and defend the view that one important reason why there are so many different languages in the world is because all human beings are equipped with a detailed and rigid Universal Grammar. At first, this sounds paradoxical to most people. Indeed, the discoveries of linguistic typological research are often taken to tell against Chomskian notions of Universal Grammar. However, a strong case can be made that these facts show just the opposite.

To appreciate the logic of the situation, consider whether human languages are more like piles of sand or Swiss watches. Many people think of languages as loose connections of many more or less unconnected words, morphemes, inflections, and syntactic constructions. These various pieces come into languages and pass out of them in a quasi-continuous fashion by the various processes of historical change. If this picture is more or less correct, then languages are similar to piles of sand, since those are a paradigm case of objects made up of many small parts without any rigid or prespecified relationships to each other. Now one striking fact about piles of sand is that they all look pretty much the same. Sometimes beach-side artists work hard to produce striking counterexamples, but after a few hours even the most impressive dragons, mermaids, and castles revert to looking like a basic mound. From this viewpoint, distinguishing traits like English Exceptional Case Marking structures or Dyirbal ergative Case marking would be expected to come and go, with their presence or absence having little effect on the language as a whole, and all languages tending toward a relatively homogeneous mush.

On the other hand, suppose that language is more like a Swiss watch, which has a complex and rigid internal structure. One important property of this kind of system is that small changes in one localized part can be passed on, even magnified, so that they have large effects on the behavior of the whole. For example, a small twist of the knob on one side can change the configurations of the hands on the face. A small change in the tension of the mainspring could change the speed at which those hands turn. A small gear removed or added anywhere in the watch could cause the whole system to stop. My claim

> is that this is similar to what happens with human language: small changes in the properties of a single word or class of words, or the addition of a single grammatical principle can have large scale repercussions on the entire language. This happens precisely because everything is bound together into a structured whole by a relatively complex and rigid unified universal grammar. That is why there are so many languages, and why those languages are sometimes so different in their grammars.

The watch vs. the sand pile is a useful contrast for the message that I want to convey. Baker sides with the clock model, insisting on a "detailed and rigid" Universal Grammar, one that (in contrast to the sand pile) "prespecifies relationships" (among lexical units), one where different languages are, in a deep sense, quite different from one another. A Universal Grammar that has a "complex and rigid" internal structure, of the sort defended by Mark Baker, corresponds to the view of UG approached from above (in the sense of Chomsky (2007)), and is very reminiscent of the view put forth in Chomsky (1981).

The image of the watch is also an image made famous in biology by Paley, the author of *Natural Theology*, who greatly influenced Darwin.[4]

> In crossing a heath, suppose I pitched my foot against a stone, and were asked how the stone came to be there; I might possibly answer, that, for any thing I knew to the contrary, it had lain there for ever: nor would it perhaps be very easy to show the absurdity of this answer. But suppose I had found a watch upon the ground, and it should be inquired how the watch happened to be in that place; I should hardly think of the answer which I had before given, that, for any thing I knew, the watch might have always been there. Yet why should not this answer serve for the watch as well as for the stone?

Paley is now remembered for offering the watch metaphor (the image of intricate complexity) as an argument in favor of the existence of a watchmaker, which Darwin destroyed. Darwin did so by appealing to Adam Smith's metaphor of the invisible hand. But it is important to stress that although Darwin eliminated the need for a watchmaker, he did not eliminate the watch. He

[4] Darwin has this to say about his reading of Paley in his *Autobiography*.

> In order to pass the B.A. examination, it was, also, necessary to get up Paley's Evidences of Christianity, and his Moral Philosophy ... The logic of this book and as I may add of his Natural Theology gave me as much delight as did Euclid. The careful study of these works, without attempting to learn any part by rote, was the only part of the Academical Course which, as I then felt and as I still believe, was of the least use to me in the education of my mind. I did not at that time trouble myself about Paley's premises; and taking these on trust I was charmed and convinced of the long line of argumentation.

argued for a blind watchmaker (to use Richard Dawkins's apt characterization), but left the watch intact. That is to say, Darwin did not deny the existence of intricate complexity. And modern biologists have followed him in this respect. In fact, most of them would subscribe to the view that "the theory of evolution by cumulative natural selection is the only theory we know of that is in principle capable of explaining the evolution of organized complexity" (Dawkins (1996)). The designer was replaced, but the design stayed the same.

Very occasionally, however, a few brave souls have attacked the appropriateness of the watch metaphor and the idea of intricate, built-in complex design. David Hume, in his *Dialogues Concerning Natural Religion*, writes the following:

> The world, says he, resembles the works of human contrivance; therefore its cause must also resemble that of the other. Here we may remark, that the operation of one very small part of nature, to wit man, upon another very small part, to wit that inanimate matter lying within his reach, is the rule by which Cleanthes judges of the origin of the whole; and he measures objects, so widely disproportioned, by the same individual standard. But to waive all objections drawn from this topic, I affirm, that there are other parts of the universe (besides the machines of human invention) which bear still a greater resemblance to the fabric of the world, and which, therefore, afford a better conjecture concerning the universal origin of this system. These parts are animals and vegetables. The world plainly resembles more an animal or a vegetable, than it does a watch or a knitting-loom. Its cause, therefore, it is more probable, resembles the cause of the former. The cause of the former is generation or vegetation. The cause, therefore, of the world, we may infer to be something similar or analogous to generation or vegetation.

A more explicit attack of the watch metaphor, one that I find particularly compelling, was made some 200 years later by Alan Turing. To linguists, the name of Alan Turing is inextricably linked to the conception of computation (the Turing machine) and, perhaps, also to the role of language in cognition (the Turing test). But I would like to suggest that to biolinguists (at least those taking the internalist perspective advocated by Chomsky via the term *I*, linguistics), Turing should be first and foremost remembered for his contribution to the field of morphogenesis (a term Turing himself introduced).

In his seminal (1952) paper, Turing addressed what he clearly saw as a (if not, *the*) central problem of biology, namely, how the cell (the zygotic cell of conception) manages, through strictly chemical and physical means, to grow into the far more complex structures of the fetus, the baby, and the mature organism, creating all along *new* information and structure. Turing went on to

propose a mechanism, a reaction-diffusion process, that is at the heart of much current work on biophysics. Turing describes his approach thus:

> Unless we adopt a vitalistic and teleological conception of living organisms, or make extensive use of the plea that there are important physical laws as yet undiscovered relating to the activities of organic molecules, we must envisage a living organism as a special kind of system to which the general laws of physics and chemistry apply. And because of the prevalence of homologies of organization, we may well suppose, as D'Arcy Thompson has done, that certain physical processes are of very general occurrence. (Turing and Wardlaw (1992, 45))

In a remark to Robin Gandy, Turing explicitly states that his new ideas were intended to "defeat the argument from design" (cited from Hodges (1983, 431)). Turing here was not alluding to Paley. He was, I'm sure, assuming that Darwin had put the conscious designer to rest. As Leiber (2002, 86) correctly notes, Turing, rather, endorsed the D'Arcy Thompson view that the teleological "evolutionary explanations" endemic to (neo-)Darwinian adaptationist biology are (to quote Leiber) "non-fundamental, fragile, misdirected, and at best mildly heuristic." Or as Saunders interprets Turing's remark, "The primary task of the biologist is to discover the set of forms that are likely to appear [for] only then is it worth asking which of them will be selected" (1992, xii). Turing was thus addressing the question that Darwin did not address, namely the origin of species (or, more generally, the origin of forms, the variety on which selection feeds). Interestingly, Turing wanted as much as possible to show that this central question could be studied by appealing to "strictly chemical and physical means." Furthermore, these were anticipated to be "of very general occurrence." His was thus an inquiry into the explanatory power of "third factor" principles (to use Chomsky's (2005) terminology).

Turing was approaching complexity as something to construct, and explain by simple means. He was not denying the existence of an additional layer of historical baggage (an extra layer of complexity), but he clearly saw that not all complexities are equal. Some are more understandable than others. Only the complexities that can be explained by strictly physical and chemical means, i.e. those complexities that can be studied in a lawful, a-historical, universal and generic fashion, can be the subject matter of a science called biology. The rest must be left to the historians.

In so doing, Turing was outlining[5] a research program that has grown into an important field of study: Complex Systems, where themes of self-organization,

[5] In so doing, Turing was also harking back to a research tradition that neo-Darwinians had relegated to the furthest margins of biology, the Rationalist Morphology tradition of Goethe,

emergence, chaos, and dynamics figure prominently. I like to think of this field as "population physics" (to be contrasted[6] with "population biology," the fundamentally novel perspective introduced by Darwin, according to Mayr (1959), and much subsequent work), as it studies the behavior of large numbers of entities over which it averages (for a very useful survey, see Ball (2006)). The field of Complex Systems initiated by Turing has grown and diversified. It received key contributions from big thinkers influenced by Turing such as Ilya Prigogine,[7] who named Turing structures "dissipative structures," Brian Goodwin, Stuart Kauffman, Stuart Newman, Ricard Solé, and many others.

The various approaches to Complex Systems all explore the conditions under which structure and order can form, spontaneously, out of a homogeneous medium, avoiding (very locally, and temporarily) the inevitable grasp of the second law of thermodynamics. One of these approaches, pioneered by Per Bak (see Bak *et al.* (1988), Bak (1996)), has focused on the phenomenon of 'self-organized criticality' and interestingly, has used the sand pile (one of the images used by Mark Baker in the quote above) to convey its essential properties.

Bak *et al.* (1988) chose the image of the sand pile to show how complex pattern formation can arise robustly irrespective of the fine-grained details of the system: keep adding grains of sand onto a sand aggregate and at some point you will reach a critical sand mass that will cause an avalanche (or a suite of avalanches) until the system returns to a 'barely stable' state, one where avalanches are likely to be triggered soon afterwards if sand is being added.

The image of the sand pile in the context of self-organized criticality is important when one considers Baker's quote above. Yes, a sand pile appears to be made up of (quasi-)identical grains of sand, loosely connected, and without any pre-specified pattern. But Bak and colleagues showed that out of this homogeneous state patterns (of the sort Baker alluded to when he is talking about the Swiss watch) do emerge, with "small changes passed on, and magnified, so that they have large effects on the behavior of the whole."

Geoffroy, Owen, and D'Arcy Thompson, to whose *On Growth and Form* Turing alludes in the quote above. I discuss this tradition in Boeckx (2009a), which I relate to the rationalist cognitive science tradition that Chomsky (1966) called "Cartesian Linguistics." Both of these traditions inform some current Evo-Devo trends in biology, and the renewed focus on Biolinguistics in current linguistics. For more on this, see Boeckx (2010b). For relevant discussion, see also Benítez-Burraco and Longa (2010); Medina (2010).

6 On the contrast between physics and biology I am alluding to here, see Boeckx (2006, chap. 4).

7 Prigogine attended some of Turing's lectures on morphogenesis, and is reported to have said that he spent a day in vigorous conversation with Turing (see Hodges (1983, 564) and Dupuy (2009, 130), who cites a paper by Isabelle Stengers on this connection which, despite my best efforts, I have not been able to access).

What this means is that complex-wide repercussions triggered by small, seemingly insignificant changes robustly emerge without requiring detailed, pre-specificed, rigidly encoded instructions ("grammar" or "lexicon"). The complex whole is the output of simple processes and interactions, rather than the result of complex specifications.

I believe that this was the essence of what Turing wanted to convey in his 1952 paper, and I think that this is the message that (bio)linguists (and most biologists) have not yet assimilated. It is this failure to assimilate Turing's lesson (perhaps because it appears to conflict with the table of instructions that is part of the computational vision that Turing brought forth in his famous 1936 paper, and that is at the heart of classical cognitive science, the so-called computational theory of mind?) that invariably draws us back to preformationism, (thinly) disguised as genocentrism (for biologists) and lexicocentrism (for linguists) (and, come to think of it, phrenology for neuroscientists). Sand piles (think of hourglasses) and Swiss watches may perform the same function, but depending on which metaphor one chooses, it may lead to irreducible complexity or beyond explanatory adequacy. Since simplicity is the only road to scientific progress (i.e. progress in explanation and understanding), we should favor the model that enables us to built complexity simply, even if it means leaving out of the explanatory scheme quite a few irreducibly complex phenomena.

Appendix 3
More on the loss of syntactic variation

Contrary to what I have argued in Chapter 4 is the right position to adopt, a few linguists continue to insist on the existence of what Snyder (2011) calls "constructive parameters." Such parameters, within which, according to Snyder, syntactic parameters fall, have "the effect of adding new 'building blocks' or new structure-building operations."

For example, Snyder puts forth his well-known Compounding Parameter as an example of such constructive parameter. Snyder's Compounding Parameter rests on the observation that languages differ sharply in whether they allow endocentric, bare-root compounding as a fully productive process. Thus, whereas English allows for things like *university lab space committee decision*, Catalan does not. Snyder noted that English-type languages display certain constructions that are systematically absent from Catalan-type languages, such as Verb–NP–particle constructions, adjectival resultative constructions, and so on. On the basis of this, Snyder put forth the idea that English-type languages allow for a "General Modification" rule[1] that is lacking in Catalan-type languages. It is in this sense that the Compounding Parameter qualifies as a "constructive parameter," since "its positive setting provides a semantic composition rule for syntactic combinations that would otherwise be uninterpretable" (Snyder (2011)).

[1] I do not think that the specific details of Snyder's formulation of the rule matter for the point that I am making in the text, but for the sake of concreteness, here is Snyder's latest formulation of his General Modification rule (from Snyder (2011)):

> If α and β are syntactic sisters under γ, where α is the head of γ and denotes a kind, then interpret γ semantically as a subtype of the kind α, and as standing in a pragmatically suitable relation to the denotation of β.

There are, however, several problems with Snyder's proposal, all of which are characteristic of the difficulties suffered by parameters (in the classical sense) over the years. First, the full range of grammatical consequences predicted by the parameter on the basis of data from English does not always hold across a wider range of languages, as can be gathered from Son (2006), Son and Svenonius (forthcoming) (Snyder himself recognizes that Basque is a problem for his proposal). This typically leads to a weakening of the parameter (from biconditional, 'Property X iff Property Y' statements to weaker, unidirectional, 'Property X if Property Y' claims), leaving open how the child figures out if indeed the relation between property X and property Y holds in her language. Second, a proposal like Snyder's faces difficulties in explaining the existence of compounds in Catalan-type languages. To be sure, these are not as productive as in English, but their existence begs the question of how they were generated in the first place if the rule behind them – by hypothesis, the General Modification rule put forth by Snyder – is set to 'off' in the language. Lack of productivity is a matter of language use. Third, Snyder's proposal reveals the importance of properly formulated principles needed to embed parameters in. In a restrictive semantic framework such as Pietroski (2005), the limited repertoire of interpretive operations makes it impossible for languages to 'deactivate' some of them. Indeed, it is difficult to see how a language would be able to completely do away without as general a rule as General Modification. Fourth, there exist much more 'surfacy' explanations for the highly limited availability of compounds of the English type in Catalan-type languages, such as Tokizaki (2010, 2011), who shows how the cross-linguistic variation that Snyder is concerned with can be predicted on the basis of the canonical word-stress location in languages (specifically, the difference between right-oriented stress languages and right-edge stress languages).

As Sugisaki (2011) points out, Tokizaki's analysis suggests that the core difference between English and Catalan is not due to a constructive parameter, but merely a result of "externalization" – a possibility that is gaining support in the literature concerned with cross-linguistic variation (witness Mathieu (2011); Richards (2010)), and to which I return below. Similar considerations hold for the much-discussed difference between verb-framed and satellite-framed languages, first discussed by Talmy and often related to Snyder's Compounding parameter, in light of proposals like Acedo-Matellán (2010), who claims that the variation at issue depends on the morphological properties of the functional prepositional domain and not on the availability of a syntactic process of manner incorporation. That the morphophonological properties of prepositions would be relevant in the context of Snyder's Compounding Parameter is

clear from the fact that Catalan-type languages insert prepositional material to support compounds.

Incidentally, the compounding parameter could be understood (in line with the overall argument in Chapter 4) as a point of UG underspecification: when two categories of the same type merge with one another, the point of symmetry must be resolved (see Richards's (2010)) Identity Avoidance condition), either by immediate external Merge (insertion of a preposition between the two elements, the Romance pattern in the case of compounds), or by internal Merge (movement of one of the two elements, see Moro (2000)) upon insertion (external merge) of a higher head. At bottom, these are Spell-Out options.

In a vein very similar to Snyder's assertion, Baker (2008a) contends that there are large-scale patterns in crosslinguistic variation that require "parameters within the statements of the general principles that shape natural language syntax." Baker (2011) cites the following proposals, mostly taken from his own work, as illustrations of such syntactic parameters:

- The symmetrical object parameter (Baker 1988)
  Languages can have {1, more} "objects" (=structural Acc case).
- The Polysynthesis Parameter (Baker 1996)
  Languages {must, need not} express all theta-roles as morphemes on the verb.
- The case dependence of agreement parameter (Baker 2008b)
  Functional heads {must, need not} assign case to an NP they agree with.
- The direction of agreement parameter (Baker 2008b)
  The goal of agreement {must, need not} c-command the agreeing head.
- Parameterization in the minimal link condition (Baker and Collins 2006)
  {The closest, any} NP can move into the Specifier of a functional head.
- Parameterization in the Case filter, whether NPs have case (Diercks 2011)
  NPs {are, are not} generated with an unvalued case feature.
- Parameterization in how case is assigned (Baker and Vinokurova 2010)
  Object case and agreement {are, are not} two sides of the same coin.

Going into the details of each of these proposals would require many pages of illustrations, and I won't do this here. I will limit myself to a few general remarks. The first one is that it is quite striking that many of the parameters listed above pertain to morphological realization of case and agreement markers, whose syntactic status has been questioned (see, among others, Bobaljik (2008)), and can relatively straightforwardly be reformulated as PF/realizational parameters (i.e., Spell-Out rules). Second, some of the

parameters proposed by Baker clearly exploit a lack of (one hopes, temporary) restrictiveness in syntactic theory. For example, Baker exploits several of the theories of case assignment of the literature, and claims that these various theories correspond to different parametric settings. But the intention of the proponents of such theories is clearly more universalist than this. Most of them would assume that in the fullness of time only one option for case assignment/agreement will be valid. The relevant syntactic parameters are therefore expected to go away. It is worth noting in this context that Baker is not alone in exploiting an unresolved theoretical ambiguity and turning it into a parameter. Lasnik (2000) did the same in trying to capture the well-known difference in verb placement differentiating English from French. As I wrote in Boeckx (2012a), "[i]t is indeed quite remarkable to see that all too often it is only lack of understanding that leads one to claim that a certain property attributed to the language faculty is taken to be parametrizable. It is as if variation were the default." But the recent history of syntactic theory leads us to expect that the better we motivate our syntactic constructs, the more they will turn out to be invariant – which is one of the main reasons for me to claim that narrow syntax is completely immune to variation, and that parameters of the sort Baker has formulated don't hold.

The third remark I would like to make concerning the type of parameters Baker posits is that the one that is arguably the best worked out of all of them, his Polysynthesis Parameter (Baker, 1996), has been subject to strong criticism, and has been shown to be clearly inadequate. For instance, several authors have pointed out that languages traditionally characterized as polysynthetic do not display the full range of properties predicted by Baker's parameters. In part due to this fact, but also due to the fact that some of the main signatures of polysynthesis such as Noun Incorporation vary so much across polysynthetic languages, some have argued in favor of a reformulation of the Polysynthesis parameter in terms of microparameters (see, e.g., Adger *et al.* (2009); Legate (2002)), which are much more easily recast as realizational rules (see, e.g., Mathieu and Barrie (2011)).

Last, but not least, it seems clear to me that the theoretical syntax community has (if only tacitly) decided that the reasonable success of accounting for principles like Relativized Minimality in terms of deeper design properties in recent years renders parametrizable versions of such principles unacceptable, even if such versions offer adequate descriptions of the data (though, even here, alternatives readily exist; cf. Jeong (2007) and Schneider-Zioga (2013) for reanalyses of the facts that led Baker and Collins (2006) to propose their specific parameterization in the minimal link condition). To repeat the words of van Riemsdijk (2008, 243f.), already quoted in Chapter 4:

> One of the main problems that we now face is the question of how the actual repercussions of such highly general principles of physical/biological organization in the grammar of specific languages can be insightfully represented . . . It would be absurd to propose that the constraint[s] [them]sel[ves] [are] parametrized.

Baker (2011) writes that "collecting a fuller range of good cases [of parameters such as those listed above] should help with the high-level theoretical work of discerning what can be a parameter and why." But it seems to me that we already have enough of an understanding of what syntactic principles are to be confident that parametrization of the minimal link condition can be excluded. So, I strongly disagree with Baker when he says that "we should remain open to the possibility of deeper/more extreme parameterization, at least until we know more about crosslinguistic syntactic variation at the highest level." There is, of course, always more to be learned, but I find it remarkable that Baker, who is on the one hand so confident about the robustness of syntactic principles as to be able to write (in support of macroparametric analysis, as opposed to microparametric analysis) that "it is already feasible to compare unrelated languages in an interesting way. This is possible because of the universal principles, which constrain crosslinguistic variation so that the dangers of incomparability and undiscernable interfering variation are not so dire," can be so *un*confident about their robustness as to entertain parametrized versions of them.

In sum, I think it is fair to conclude that the position advocated by Snyder and Baker, a position very close to the original idea of parameter in generative grammar, is now a minority view. At the conceptual, or theoretical, level, we have come to understand that (to borrow an observation of Kayne's, formulated at a meeting in Barcelona in January 2010) "some properties of the language faculty are too deeply built in to be possible loci of variation."[2]

As I observed in Chapter 4, points of variation are confined to the margins of narrow syntax, especially the morphophonological component (PF).

(1) Locus of variation
All 'parameters' reduce to realizational options

Choosing among these options is rendered necessary by the need to externalize structures constructed by an underspecified syntactic component. We certainly

[2] "Limiting syntactic parameters to features of functional heads is also intended to exclude the possibility that there could be a syntactic parameter that is a feature of no element of the lexicon at all, e.g. there could presumably not be a parameter of the sort 'language L has or does not have bottom-to-top derivations.'" (Kayne (2005)).

do not yet have a complete typology of such rules, but works focusing on the articulation of the post-syntactic morphological component of the sort found in Distributed Morphology or Nanosyntax, and discussed in Chapter 3, are beginning to provide a good idea of what such realizational options amount to.

One interesting illustration of the point just made comes from Duguine's work on the nature of *pro*-drop (Duguine (2013)). Duguine shows that null elements that traditionally fall under the rubric of *pro*-drop, such as null subjects, do not owe their existence to a particular parameter setting. Rather, null elements are the null option, the default. They are just a special case of ellipsis. What needs to be accounted for is the impossibility of null elements (ellipsis). This, Duguine argues, is due to the failure to license ellipsis in the PF-component. Interestingly, Duguine concludes that DP-ellipsis of the sort giving rise to *pro*-drop is ruled by two formal conditions: the DICE and the ϕICE:

(2) D-Identity Condition on DP-Ellipsis (DICE)
A DP is eligible for ellipsis *iff* it stands in a D-Agree relation with a head

(3) Φ-Identity Constraint on DP-ellipsis (ϕICE)
A DP can be elided *iff* the values of its ϕ-features are identical to the values of the ϕ-features present on the head with which it has established a D-Agree relation

As Duguine points out, this state of affairs is reminiscent of Rizzi's original proposal concerning *pro*-drop (Rizzi (1986)), which also posited that *pro*-drop is ruled by two conditions, one based on Case – the Licensing Condition – and another based on the ϕ-features on the agreement morpheme – the Identification Condition. Why should this be? I'd like to argue that the fact that studies on *pro*-drop that are otherwise built from different premises wind up proposing two very similar conditions is due to the architecture of the PF-component. As we saw in Chapter 3, the PF-wing of the grammar is quite complex. Prior to phonology proper, it consists of two lexical insertion steps, L-match and L-insert. Since both steps are involved in lexical realization, it is no surprise to see DP-ellipsis licensing consisting of a matching condition (at the L-match stage) and a lexical selection condition (L-insert). In other words, the very fact that we find two conditions governing *pro*-drop argues for a PF-treatment of the latter, because these conditions correspond to the sort of architecture of the PF-component that has been independently argued for.

I could provide many other examples of this kind: the number of proposals reanalyzing traditional syntactic parameters in PF terms is indeed growing. Consider Safir's (2014) conclusion regarding variation in the domain of binding ("there is only one true anaphor in natural language which takes many [morphological] shapes") (see also Drummond (2011)), or Jenks's (2012) treatment of classifier- vs. non-classifier languages in terms of 'spanning' rules ("by approaching analytic languages in terms of spanning, a complex web of realizational spanning rules might be found lurking beneath their spartan morphology"). These studies point to the fact that although syntax 'projects' the lexicon (and not the other way around), phonology has a much larger impact on the final form of a linguistic utterance than is generally thought, because of the fact that a minimal syntax does not provide structures that are immediately ready for externalization.

Before concluding this appendix, I would like to address an interesting point made by Roberts (2010b, 2011), which I reproduce here from Boeckx and Leivada (2014). While recognizing the existence of realizational 'parameters,' Roberts thinks that it would be wrong to limit variation to the PF-component of the grammar, as he sees no non-stipulative way to exclude syntactic parameters in current minimalist models, hence, he claims, such syntactic options for variation should be exploited as well. I have already addressed part of his argument in Chapter 4, pointing out that once lexicocentrism is eliminated, there exist non-stipulative arguments to exclude syntactic variation. Here I'd like to focus on another part of Roberts's argument in favor of the existence of syntactic, non-PF-parameters. Roberts takes realizational parameters to be "dumb" and to be unable to give rise to parametric hierarchies.

The classical notion of 'parameter' was indeed meant to be used to make certain predictions with respect to the existence of specific parametric paths, guiding language acquisition. The problem for such a view is not logical (reducing the cost of acquisition would be a great feature), but as discussed in Boeckx and Leivada (2013), when a realistic number of parameters and parametric dependencies are taken into account, the deterministic nature of the paths defined by UG disappears, revealing a complex, subway-map-like network that cannot be assumed to guide the learner in any straightforward way. But aside from this, Roberts' argument that PF-'parameters' cannot give rise to (parametric) hierarchies is flawed because even for him there is nothing properly syntactic to construct the hierarchies. As Roberts himself has made explicit in a number of works, parametric hierarchies are "emergent properties" (Roberts (2011); Roberts and Holmberg (2009)). They are constructed

on the basis of markedness principles, crucially not syntactic principles. As Roberts (2011) states, markedness principles are "not grammatical principles but acquisition strategies" that derive from third factor considerations. If so, these principles could 'structure' variation (giving rise to hierarchies) in the PF-component.

Roberts provides another argument in favor of syntactic variation: for Roberts, 'PF parameters' are symmetrical and a parameter P is a non-PF one iff the realized variation defined by P contains a (typological) gap. Notice right away that this characterization is not derived by Roberts. That is, it's a stipulative way of defining PF and non-PF parameter. In other words, it's a stipulative way to rule in the existence of syntactic parameter. Why couldn't PF parameters give rise to typological gaps, given the existence of typological gaps in uncontroversially phonological phenomena?

Roberts provides an example based on word order restrictions to support his claim, but as discussed in Boeckx and Leivada (2014), empirically, it is not even clear that his example based on word order achieves what he wants it to do. Roberts notes that a seemingly symmetric parameter like head-final/head-initial gives rise to asymmetries such as the Final-over-Final Constraint (FOFC) that make the parameter syntactic. FOFC is defined as follows:

(4) A head-final phrase cannot immediately dominate a head-initial phrase in a single extended projection.
Impossible: $[_{XP}\ [_{YP}$ Y ZP] X]

Roberts calls FOFC the "signature asymmetry" that shows the word-order/linearization parameter to be a non-PF parameter. However, exploring the predictions of FOFC across different languages, one observes that, contra (4), head-final VPs may immediately dominate head-initial PPs in verb-second languages, as Biberauer *et al.* (2007) also acknowledge. This observation rendered the following modification of FOFC necessary:

(5) If a phase-head PH has an EPP feature, all the heads in its complement domain with which it agrees in categorial features must have an EPP feature.
Impossible: $[_{VP}\ [_{VP}$ V OBJ] V]

And yet, even this formulation is not immune to counterexamples either. More specifically, '2-3-1' orderings of verbal clusters (modal2-verb3-auxiliary1 sequences in Afrikaans, varieties of Swiss German, and West Flemish,

discussed in Biberauer (n.d.)) correspond to head-final, head-initial, and head-final patterns for VP, ModP, and AuxP respectively and they therefore give rise to FOFC violations. This is an important observation if FOFC is portrayed as the signature asymmetry that provides evidence in favor of syntactic parameters. If the allegedly robust asymmetries can be violated, it's not clear what conclusion one can draw from FOFC. In line with the proposal put forth in this book, FOFC constraints have been approached in PF terms. For example, Etxepare and Haddican (2013) suggest that Basque verb clusters favor a PF-based analysis of FOFC effects as opposed to a syntactic one. In sum, even "signature asymmetries" that at first look like they favor syntactic treatments may reinforce the claim that the sole locus of grammatical variation lies in the PF component.

References

Abels, K. (2003). Successive cyclicity, anti-locality, and adposition stranding. Ph.D. thesis, University of Connecticut.

Abels, K. and Bentzen, K. (2009). Are movement paths punctuated or uniform? *Catalan Journal of Linguistics*, **8**, 19–40.

Abney, S. P. (1987). The noun phrase in its sentential aspect. Ph.D. thesis, MIT.

Aboh, E. O. (2009). Clause structure and verb series. *Linguistic Inquiry*, **40**(1), 1–33.

Acedo-Matellán, V. (2010). Argument structure and the syntax–morphology interface. A case study in Latin and other languages. Ph.D. thesis, Universitat de Barcelona.

Acquaviva, P. and Panagiotidis, P. (2012). Lexical decomposition meets conceptual atomism. MS, University College Dublin and University of Cyprus.

Adger, D. (2003). *Core Syntax: A Minimalist Approach*. Oxford University Press.

(2010). A minimalist theory of feature structure. In *Features: Perspectives on a Key Notion in Linguistics* (ed. A. Kibort and G. Corbett), pp. 185–218. Oxford University Press.

(2013a). Constructions and grammatical explanation: Comments on Goldberg. *Mind & Language*, **28**, 466–478.

(2013b). *The Syntax of Substance*. Cambridge, MA: MIT Press.

Adger, D., Harbour, D., and Watkins, L. J. (2009). *Mirrors and Macroparameters: Phrase Structure beyond Free Word Order*. Cambridge University Press.

Adger, D. and Svenonius, P. (2011). Features in minimalist syntax. In *The Oxford Handbook of Linguistic Minimalism* (ed. C. Boeckx), pp. 27–51. Oxford University Press.

Amritavalli, R. and Jayaseelan, K. A. (2003). The genesis of syntactic categories and parametric variation. In *Generative Grammar in a Broader Perspective: Proceedings of the 4th GLOW in Asia*, pp. 19–41.

Anderson, S. R. (2004). *Doctor Dolittle's Delusion: Animals and the Uniqueness of Human Language*. New Haven, CT: Yale University Press.

Arad, M. (2005). *Roots and Patterns: Hebrew Morpho-Syntax*. Dordrecht: Kluwer.

Arbib, M. (n.d.). Neurolinguistics. MS, University of Southern California.

Arregi, K. and Nevins, A. (2012). *Morphotactics: Basque Auxiliaries and the Structure of Spellout*. Dordrecht: Springer.

(2013). Contextual neutralization and the elsewhere principle. In *Morphemes for Morris Halle* (ed. A. Marantz and O. Matushansky), pp. 199–222. Cambridge, MA: MIT Press.

Arsenijević, B. (2009). Clausal complementation as relativization. *Lingua*, **119**(1), 39–50.

(2013). The syntactic triangle: phases, categories and reference. MS, University of Niš.

Arsenijević, B. and Hinzen, W. (2010). Recursion as a human universal and as a primitive. *Biolinguistics*, **4**, 165–173.

(2012). On the absence of x-within-x recursion in human grammar. *Linguistic Inquiry*, **43**(3), 423–440.

Bach, E. and Horn, G. M. (1976). Remarks on "conditions on transformations." *Linguistic Inquiry*, **7**(2), 265–299.

Bak, P. (1996). *How Nature Works: The Science of Self-organized Criticality*. New York: Copernicus.

Bak, P., Tang, C., and Wiesenfeld, K. (1988). Self-organized criticality. *Physical Review A*, **38**(1), 364–374.

Baker, M. (1988). *Incorporation: A Theory of Grammatical Function Changing*. University of Chicago Press.

(1996). *The Polysynthesis Parameter*. Oxford University Press.

(1999). On the interplay of the universal and the particular: Case study of Edo. In *Proceedings of CLS 35: The Panels*, pp. 265–289. Chicago Linguistic Society.

(2001). *The Atoms of Language*. New York: Basic Books.

(2003). *Lexical Categories: Verbs, Nouns, and Adjectives*. Cambridge University Press.

(2005). The innate endowment for language: Underspecified or overspecified? In *The Innate Mind* (ed. P. Carruthers, S. Laurence, and S. Stich), pp. 156–174. Oxford University Press.

(2008a). The macroparameter in a microparametric world. In *The Limits of Variation* (ed. T. Biberauer), pp. 351–373. Amsterdam: John Benjamins.

(2008b). *The Syntax of Agreement and Concord*. Cambridge University Press.

(2011). Principles and parameters set out from Europe. Presented at 50 Years of Linguistics at MIT.

Baker, M. and Collins, C. (2006). Linkers and the internal structure of vP. *Natural Language & Linguistic Theory*, **24**(2), 307–354.

Baker, M. and Vinokurova, N. (2010). Two modalities of case assignment in Sakha. *Natural Language & Linguistic Theory*, **28**, 593–642.

Balari, S., Boeckx, C., and Lorenzo, G. (2012). On the feasibility of biolinguistics: Koster's word-based challenge and our 'natural computation' alternative. *Biolinguistics*, **6**(2), 205–221.

Balari, S. and Lorenzo, G. (2013). *Computational Phenotypes: Towards an Evolutionary Developmental Biolinguistics*. Oxford University Press.

Ball, P. (2006). *Critical Mass: How One Thing Leads to Another*. New York: Farrar Straus & Giroux.

Barner, D., Wood, J., Hauser, M., and Carey, S. (2008). Evidence for a non-linguistic distinction between singular and plural sets in rhesus monkeys. *Cognition*, **107**(2), 603–622.

Behar, D. M., Villems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E., Scozzari, R., Makkan, H., Tzur, S., Comas, D. *et al.* (2008). The dawn of human matrilineal diversity. *American Journal of Human Genetics*, **82**(5), 1130–1140.

Béjar, S. (2003). Phi-syntax: A theory of agreement. Ph.D. thesis, University of Toronto.

Belletti, A. (2004). Aspects of the low IP area. In *The Structure of CP and IP* (ed. L. Rizzi), pp. 16–51. Oxford University Press.

Benítez-Burraco, A. (2009). *Genes y lenguaje. Aspectos ontogenéticos, filogenéticos y cognitivos*. Reverté, Barcelona.

Benítez-Burraco, A. and Longa, V. (2010). Evo-Devo – of course, but which one? *Biolinguistics*, **4**, 308–323.

Berwick, R. C. (2011). All you need is merge: biology, computation and language from the bottom-up. In *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty* (ed. A. M. Di Sciullo and C. Boeckx), pp. 461–491. Oxford University Press.

Berwick, R., Beckers, G., Okanoya, K., and Bolhuis, J. (2012). A bird's eye view of human language evolution. *Frontiers in Evolutionary Neuroscience*, **4**(5).

Berwick, R. C. and Chomsky, N. (2011). The biolinguistic program: The current state of its development. In *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty* (ed. A. M. Di Sciullo and C. Boeckx), pp. 19–41. Oxford University Press.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In *Cognition and the Development of Language* (ed. J. R. Hayes), pp. 279–362. New York: Wiley.

Bhatt, R. and Walkow, M. (2013). Locating agreement in grammar: An argument from agreement in conjunctions. *Natural Language and Linguistic Theory*, **31**(4), 951–1013.

Biberauer, T. (ed.) (2008). *The Limits of Syntactic Variation*. Amsterdam: John Benjamins.

(n.d). Disharmonic word order, quirky morphology and the Afrikaans verb cluster. MS, University of Cambridge.

Biberauer, T., Holmberg, A., and Roberts, I. G. (2007). Structure and linearization in disharmonic word orders. Paper presented at the 17th Colloquium of Generative Grammar, Girona.

Biberauer, T. and Richards, M. (2006). True optionality: When the grammar doesn't mind. In *Minimalist Essays* (ed. C. Boeckx), pp. 35–67. Amsterdam: John Benjamins.

Blevins, J. (2004). *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge University Press.

Bobaljik, J. D. (2008). Where's Phi? Agreement as a post-syntactic operation. In *Phi Theory: Phi-Features across Modules and Interfaces* (ed. D. Harbour, D. Adger, and S. Béjar), pp. 295–328. Oxford University Press.

Bobaljik, J. D. and Thráinsson, H. (1998). Two heads aren't always better than one. *Syntax*, **1**, 37–71.

Boeckx, C. (2003). *Islands and Chains*. Amsterdam: John Benjamins.

(2005). Generative grammar and modern cognitive science. *Journal of Cognitive Science*, **6**, 45–54.

(2006). *Linguistic Minimalism: Origins, Concepts, Methods, and Aims*. Oxford University Press.

(2008a). *Aspects of the Syntax of Agreement*. London: Routledge.

(2008b). *Bare Syntax*. Oxford University Press.
(2008c). The person case constraint and patterns of exclusivity. In *Agreement Restrictions* (ed. R. D'Alessandro, S. Fischer, and G. Hrafnbjargarson), pp. 87–101. Berlin: Mouton de Gruyter.
(2008d). *Understanding Minimalist Syntax: Lessons from Locality in Long-distance Dependencies*. Oxford: Blackwell.
(2009a). Cartesian biolinguistics. SOLIFIC lecture, Sophia University, July 2009. MS, ICREA – Universitat Autònoma de Barcelona.
(2009b). *Language in Cognition: Uncovering Mental Structures and the Rules behind Them*. Malden, MA: Wiley-Blackwell.
(2009c). The locus of asymmetry in UG. *Catalan Journal of Linguistics*, **8**, 41–53.
(2009d). On long-distance agree. *Iberia*, **1**, 1–32.
(2009e). On the nature of merge. In *Of Minds and Language: A Basque Encounter with Noam Chomsky* (ed. M. Piattelli-Palmarini, P. Salaburu, and J. Uriagereka), pp. 44–57. Oxford University Press.
(2009f). Some notes on the syntax–thought interface. In *Proceedings of the Sophia University Linguistic Society 24*, pp. 92–103. Sophia University Linguistic Society.
(2010a). Linguistic minimalism. In *Oxford Handbook of Linguistic Analysis* (ed. B. Heine and H. Narrog), pp. 485–505. Oxford University Press.
(2010b). Syntactic order for free. Presented at the 10th European Conference on Complex Systems, Lisbon University Institute, workshop "Modelling the non-separability of a very complex world."
(2011a). Approaching parameters from below. In *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty* (ed. A.-M. D. Sciullo and C. Boeckx), pp. 205–221. Oxford University Press.
(2011b). The emergence of the language faculty, from a biolinguistic point of view. In *The Oxford Handbook of Language Evolution* (ed. M. Tallerman and K. Gibson), pp. 492–501. Oxford University Press.
(2011c). Some reflections on Darwin's problem in the context of Cartesian biolinguistics. In *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty* (ed. A.-M. Di Sciullo and C. Boeckx), pp. 42–64. Oxford University Press.
(2012a). Phases beyond explanatory adequacy. In *Phase Theory: Developing the Framework* (ed. A. Gallego), pp. 45–66. Berlin: Mouton de Gruyter.
(2012b). *Syntactic Islands*. Cambridge University Press.
(2013a). Biolinguistics: Facts, fiction, forecast. *Biolinguistics*, **7**, 316–328.
(2013b). Biolinguistics: Forays into human cognitive biology. *Journal of Anthropological Sciences*, **91**, 63–89.
(2013c). Merge: Biolinguistic considerations. *English Linguistics*, **30**(2), 463–483.
(2014a). The roots of current biolinguistic thought: Revisiting the "Chomsky–Piaget debate" in the context of the revival of biolinguistics. *Teorema*, **33**, 83–94.
(2014b). What can an extended synthesis do for biolinguistics? On the needs and benefits of eco-evo-devo program. In *The Evolution of Communication in Primates:*

A Multidisciplinary Approach (ed. N. Gontier and M. Pina), pp. 313–326. Dordrecht: Springer.

(2014c). What Principles & Parameters got wrong. In *Linguistic Variation and the Minimalist Program* (ed. C. Picallo). Oxford University Press, Oxford.

(forthcoming). Considerations pertaining to the nature of logodiversity, or how to construct a parametric space without parameters. In *Rethinking Parameters* (ed. Luis Eguren *et al.*). Oxford University Press.

Boeckx, C. and Benítez-Burraco, A. (2014). The shape of the language-ready brain. *Frontiers in Psychology*, **5**, 282.

Boeckx, C. and Grohmann, K. K. (2007). Putting phases in perspective. *Syntax*, **10**, 204–222.

Boeckx, C. and Hornstein, N. (2008). Superiority, reconstruction and islands. In *Foundational Issues in Linguistics* (ed. C. Otero, R. Freidin, and M.-L. Zubizarreta), pp. 197–225. Cambridge, MA: MIT Press.

Boeckx, C., Hornstein, N., and Nunes, J. (2010). *Control as Movement*. Cambridge University Press.

Boeckx, C. and Leivada, E. (2013). Entangled parametric hierarchies: Problems for an overspecified universal grammar. *PLoS One*, **8**(9), e72357.

(2014). On the particulars of universal grammar: Implications for acquisition. *Language Sciences* [in press], ICREA & Universitat de Barcelona.

Boeckx, C. and Martín, T. (2013). *El clitic datiu es mes que un clitic*. Pages editors, Lleida.

Boeckx, C. and Martínez-Álvarez, A. (2013). A multi-step algorithm for serial order: Converging evidence from linguistics and neuroscience. Presented at GLOW 36, Lund.

Boeckx, C., Martins, P. T., and Leivada, E. (in press). Biolinguistics. In *The Cambridge Handbook of Syntactic Change* (ed. I. Roberts and A. Ledgeway). Cambridge University Press.

Boeckx, C. and Piattelli-Palmarini, M. (2005). Language as a natural object; linguistics as a natural science. *Linguistic Review*, **22**(2–4), 467–471.

(2007). Linguistics in cognitive science: state of the art amended. *Linguistic Review*, **24**(4), 403–415.

Boeckx, C. and Samuels, B. (2009). What emerges from merge in phonology. Presented at the 6th Old World Conference on Phonology, Edinburgh, UK.

Boeckx, C. and Uriagereka, J. (2007). Minimalism. In *The Oxford Handbook of Linguistic Interfaces* (ed. G. Ramchand and C. Reiss), pp. 541–573. Oxford University Press.

Borer, H. (1984). *Parametric Syntax*. Dordrecht: Foris.

(2003). Exo-skeletal vs. endo-skeletal explanations: Syntactic projections and the lexicon. In *The Nature of Explanation* (ed. J. Moore and M. Polinsky), pp. 31–67. Chicago: CSLI Publications.

(2005). *Structuring Sense (2 vols.)*. Oxford University Press.

Bošković, Ž. (1994). D-structure, theta criterion, and movement into theta positions. *Linguistic Analysis*, **24**, 247–286.

(2002). A-movement and the EPP. *Syntax*, **5**, 167–218.

(2005). On the locality of left branch extraction and the structure of NP. *Studia Linguistica*, **59**(1), 1–45.
(2007). On the locality and motivation of move and agree: An even more minimal theory. *Linguistic Inquiry*, **38**(4), 589–644.
(2008a). The NP/DP analysis and Slovenian. In *Novi Sad Generative Syntax Workshop*, pp. 53–73. University of Novi Sad.
(2008b). What will you have, DP or NP? In *Proceedings of NELS 37*. GLSA.
(2010). NPs and clauses. MS, University of Connecticut.
(2011). On valued uninterpretable features. In *Proceedings of NELS 39*, pp. 109–120. GLSA Publications: University of Massachusetts.
Bošković, Ž. and Lasnik, H. (2007). *Minimalist Syntax: The Essential Readings*. Malden, MA: Wiley-Blackwell.
Brame, M.K. (1976). *Conjectures and Refutations in Syntax and Semantics*. London: North Holland.
Briscoe, E. and Feldman, J. (2006). Conceptual complexity and the bias–variance tradeoff. MS, Rutgers University.
Brody, M. (2000). Mirror theory: Syntactic representation in perfect syntax. *Linguistic Inquiry*, **29**, 367–398.
(2003). *Towards an Elegant Syntax*. New York: Routledge.
Browning, M. (1987). Null operator constructions. Ph.D. thesis, MIT.
Bruening, B. (2008). Selectional asymmetries between CP and DP suggest that the DP hypothesis is wrong. MS, University of Delaware.
Bufill, E. and Carbonell, E. (2004). Are symbolic behaviour and neuroplasticity an example of gene–culture coevolution? *Revista de neurologia*, **39**(1), 48–55.
Butler, J. (2004). Phase structure, phrase structure, and quantification. Ph.D. thesis, University of York.
Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge University Press.
Bye, P. and Svenonius, P. (2012). Non-concatenative morphology as epiphenomenon. In *The Morphology and Phonology of Exponence* (ed. J. Trommer), pp. 427–495. Oxford University Press.
Caha, P. (2009). The nanosyntax of case. Ph.D. thesis, Universitetet i Tromsø.
Caponigro, I. and Polinsky, M. (2011). Relative embeddings: A Circassian puzzle for the syntax/semantics interface. *Natural Language and Linguistic Theory*, **29**, 71–122.
(forthcoming). Almost everything is relative in the Caucasus. In *Proceedings of SALT 18*.
Carnie, A. (2008). *Constituent Structure*. Oxford University Press.
Carruthers, P. (2006). *The Architecture of the Mind*. Oxford University Press.
Carstens, V. (2005). Agree and EPP in Bantu. *Natural Language and Linguistic Theory*, **23**(2), 219–279.
Castañeda, H.-N. (1967). Actions, imperatives, and obligations. In *Proceedings of the Aristotelian Society*, **68**, pp. 25–48.
Cheney, D. L. and Seyfarth, R. M. (1990). *How Monkeys See the World*. University of Chicago Press.
(2007). *Baboon Metaphysics: The Evolution of a Social Mind*. University of Chicago Press.

Cheng, L. L. S. and Sybesma, R. (1999). Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry*, **30**(4), 509–542.

(2005). Classifiers in four varieties of Chinese. In *The Oxford Handbook of Comparative Syntax* (ed. G. Cinque and R. Kayne), pp. 259–292. Oxford University Press.

Chierchia, G. (1998). Reference to kinds across languages. *Natural Language Semantics*, **6**, 339–405.

Chomsky, N. (1955). The logical structure of linguistic theory. MS, Harvard University, Cambridge, MA. [published in part in 1975. New York: Plenum]

(1957). *Syntactic Structures*. The Hague: Mouton.

(1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

(1966). *Cartesian Linguistics*. New York: Harper & Row. [third edition, with introduction by J. McGilvray, 2009, Cambridge University Press].

(1970). Remarks on nominalization. In *Readings in English Transformational Grammar* (ed. R. Jacobs and P. Rosenbaum), pp. 184–221. Waltham, MA: Ginn.

(1972). *Language and Mind*. New York: Harcourt Brace Jovanovich.

(1973). Conditions on transformations. In *A Festschrift for Morris Halle* (ed. S. Anderson and P. Kiparsky), pp. 232–286. New York: Holt, Rinehart and Winston.

(1977a). *Essays on Form and Interpretation*. Amsterdam: Elsevier.

(1977b). On wh-movement. In *Formal Syntax* (ed. P. Culicover, T. Wasow, and A. Akmajian), pp. 71–132. New York: Academic Press.

(1980). *Rules and Representations*. New York: Columbia University Press.

(1981). *Lectures on Government and Binding*. Dordrecht: Foris.

(1986). *Knowledge of Language*. New York: Praeger.

(1988). *Language and Problems of Knowledge: The Managua Lectures*. Cambridge, MA: MIT Press.

(1993). A minimalist program for linguistic theory. In *The View from Building 20* (ed. K. Hale and S. J. Keyser), pp. 1–52. Cambridge, MA: MIT Press.

(1994). Bare phrase structure. *MIT Occasional Papers in Linguistics*, **5**. Reprinted in G. Webelhuth, ed. (1995), *Government and Binding Theory and the Minimalist Program*, 383–439. Oxford: Blackwell.

(1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

(2000a). Minimalist inquiries: The framework. In *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik* (ed. R. Martin, D. Michaels, and J. Uriagereka), pp. 89–155. Cambridge, MA: MIT Press.

(2000b). *New Horizons in the Study of Language and Mind*. Cambridge University Press.

(2001). Derivation by phase. In *Ken Hale: A Life in Language* (ed. M. Kenstowicz), pp. 1–52. Cambridge, MA: MIT Press.

(2004). Beyond explanatory adequacy. In *Structures and Beyond* (ed. A. Belletti), pp. 104–131. New York: Oxford University Press.

(2005). Three factors in the language design. *Linguistic Inquiry*, **36**, 1–22.

(2007). Approaching UG from below. In *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Semantics* (ed. U. Sauerland and H.-M. Gärtner), pp. 1–30. Berlin: Mouton de Gruyter.

(2008). On phases. In *Foundational Issues in Linguistics* (ed. C. Otero, R. Freidin, and M.-L. Zubizarreta), pp. 133–166. Cambridge, MA: MIT Press.
(2012a). Foreword. In *Phases: Developing the Framework* (ed. Á. Gallego), pp. 1–7. Berlin: Mouton de Gruyter.
(2012b). *The Science of Language: Interviews with James McGilvray*. Cambridge University Press.
(2013). Problems of projection. *Lingua*, **130**, 33–49.
Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English*. Harper & Row, New York. Reprinted in 1991. Cambridge, MA: MIT Press.
Chomsky, N., Halle, M., and Lukoff, F. (1956). On accent and juncture in English. In *For Roman Jakobson: Essays on the Occasion of His Sixtieth Birthday*, pp. 65–80. The Hague: Mouton.
Chomsky, N. and Lasnik, H. (1977). Filters and Control. *Linguistic Inquiry*, **8**(3), 425–504.
Chung, S. (2012). On reaching agreement late. In *Proceedings of CLS 48*.
Cinque, G. (1999). *Adverbs and Functional Heads*. Oxford University Press.
(2002). "We had to meet in cafés to read *Syntactic Structures*." *Glot International*, **6**, 190–193.
(2013). *Typological Studies. Word Order and Relative Clauses*. New York: Routledge.
Cinque, G. and Rizzi, L. (2010). The cartography of syntactic structures. In *Oxford Handbook of Linguistic Analysis* (ed. B. Heine and H. Narrog), pp. 51–65. Oxford University Press.
Collins, C. (2002). Eliminating labels. In *Derivation and Explanation in the Minimalist Program* (ed. S. D. Epstein and T. D. Seely), pp. 42–64. Oxford: Blackwell.
(2005). A smuggling approach to the passive in English. *Syntax*, **8**(2), 81–120.
Collins, C. and Ura, H. (2001). Eliminating phrase structure. MS, Cornell University and Kwansei Gakuin University.
Cooke, J. and Zeeman, E. C. (1976). A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. *Journal of Theoretical Biology*, **58**(2), 455–476.
Cottingham, J. (1978). 'A brute to the brutes?': Descartes' treatment of animals. *Philosophy*, **53**(206), 551–559.
Cuervo, M. C. (2003). Datives at large. Ph.D. thesis, MIT.
Culbertson, J. (2010). Learning biases, regularization, and the emergence of typological universals in syntax. Ph.D. thesis, Johns Hopkins University.
Culicover, P. (1999). *Syntactic Nuts: Hard Cases, Syntactic Theory, and Language Acquisition*. Oxford University Press.
Culicover, P. W. and Jackendoff, R. (2005). *Simpler Syntax*. Oxford University Press.
Davidson, D. (1967). The logical form of action sentences. In *Essays on Actions and Events* (ed. D. Davidson), pp. 105–148. Oxford: Clarendon Press.
Dawkins, R. (1996). *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe Without Design*. New York: Norton.
(2006). *The God Delusion*. London: Black Swan.
De Belder, M. and van Craenenbroeck, J. (2011). How to merge a root. MS, Brussels and Utrecht.

(2013). On vocabulary insertion. MS, HU Brussels.

De Vries, M. (2009). On multidominance and linearization. *Biolinguistics*, **3**(4), 344–403.

Deacon, T. W. (2006). Emergence: The hole at the wheel's hub. In *The Re-emergence of Emergence: The Emergentist Hypothesis from Science to Religion* (ed. P. Clayton and P. Davies), pp. 111–150. Oxford University Press.

(2010). A role for relaxed selection in the evolution of the language capacity. *Proceedings of the National Academy of Sciences*, **107**(Supplement 2), 9000–9006.

Dediu, D. and Levinson, S. C. (2013). On the antiquity of language: The reinterpretation of neandertal linguistic capacities and its consequences. *Frontiers in Psychology*, **4**, 397.

Dehaene, S. (1997). *The Number Sense*. Oxford University Press.

Demirdache, H. and Uribe-Etxebarria, M. (2000). The primitives of temporal relations. In *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik* (ed. R. Martin, D. Michaels, and J. Uriagereka), pp. 157–186. Cambridge, MA: MIT Press.

Di Sciullo, A. M. and Boeckx, C. (2011). Contours of the biolinguistic research agenda. In *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty* (ed. A.-M. D. Sciullo and C. Boeckx), pp. 1–16. Oxford University Press.

Diamond, J. M. (1992). *The Third Chimpanzee: The Evolution and Future of the Human Animal*. New York: HarperCollins.

Diercks, M. (2011). Parameterizing case: Evidence from Bantu. MS, Pomona College.

Dillon, B. and Idsardi, W. (2009). Investigating statistical approaches to building a phonology. MS, University of Maryland.

Dowty, D. (1989). On the semantic content of the notion of "thematic role." In *Properties, Types and Meaning* (ed. G. Chierchia, B. H. Partee, and R. Turner), pp. 69–129. Dordrecht: Kluwer.

Drummond, A. (2011). Binding phenomena within a reductionist theory of grammatical dependencies. Ph.D. thesis, University of Maryland.

Duguine, M. (2013). Null arguments and linguistic variation: A minimalist analysis of *pro*-drop. Ph.D. thesis, Euskal Herriko Unibertsitatea and Université de Nantes.

Dupuy, J.-P. (2009). *On the Origins of Cognitive Science: The Mechanization of the Mind* (2nd edn). Cambridge, MA: MIT Press.

Embick, D. (2010). *Localism versus Globalism in Morphology and Phonology*. Cambridge, MA: MIT Press.

Embick, D. and Noyer, R. (2007). Distributed morphology and the syntax/morphology interface. In *The Oxford Handbook of Linguistic Interfaces* (ed. G. Ramchand and C. Reiss), pp. 289–324. Oxford University Press.

Emonds, J. E. (1985). *A Unified Theory of Syntactic Categories*. Dordrecht: Foris.

(2008). Valuing v-features and n-features: What adjuncts tell us about case, agreement, and syntax in general. In *Merging Features: Computation, Interpretation, and Acquisition* (ed. J. Brucart, A. Gavarró, and J. Solà), pp. 194–214. Oxford University Press.

Endress, A. D., Cahill, D., Block, S., Watumull, J., and Hauser, M. D. (2009). Evidence of an evolutionary precursor to human language affixation in a non-human primate. *Biology Letters*, **5**(6), 749–751.

Endress, A. D., Carden, S., Versace, E., and Hauser, M. D. (2010). The apes' edge: Positional learning in chimpanzees and humans. *Animal Cognition*, **13**(3), 483–495.

Endress, A. D., Dehaene-Lambertz, G., and Mehler, J. (2007). Perceptual constraints and the learnability of simple grammars. *Cognition*, **105**(3), 577–614.

Endress, A. D. and Mehler, J. (2009). Primitive computations in speech processing. *Quarterly Journal of Experimental Psychology*, **62**(11), 2187–2209.

Endress, A. D., Nespor, M., and Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences*, **13**(8), 348–353.

Epstein, S. D. (2003 [2007]). On i(nternalist)–functional explanation in minimalism. *Linguistic Analysis*, **33**(1–2), 20–53.

Epstein, S. D., Kitahara, H., and Seely, T. D. (2012). Structure Building that can't be. In *Ways of Structure Building* (ed. M. Uribe-Etxebarria and V. Valmala), pp. 253–270. Oxford University Press.

Epstein, S. D. and Seely, T. D. (2002). Rule applications as cycles in a level-free syntax. In *Derivation and Explanation in the Minimalist Program* (ed. S. D. Epstein and T. D. Seely), pp. 65–89. Oxford: Blackwell.

(2006). *Derivations in Minimalism*. Cambridge University Press.

Etxepare, R. and Haddican, B. (2013). Repairing final-over-final constraint violations: Evidence from Basque verb clusters. Paper presented at GLOW 36.

Evans, N. and Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, **32**(5), 429–492.

Fasanella-Seligrat, A. (2011). Els problemes de Plató, de Darwin i de Greenberg. *Els Marges*, **94**, 18–36.

Fasanella-Seligrat, A. and Fortuny, J. (2011). Deriving linguistic variation from learnability conditions in a parametric approach to UG. Presented at the workshop Formal Grammar and Syntactic Variation, Universidad Autónoma de Madrid.

Feldman, J. *et al.* (2000). Minimization of Boolean complexity in human concept learning. *Nature*, **407**(6804), 630–632.

Felix, S. (2010). Me and Chomsky. Remarks from someone who quit. In *Language and Logos* (ed. T. Hanneforth and G. Fanselow), pp. 64–71. Berlin: Akademie Verlag.

Ferrer i Cancho, R. and Forns, N. (2009). The self-organization of genomes. *Complexity*, **15**, 34–36.

Fitch, W. T. (2009). Prolegomena to a future science of biolinguistics. *Biolinguistics*, **3**(4), 283–320.

Fodor, J. A. (1998). *In Critical Condition: Polemical Essays on Cognitive Science and the Philosophy of Mind*. Cambridge, MA: MIT Press.

Fodor, J. D. (1998). Unambiguous triggers. *Linguistic Inquiry*, **29**(1), 1–36.

Fortuny, J. (2008). *The Emergence of Order in Syntax*. John Benjamins, Amsterdam.

Fox, D. (2000). *Economy and Semantic Interpretation*. Cambridge, MA: MIT Press.

(2002). Antecedent-contained deletion and the copy theory of movement. *Linguistic Inquiry*, **33**(1), 63–96.

Frampton, J. and Gutmann, S. (2002). Crash-proof syntax. In *Derivation and Explanation in the Minimalist Program* (ed. S. D. Epstein and T. D. Seely), pp. 90–105. Oxford: Blackwell.

Freidin, R. (1999). Cyclicity and minimalism. In *Working Minimalism* (ed. S. D. Epstein and N. Hornstein), pp. 95–126. Cambridge, MA: MIT Press.

Fujita, K. (2009). A prospect for evolutionary adequacy: Merge and the evolution and development of human language. *Biolinguistics*, **3**(2), 128–153.

Fukui, N. (2006). *Theoretical Comparative Syntax*. London: Routledge.

Fukui, N. and Speas, M. (1986). Specifiers and projections. *MIT Working Papers in Linguistics*, **8**, 128–172.

Fukui, N and Zushi, M. (2008). On certain differences between noun phrases and clauses. In *Essays on Nominal Determination* (ed. H. H. Müller and A. Klinge), pp. 265–286. Amsterdam: John Benjamins.

Gallego, Á. (2009). A note on C. MS, Universitat Autònoma de Barcelona.

(2011a). Lexical items and feature bundling. Presented at the workshop on Formal Grammar and Syntactic Variation, Universidad Autónoma de Madrid.

(2011b). *Phase Theory*. Amsterdam: John Benjamins.

(ed.) (2012). *Phases: Developing the Framework*. Berlin: Mouton de Gruyter.

Gallego, Á. and Bosque, I. (2011). A subextraction puzzle and its consequences for cyclicity. MS, UCM & UAB.

Georgi, D. and Müller, G. (2010). Noun phrase structure by reprojection. *Syntax*, **13**(1–36).

Gervain, J. and Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology*, **61**, 191–218.

Gierer, A. and Meinhardt, H. (1972). A theory of biological pattern formation. *Biological Cybernetics*, **12**(1), 30–39.

Giorgi, A. and Pianesi, F. (1997). *Tense and Aspect: From Semantics to Morphosyntax*. Oxford University Press.

Gould, S. J. (1997). Darwinian fundamentalism. *New York Review of Books*, **44**, 34–37.

Grimshaw, J. (1991). Extended projection. MS, Brandeis University. [reprinted in 2005, *In Words and Structure. CSLI Lecture Notes Number 151*. Stanford, CA: CSLI Publications]

Grohmann, K. K. (2003). *Prolific Domains. On the Anti-Locality of Movement Dependencies*. Amsterdam: John Benjamins.

(2011). Anti-locality. In *Oxford Handbook of Linguistic Minimalism* (ed. C. Boeckx), pp. 260–290. Oxford University Press.

Guardiano, C. and Longobardi, G. (2005). Parametric comparison and language taxonomy. In *Grammaticalization and Parametric Variation* (ed. M. Batllori, M.-L. Hernanz, C. Picallo, and F. Roca), pp. 149–174. Oxford University Press.

Haegeman, L. (1994). *Introduction to Government and Binding Theory*. Malden, MA: Blackwell.

Hale, K. and Keyser, S. J. (1993). On argument structure and the lexical expression of grammatical relations. In *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger* (ed. K. Hale and S. J. Keyser), pp. 53–110. Cambridge, MA: MIT Press.

(2002). *Prolegomenon to a Theory of Argument Structure*. Cambridge, MA: MIT Press.

Halle, M. and Marantz, A. (1993). Distributed morphology and the pieces of inflection. In *The View from Building 20* (ed. K. Hale and S. J. Keyser), pp. 111–176. Cambridge, MA: MIT Press.

(1994). Some key features of distributed morphology. *MIT Working Papers in Linguistics*, **21**, 275–288.

Harbour, D. (2006). The elimination of geometry. MS, Queen Mary University of London.

(2009). The semantics, and generality, of features: Or, how not to construct a theory of cognitive evolution. MS, Queen Mary University of London.

Harley, H. (2013). External arguments and the mirror principle: On the independence of voice and v. *Lingua*, **125**, 34–57.

(forthcoming). On the identity of roots. *Theoretical Linguistics*.

Harley, H. and Noyer, R. (1999). Distributed morphology. *Glot International*, **4**(4), 3–9.

Harley, H. and Ritter, E. (2002a). Person and number in pronouns: A feature-geometric analysis. *Language*, 482–526.

(2002b). Structuring the bundle: A universal morphosyntactic feature geometry. In *Pronouns – Grammar and Representation* (ed. H. Weise and H. Simon), pp. 23–39. Amsterdam: John Benjamins.

Haspelmath, M. (2008). Parametric versus functional explanations of syntactic universals. In *The Limits of Syntactic Variation* (ed. T. Biberauer), pp. 75–107. John Benjamins, Amsterdam.

(2010). Comparative concepts and descriptive categories in cross-linguistic studies. *Language*, **86**(3), 663–687.

Hauser, M. D. (2001). *Wild Minds: What Animals Really Think*. New York: Owl Books.

(2009). The possibility of impossible cultures. *Nature*, **460**(7252), 190–196.

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, **298**, 1569–1579.

Heine, B. and Kuteva, T. (2007). *The Genesis of Grammar: A Reconstruction*. Oxford University Press.

Heine, B. and Narrog, H. (eds.) (2010). *Oxford Handbook of Linguistic Analysis*. Oxford University Press.

Herburger, E. (2000). *What Counts: Focus and Quantification*. Cambridge, MA: MIT Press.

Hinzen, W. (2006). *Mind Design and Minimal Syntax*. Oxford University Press.

(2007). *An Essay on Naming and Truth*. Oxford University Press.

(2008). Prospects for an explanatory theory of semantics. *Biolinguistics*, **2**(4), 348–363.

(2011a). Emergence of a systemic semantics through minimal and underspecified codes. In *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty* (ed. A. M. Di Sciullo and C. Boeckx), pp. 417–439. Oxford University Press.

(2011b). Language and thought. In *The Oxford Handbook of Linguistic Minimalism* (ed. C. Boeckx), pp. 499–522. Oxford University Press.

(2012). Phases and semantics. In *Phases: Developing the Framework* (ed. Á. Gallego), pp. 309–342. Berlin: Mouton de Gruyter.
(2013). Narrow syntax and the language of thought. *Philosophical Psychology*, **26**(1), 1–23.
Hinzen, W. and Poeppel, D. (2011). Semantics between cognitive neuroscience and linguistic theory: Guest editors' introduction. *Language and Cognitive Processes*, **26**(9), 1297–1316.
Hinzen, W. and Reichard, U. (2011). The grammaticalization of thought. MS, University of Durham.
Hinzen, W. and Sheehan, M. (2013). *The Philosophy of Universal Grammar*. Oxford University Press.
Hiramatsu, K. (2000). Accessing linguistic competence: Evidence from children's and adults' acceptability judgments. Ph.D. thesis, University of Connecticut.
Hodges, A. (1983). *The Enigma*. New York: Simon & Schuster.
Holmberg, A. (2010). Parameters in minimalist theory: The case of Scandinavian. *Theoretical Linguistics*, **36**(1), 1–48.
Holmberg, A. and Platzack, C. (1995). *The Role of Inflection in Scandinavian Syntax*. Oxford University Press.
Hornstein, N. (2001). *Move! A Minimalist Theory of Construal*. Oxford: Blackwell.
(2009). *A Theory of Syntax*. Cambridge University Press.
Hornstein, N., Nunes, J., and Grohmann, K. K. (2006). *Understanding Minimalism*. Cambridge University Press.
Hornstein, N. and Pietroski, P. M. (2009). Basic operations: Minimal syntax–semantics. *Catalan Journal of Linguistics*, **8**, 113–139.
Hornstein, N. and Uriagereka, J. (2002). Reprojections. In *Derivation and Explanation in the Minimalist Program* (ed. S. D. Epstein and T. D. Seely), pp. 106–132. Oxford: Blackwell.
Huang, C.-T. J. (1982). Logical Relations in Chinese and the theory of grammar. Ph.D. thesis, MIT.
(2005). Analyticity. Class lectures, MIT/Harvard LSA Summer Institute.
Hurford, J. R. (2007). *The Origins of Meaning*. Oxford University Press.
Idsardi, W. and Raimy, E. (2013). Three types of Linearization and the temporal aspects of speech. In *Principles of linearization* (ed. T. Biberauer and I. Roberts), pp. 31–56. Berlin: Mouton de Gruyter.
Irurtzun, A. (2007). The grammar of focus at the interfaces. Ph.D. thesis, Euskal Herriko Unibertsitatea.
Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.
(1997). *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.
(2002). *Foundations of Language*. Oxford University Press.
(2005). Alternative minimalist visions of language. In *Proceedings from the 41st Annual Meeting of the Chicago Linguistic Society*, **2**, pp. 189–226.
(2010). *Meaning and the Lexicon: The Parallel Architecture 1975–2010*. Oxford University Press.
(2011). What is the human language faculty? Two views. *Language*, **87**(3), 586–624.

Jenks, P. (2010). Evidence for the syntactic diversity of numeral classifiers. MS, Harvard University.

(2012). Definite spans and blocking in classifier languages. MS, UC Berkeley.

Jeong, Y. (2007). *Applicatives: Structure and Interpretation from a Minimalist Perspective*. Amsterdam: John Benjamins.

(2011). Re-examining the "NP/DP" parameter in light of the diversity of East-Asian classifiers. In *Proceedings of the 12th Annual Tokyo Conference on Psycholinguistics* (ed. Y. Otsu), Tokyo, pp. 113–131.

Jiang, L. J. (2012). Nominal arguments and language variation. Ph.D. thesis, Harvard University.

Johnson, K. (2010). A remerge theory of movement. MS, University of Massachusetts at Amherst.

(2012). Towards deriving differences in how *wh*-movement and QR are pronounced. *Lingua*, **122**(6), 529–553.

Jurka, J. (2010). The importance of being a complement: CED effects revisited. Ph.D. thesis, University of Maryland.

Kagawa, H., Yamada, H., Lin, R.-S., Mizuta, T., Hasegawa, T., and Okanoya, K. (2012). Ecological correlates of song complexity in white-rumped Munias: The implication of relaxation of selection as a cause for signal variation in birdsong. *Interaction Studies*, **13**(2), 263–284.

Kato, Takaomi (2007). Symmetries in coordination. Ph.D. dissertation. Harvard University.

Katz, J. and Pesetsky, D. (2009). The identity thesis for language and music. MS, MIT.

Kayne, R. S. (1984). *Connectedness and Binary Branching*. Dordrecht: Foris.

(1994). *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.

(1997). The English complementizer *of*. *Journal of Comparative Germanic Linguistics*, **1**(1), 43–54.

(2005). *Movement and Silence*. Oxford University Press.

(2010). Why isn't *this* a complementizer? In *Principles and Contrasts* (ed. R. Kayne), pp. 190–227. Oxford University Press.

(2011). Antisymmetry and the lexicon. In *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty* (ed. A.-M. di Sciullo and C. Boeckx), pp. 329–353. Oxford University Press.

Kemp, C., Perfors, A., and Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, **10**(3), 307–321.

Kibort, A. and Corbett, G. (ed.) (2010). *Features. Perspectives on a Key Notion in Linguistics*. Oxford University Press.

Kinsella, A. R. (2009). *Language Evolution and Syntactic Theory*. Cambridge University Press.

Koopman, H. (2005). On the parallelism of DPs and clauses: Evidence from Kisongo Maasai. In *Verb First: On the Syntax of Verb-Initial Languages* (ed. A. Carnie, S. Dooley, and H. Harley), pp. 281–301. Amsterdam: John Benjamins.

Koster, J. (1987). *Domains and Dynasties: The Radical Autonomy of Syntax*. Dordrecht: Foris.

(2009). Ceaseless, unpredictable creativity: Language as technology. *Biolinguistics*, **3**(1), 61–92.

(2010). Language and tools. MS, Universiteit Groningen.
Krifka, M. (1992). Thematic relations as links between nominal reference and temporal constitution. In *Lexical Matters* (ed. I. Sag and A. Szabolcsi), pp. 29–53. Stanford, CA: CSLI Publications.
Kural, M. (2005). Tree traversal and word order. *Linguistic Inquiry*, **36**(3), 367–387.
Lahne, A. (2008). Where there is fire there is smoke. Local modelling of successive cyclic movement. Ph.D. thesis, Universität Leipzig.
Lakoff, G. (1970). *Irregularity in Syntax*. New York: Holt, Rinehart and Winston.
Landau, I. (2013). Agreement at PF: An argument from partial control. MS, Ben Gurion University.
Langacker, R. W. (1987). Nouns and verbs. *Language*, **63**, 53–94.
(1999). *Grammar and Conceptualization*. Volume XIV. Berlin: Mouton de Gruyter.
Larson, R. K. (1987). "Missing prepositions" and the analysis of English free relative clauses. *Linguistic Inquiry*, **18**(2), 239–266.
(1988). On the double object construction. *Linguistic Inquiry*, **19**(3), 335–391.
Lasnik, H. (1999). *Minimalist Analysis*. Oxford: Blackwell.
(2000). *Syntactic Structures Revisited*. Cambridge, MA: MIT Press.
(2001). When can you save a structure by destroying it? In *Proceedings of NELS 31*, pp. 301–320. Amherst, MA: GLSA.
(2002). "All I ever wanted to be was a teacher!" An interview conducted by L. Cheng and R. Sybesma. *Glot International*, **6**(9–10), 320–328.
(2005). Review of Jason Merchant, *The Syntax of Silence*. *Language*, **81**, 259–265.
(2006). Conceptions of the cycle. In *Wh-movement: Moving On* (ed. L. Cheng and N. Corver), pp. 197–216. Cambridge, MA: MIT Press.
Lasnik, H. and Saito, M. (1992). *Move α*. Cambridge, MA: MIT Press.
Legate, J. A. (2002). Walpiri: theoretical implications. Ph.D. thesis, MIT.
(2011). Under-inheritance. MS, University of Pennsylvania. [presented at NELS42]
Legate, J. A. and Yang, C. (2012). Assessing child and adult grammar. In *Rich Languages from Poor Inputs* (ed. R. Berwick and M. Piattelli-Palmarini), pp. 168–182. Oxford University Press.
Leiber, J. (2002). Philosophy, engineering, biology, and history: A vindication of Turing's views about the distinction between the cognitive and physical sciences. *Journal of Experimental & Theoretical Artificial Intelligence*, **14**(1), 29–37.
Lenneberg, E. H. (1963). The relationship of language to the formation of concepts. In *Boston Studies in the Philosophy of Science (Proceedings of the Boston Colloquium for the Philosophy of Science 1961/1962)* (ed. M. Wartofsky), pp. 48–54. Dordrecht: Reidel.
(1964). A biological perspective of language. In *New Directions in the Study of Language* (ed. E. Lenneberg), pp. 65–88. Cambridge, MA: MIT Press.
Leu, T. (2008). The internal syntax of determiners. Ph.D. thesis, NYU.
(2012). The indefinite article – indefinite? – indefitive? – article? *University of Pennsylvania Working Papers in Linguistics*, **18**(1), 19.
Lewontin, R. C. (2000). *The Triple Helix: Gene, Organism, and Environment*. Cambridge, MA: Harvard University Press.

Lohndal, T. (2012). Without specifiers: Phrase structure and events. Ph.D. thesis, University of Maryland.

Lohndal, T. and Uriagereka, J. (2010). The logic of parametric theory. *Theoretical Linguistics*, **36**, 69–76.

Longa, V. M. and Lorenzo, G. (2008). What about a (really) minimalist theory of language acquisition? *Linguistics*, **46**(3), 541–570.

(2012). Theoretical linguistics meets development: Explaining FL from an epigeniticist point of view. In *Language from a Biological Point of View: Current Issues in Biolinguistics* (ed. C. Boeckx, M. Horno, and J. Mendívil Giró), pp. 52–84. Cambridge, UK: Cambridge Scholars.

Longa, V. M., Lorenzo, G., and Rigau, G. (1996). Expressing modality by recycling clitics. *Catalan Working Papers in Linguistics*, **5**(1), 67–79.

(1998). Subject clitics and clitic recycling: Locative sentences in some Iberian Romance languages. *Journal of Linguistics*, **34**, 125–164.

Longobardi, G. (1994). Reference and proper names: A theory of N-movement in syntax and logical form. *Linguistic Inquiry*, **25**(4), 609–665.

(2004). Methods in parametric linguistics and cognitive history. *Linguistic Variation Yearbook*, **3**, 101–138.

(2005a). A minimalist program for parametric linguistics? In *Organizing Grammar* (ed. H. Broekhuis, N. Corver, R. Huybregts, U. Kleinhenz, and J. Koster), pp. 407–414. Berlin: Mouton de Gruyter.

(2005b). Toward a unified grammar of reference. *Zeitschrift für Sprachwissenschaft*, **24**(1), 5–44.

(2006). Reference to individuals, person, and the variety of mapping parameters. In *Essays on Nominal Determination* (ed. H. H. Müller and A. Klinge), pp. 189–211. Amsterdam: John Benjamins.

Longobardi, G. and Guardiano, C. (2011). The biolinguistic program and historical reconstruction. In *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty* (ed. A. M. Di Sciullo and C. Boeckx), pp. 266–304. Oxford University Press.

Lorenz, K. (1959). Psychologie und Stammesgeschichte. In *Evolution der Organismen* (ed. G. Heberer), pp. 131–170. Stuttgart: Fischer.

Lorenzo, G. (2013). Beyond developmental compatibility: A note on generative linguistics and the developmentalist challenge. *Teorema*, **32**, 29–44.

Lorenzo, G. and Longa, V. M. (2003). Minimizing the genes for grammar: The minimalist program as a biological framework for the study of language. *Lingua*, **113**(7), 643–657.

(2009). Beyond generative geneticism: Rethinking language acquisition from a developmentalist point of view. *Lingua*, **119**(9), 1300–1315.

McBrearty, S. and Brooks, A. S. (2000). The revolution that wasn't: A new interpretation of the origin of modern human behavior. *Journal of Human Evolution*, **39**(5), 453–563.

Manzini, M. R. and Savoia, L. M. (2011). *Grammatical Categories: Variation in Romance Languages*. Volume 128. Cambridge University Press.

Marantz, A. (1984). *On the Nature of Grammatical Relations*. Cambridge, MA: MIT Press.

(1991). Case and licensing. In *Proceedings of ESCOL'91*, pp. 234–253.
(1995). The minimalist program. In *Government and Binding and the Minimalist Program* (ed. G. Webelhuth), pp. 349–382. Malden, MA: Blackwell.
(1996). Cat as a phrasal idiom. MS, MIT.
(1997). No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics*, **4**(2), 201–225.
(2000). Words. Handout, MIT.
(2005). Generative linguistics within the cognitive neuroscience of language. *Linguistic Review*, **22**(2-4), 429–446.
(2008). Words and phases. In *Phases in the Theory of Grammar* (ed. S.-H. Choe), pp. 191–222. Seoul: Dong In.
(2013). Verbal argument structure: Events and participants. *Lingua*, **130**, 152–168.

Marcus, G. (2006). Cognitive architecture and descent with modification. *Cognition*, **101**(2), 443–465.

Martín, F. J. (2012). Deconstructing Catalan object clitics. Ph.D. thesis, New York University.

Marušič, F. (2005). On non-simultaneous phases. Ph.D. thesis, Stony Brook University.

Marvin, T. (2002). Topics in the stress and syntax of words. Ph.D. thesis, MIT.

Masullo, P. J. (2008). The syntax–lexical semantics interface: Prepositionalizing motion verbs in Spanish. MS, University of Pittsburgh.

Mateu, J. (2005). Impossible primitives. In *The Compositionality of Meaning and Content* (ed. M. Werning, E. Machery, and G. Schurz), pp. 213–229. Heusenstamm: Ontos Verlag.

Mathieu, E. (2011). Wh-in-situ and external parameters. Presented at the workshop Formal Grammar and Syntactic Variation, Universidad Autónoma de Madrid.

Mathieu, E. and Barrie, M. (2011). Macroparameters don't exist: The case of polysynthesis and noun incorporation. Presented at the workshop Formal Grammar and Syntactic Variation, Universidad Autónoma de Madrid.

Mayr, E. (1959). Darwin and the evolutionary theory in biology. In *Evolution and Anthropology: A Centennial Appraisal* (ed. J. Meggers), pp. 1–10. The Anthropological Society of Washington, Washington D.C.

Medina, M. L. (2010). Two "evo-devos." *Biological Theory*, **5**, 7–11.

Mellars, P., Boyle, K., Bar-Yosef, O., and Stringer, C. (eds.) (2007). *Rethinking the Human Revolution*. Cambridge, UK: McDonald Institute Monographs.

Merchant, J. (2001). *The Syntax of Silence: Sluicing, Islands, and Identity in Ellipsis*. Oxford University Press.

Miller, G. and Chomsky, N. (1963). Finitary models of language users. In *Handbook of Mathematical Psychology* (ed. R. D. Luce, R. Bush, and E. Galanter), pp. 419–491. New York: Wiley.

Mitrović, M. (2013). The composition of logical constants. MS, University of Cambridge.

Miyagawa, S. (2010). *Why Agree; Why Move*. Cambridge, MA: MIT Press.

Moro, A. (2000). *Dynamic Antisymmetry*. Cambridge, MA: MIT Press.

Müller, G. (2010). On deriving CED effects from the PIC. *Linguistic Inquiry*, **41**, 35–82.

Muysken, P. and van Riemsdijk, H. (ed.) (1986). *Features and Projections*. Volume XXV. Berlin: de Gruyter.

Narita, H. (2010a). Phasing in full interpretation. Ph.D. thesis, Harvard University.

(2010b). The tension between explanatory and biological adequacy. A review of *Theoretical Comparative Syntax: Studies in Macroparameters*, Naoki Fukui, Routledge, London and New York (2006). *Lingua*, **120**, 1313–1323.

Newell, H. (2008). Aspects of the morphology and phonology of phases. Ph.D. thesis, McGill University.

Newmeyer, F. (2004). Against a parameter-setting approach to language variation. *Linguistic Variation Yearbook*, **4**, 181–234.

Newmeyer, F. J. (2005). *Possible and Probable Languages: A Generative Perspective on Linguistic Typology*. Oxford University Press.

Nicolis, M. (2008). The null subject parameter and correlating properties: The case of Creole languages. In *The Limits of Variation* (ed. T. Biberauer). Amsterdam: John Benjamins.

Niyogi, P. and Berwick, R. C. (1996). A language learning model for finite parameter spaces. *Cognition*, **61**(1), 161–193.

Nunes, J. (2004). *Linearization of Chains and Sideward Movement*. Cambridge, MA: MIT Press.

Obata, M. (2010). Root, successive-cyclic and feature-splitting internal merge: implications for feature-inheritance and transfer. Ph.D. thesis, University of Michigan.

Ochi, M. (2010). Classifiers and nominal structure: A parametric approach and its consequences. MS, Osaka University.

Okanoya, K. (2012). Behavioural factors governing song complexity in Bengalese Finches. *International Journal of Comparative Psychology*, **25**, 44–59.

Osborne, T., Putnam, M., and Gross, T. M. (2011). Bare phrase structure, label-less trees, and specifier-less syntax: Is minimalism becoming a dependency grammar? *Linguistic Review*, **28**(3), 315–364.

Otero, C. (1976). The dictionary in a generative grammar. Presented at the 91st Annual Convention of the MLA, New York.

(1983). Towards a model of paradigmatic grammar. *Quaderni di Semantica*, **4**, 134–144.

(1996). Head movement, cliticization, precompilation, and word insertion (comments on Uriagereka's paper). In *Current Issues in Comparative Grammar* (ed. R. Freidin), pp. 296–337. Dordrecht: Kluwer.

Ott, D. (2008). Notes on noun ph(r)ases. MS, Harvard University.

(2009a). The conceptual necessity of phases: Some remarks on the minimalist enterprise. In *Explorations of Phase Theory: Interpretation at the Interfaces* (ed. K. Grohmann), pp. 253–275. Berlin: Mouton de Gruyter.

(2009b). Stylistic Fronting as Remnant Movement. *Working Papers in Scandinavian Syntax*, **83**, 141–178.

(2009c). The evolution of I-Language: Lexicalization as the key evolutionary novelty. *Biolinguistics*, **3**(2), 255–269.

(2011). A note on free relative clauses in the theory of phases. *Linguistic Inquiry*, **42**, 183–192.

Ouali, H. (2007). Unifying agreement relations: A minimalist analysis of Berber. Ph.D. thesis, University of Michigan.

Ouhalla, J. (1991). *Functional Categories and Parametric Variation*. London: Routledge.

Pantcheva, M. (2011). Decomposing path: The nanosyntax of directional expressions. Ph.D. thesis, Universitetet i Tromsø.

Parsons, T. (1990). *Events in the Semantics of English*. Cambridge, MA: MIT Press.

Pearl, L. S. (2007). Necessary bias in natural language learning. Ph.D. thesis, University of Maryland.

Pesetsky, D. (1995). *Zero Syntax: Experiencers and Cascades*. Cambridge, MA: MIT Press.

Pesetsky, D. and Torrego, E. (2001). T-to-C movement: Causes and consequences. In *Ken Hale: A Life in Language* (ed. M. Kenstowicz), pp. 355–426. Cambridge, MA: MIT Press.

(2007). The syntax of valuation and the interpretability of features. In *Phrasal and Clausal Architecture: Syntactic Derivation and Interpretation* (ed. S. Karimi, V. Samiian, and W. Wilkins), pp. 262–294. Amsterdam: John Benjamins.

Phillips, C. (2006). The real-time status of island phenomena. *Language*, **82**(4), 795–823.

Piattelli-Palmarini, M. (1989). Evolution, selection and cognition: from 'learning' to parameter setting in biology and in the study of language. *Cognition*, **31**, 1–44.

(2010). What is language that it may have evolved, and what is evolution that ' it may apply to language? In *The Evolution of Human Language: Biolinguistic Perspectives* (ed. R. Larson, H. Déprez, and V. Yamakido), pp. 148–162. Cambridge University Press.

Piattelli-Palmarini, M. and Uriagereka, J. (2005). The evolution of the narrow language faculty: The skeptical view and a reasonable conjecture. *Lingue e Linguaggio*, **4**, 27–79.

(2011). FoxP2: A geneticist's dream; a linguist's nightmare. In *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty* (ed. A.-M. di Sciullo and C. Boeckx), pp. 100–125. Oxford University Press.

Picallo, M. C. (2006). On gender and number. MS, Universitat Autònoma de Barcelona.

(2008). On gender and number in Romance. *Lingue e Linguaggio*, **1**, 44–66.

Piera, C. (1985). On the representation of higher order complex words. In *Selected Papers from XIII Linguistic Symposium on Romance Languages, Chapel Hill, NC, 24–26 March 1983*, pp. 287–313. Amsterdam: John Benjamins.

Pietroski, P. M. (2003a). Quantification and second-order monadicity. *Philosophical Perspectives*, **17**, 259–298.

(2003b). Small verbs, complex events: Analyticity without synonymy. In *Chomsky and His Critics* (ed. N. Hornstein and L. Antony), pp. 179–214. Oxford: Wiley.

(2005). *Events and Semantic Architecture*. Oxford University Press.

(2007). Systematicity via monadicity. *Croatian Journal of Philosophy*, **7**, 343–374.

(2008). Minimalist meaning, internalist interpretation. *Biolinguistics*, **2**(4), 317–341.
(2010). Concepts, meanings and truth: First nature, second nature and hard work. *Mind & Language*, **25**(3), 247–278.
(2011). Minimal semantic instructions. In *Oxford Handbook of Linguistic Minimalism* (ed. C. Boeckx), pp. 472–498. Oxford University Press.
(2012). Language and conceptual reanalysis. In *Towards a Biolinguistic Understanding of Grammar: Essays on Interfaces* (ed. A. Di Sciullo). Ansermenn: John Benjamins.
(forthcoming). *Conjoining Meanings: Semantics without Truth Values*. Oxford University Press.
Pinker, S. and Bloom, P. (1990). Natural selection and natural language. *Behavioral and Brain Sciences*, **13**, 707–784.
Pinker, S. and Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition*, **95**(2), 201–236.
Poeppel, D. (2005). The interdisciplinary study of language and its challenges. Technical report, Jahrbuch des Wissenschaftskollegs zu Berlin.
(2012). The maps problem and the mapping problem: Two challenges for a cognitive neuroscience of speech and language. *Cognitive Neuropsychology*, **29**(1–2), 34–55.
Poeppel, D. and Embick, D. (2005). Defining the relation between linguistics and neuroscience. In *Twenty-first Century Psycholinguistics: Four Cornerstones* (ed. A. Cutler), pp. 173–189. Hillsdale, NJ: Erlbaum.
Pourquié, O. (2003). The segmentation clock: Converting embryonic time into spatial pattern. *Science Signaling*, **301**(5631), 328.
Pullum, G. K. (1996). Nostalgic views from Building 20. *Journal of Linguistics*, **32**(1), 137–147.
Pylkkänen, L. (2002). Introducing arguments. Ph.D. thesis, MIT.
(2008). *Introducing Arguments*. Cambridge, MA: MIT Press.
Ramchand, G. (2008). *Verb Meaning and the Lexicon: A First-Phase Syntax*. Cambridge University Press.
Ramchand, G. and Svenonius, P. (2008). Mapping a parochial lexicon onto a universal semantics. In *The Limits of Syntactic Variation* (ed. T. Biberauer), pp. 219–245. Amsterdam: John Benjamins.
(2013). Deriving the functional hierarchy. Presented at 36th GLOW Colloquium, Lunds Universitet.
Raposo, E. (2002). Nominal gaps with prepositional modifiers in Portuguese and Spanish: A case for quick spell-out. *Cuadernos de Lingüística del I. U. Ortega y Gasset*, **9**, 127–144.
Reinhart, T. (2006). *Interface Strategies: Optimal and Costly Computations*. Cambridge, MA: MIT Press.
Reinhart, T. and Siloni, T. (2005). The lexicon–syntax parameter: Reflexivization and other arity operations. *Linguistic Inquiry*, **36**(3), 389–436.
Reiss, C. (2003). Quantification in structural descriptions: Attested and unattested patterns. *Linguistic Review*, **20**(2–4), 305–338.
Richards, M. (2004). Object shift and scrambling in North and West Germanic: A case study in symmetrical syntax. Ph.D. thesis, University of Cambridge.

(2007). On feature inheritance: An argument from the phase impenetrability condition. *Linguistic Inquiry*, **38**(3), 563–572.
(2008a). Two kinds of variation in a minimalist system. *Linguistische Arbeits Berichte*, **87**, 133–162.
(2009a). Internal pair-merge: The missing mode of movement. *Catalan Journal of Linguistics*, **8**, 55–73.
(2009b). Stabilizing syntax: On instability, optionality, and other indeterminacies. Presented at the Syntax workshop, Universität Stuttgart, March 2010.
(2011). Deriving the edge: What's in a phase? *Syntax*, **14**, 74–95.
(2012). On feature inheritance, defective phases, and the movement–morphology connection. In *Phases: Developing the Framework* (ed. Á. Gallego), pp. 195–232. Berlin: Mouton de Gruyter.
Richards, N. (2001). *Movement in Language: Interactions and Architectures*. Oxford University Press.
(2008b). Lardil 'case stacking' and the structural/inherent case distinction. MS, MIT.
(2010). *Uttering Trees*. Cambridge, MA: MIT Press.
Rizzi, L. (1978). Violations of the wh-island condition in Italian and the subjacency condition. *Montreal Working Papers in Linguistics*.
(1982). *Italian Syntax*. Dordrecht: Foris.
(1986). Null objects in Italian and the theory of pro. *Linguistic Inquiry*, **17**(3), 501–557.
(1997). The fine structure of the left periphery. In *Elements of Grammar: Handbook in Generative Syntax* (ed. L. Haegeman), pp. 281–337. Dordrecht: Kluwer.
(2001). On the position"Int(errogative)" in the left periphery of the clause. In *Current Studies in Italian Syntax* (ed. G. Cinque and G. Salvi), pp. 287–296. Amsterdam: Elsevier.
(2004). On the cartography of syntactic structures. In *The Structure of CP and IP* (ed. L. Rizzi), pp. 3–15. Oxford University Press.
(2006). Grammatically-based target-inconsistencies in child language. In *The Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition – North America* (ed. K. Deen, J. Nomura, B. Schulz, and B. Schwartz), pp. 19–49. University of Connecticut Occasional Papers in Linguistics, 4.
(2009). Some elements of syntactic computation. In *Biological Foundations and Origin of Syntax* (ed. D. Bickerton and E. Szathmáry), pp. 63–88. Cambridge, MA: MIT Press.
(2010). On the elements of syntactic variation. Presented at the workshop on linguistic variation and the minimalist program, Barcelona, January 2010.
Rizzi, L. and Shlonsky, U. (2007). Strategies of subject extraction. In *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Semantics* (ed. U. Sauerland and H.-M. Gärtner), pp. 115–160. Berlin: Mouton de Gruyter.
Roberts, I. G. (2001). Language change and learnability. In *Language Acquisition and Learnability* (ed. S. Bertolo), pp. 81–125. Cambridge University Press.
(2010a). *Agreement and Head Movement: Clitics, Incorporation, and Defective Goals*. Cambridge, MA: MIT Press.

(2010b). On the nature of syntactic parameters: A programme for research. Presented at the 2010 Mayfest 'Bridging Typology and Acquisition.'

(2011). Parametric hierarchies: Some observations. Presented at the workshop on linguistic variation and the minimalist program, Universidad Autonoma de Madrid.

Roberts, I. G. and Holmberg, A. (2009). Introduction: Parameters in minimalist theory. In *Parametric Variation: Null Subjects in Minimalist Theory* (ed. T. Biberauer, A. Holmberg, I. Roberts, and M. Sheehan), pp. 1–57. Cambridge University Press.

Ross, J. R. (1967). Constraints on variables in syntax. Ph.D. thesis, MIT.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, **274**, 1926–1928.

Safir, K. (2014). One true anaphor. *Linguistic Inquiry*, **45**, 91–124.

Saito, M., Lin, T. H. J., and Murasugi, K. (2008). N′-ellipsis and the structure of noun phrases in Chinese and Japanese. *Journal of East Asian Linguistics*, **17**(3), 247–271.

Sakas, W. G. and Fodor, J. D. (2012). Disambiguating syntactic triggers. *Language Acquisition*, **19**(2), 83–143.

Samuels, B. (2009). The structure of phonological theory. Ph.D. thesis, Harvard University.

(2010). Consequences of phases for morphophonology. In *Phases: Developing the Framework* (ed. A. Gallego), pp. 251–282. Berlin: Walter de Gruyter.

(2011a). A minimalist program for phonology. In *The Oxford Handbook of Linguistic Minimalism* (ed. C. Boeckx), pp. 574–594. Oxford University Press.

(2011b). *Phonological Architecture: A Biolinguistic Perspective*. Oxford University Press.

(2012a). Animal minds and the roots of human language. In *Language, from a Biological Point of View* (ed. C. Boeckx, J. Mendívil Giró, and M. Horno), pp. 290–313. Cambridge UK: Cambridge Scholars.

(2012b). The emergence of phonological forms. In *Towards a Biolinguistic Understanding of Grammar* (ed. A. Di Sciullo), pp. 193–213. Amsterdam: John Benjamins.

Samuels, B., Hauser, M. D., and Boeckx, C. (in press). Do animals have Universal Grammar? A case study in phonology. In *The Oxford Handbook of Universal Grammar* (ed. I. Roberts). Oxford University Press.

Saunders, P. T. (1992). *Collected Works of A. M. Turing: Morphogenesis*. Amsterdam: North Holland.

Savescu Ciucivara, O. and Wood, J. (2010). Re-prefixation and Talmy's parameter. In *Proceedings of NELS 36*, pp. 2–13.

Schein, B. (1993). *Plurals and Events*. Volume XXIII. Cambridge, MA: MIT Press.

(2002). Events and the semantic content of thematic relations. In *Logical Form and Language* (ed. G. Preyer and G. Peter), pp. 263–344. Oxford University Press.

Schneider-Zioga, K. (2013). The linker in kinande re-examined. MS, California State University.

Searle, J. (1972). Chomsky's revolution in linguistics. *New York Review of Books*, **18**(12), 16–24.

(2006). What is language: Some preliminary remarks. MS, UC Berkeley.

Sheehan, M. (2013). Towards a general alignment parameter hierarchy. In *Proceedings of the 31st West Coast Conference on Formal Linguistics.* Somerville, MA: Gascadilla.

Shlonsky, U. (2006). Extended projection and CP cartography. *Nouveaux cahiers de linguistique française*, **27**, 83–93.

Sigurdsson, H. Á. (2004a). Meaningful silence, meaningless sounds. *Linguistic Variation Yearbook*, **4**(1), 235–259.

(2004b). The syntax of person, tense, and speech features. *Rivista di linguistica*, **16**, 219–251.

(2011). On UG and materialization. *Linguistic Analysis*, **37**, 367–388.

Šimík, R. (2011). The elimination of formal wh-features and a theory of free wh-movement. MS, Universität Potsdam.

Simpson, A. (2005). Classifiers and DP structure in Southeast Asia. In *The Oxford Handbook of Comparative Syntax* (ed. G. Cinque and R. Kayne), pp. 806–838. Oxford University Press.

Smith, N. and Law, A. (2009). On parametric (and non-parametric) variation. *Biolinguistics*, **3**(4), 332–343.

Snyder, W. (1995). Language acquisition and language variation: The role of morphology. Ph.D. thesis, MIT.

(2000). An experimental investigation of syntactic satiation effects. *Linguistic Inquiry*, **31**(3), 575–582.

(2001). On the nature of syntactic variation: Evidence from complex predicates and complex word-formation. *Language*, **77**(2), 324–342.

(2011). On language acquisition and syntactic theory. Presented at 50 Years of Linguistics at MIT.

Son, M. (2006). Directed motion and non-predicative Path P. *Nordlyd: Tromsø Working Papers on Language and Linguistics*, 176–199.

Son, M. and Svenonius, P. (forthcoming). Microparameters of cross-linguistic variation: Directed motion and resultatives. In *Proceedings of the 27th West Coast Conference on Formal Linguistics*.

Speas, P. (2010). The minimal structure of the left periphery. Presented at the Language Design conference, UQAM, Montreal, May 2010.

Speas, P. and Tenny, C. (2003). Configurational properties of point of view roles. In *Asymmetry in Grammar Vol-I: Syntax and Semantics* (ed. A.-M. di Sciullo), pp. 315–344. Amsterdam: John Benjamins.

Spelke, E. (2003). What makes us smart? Core knowledge and natural language. In *Language and Mind: Advances in the Study of Language and Thought* (ed. D. Gentner and S. Goldin-Meadow), pp. 277–311. Cambridge, MA: MIT Press.

Spelke, E. and Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, **10**(1), 89–96.

Sprouse, J. (2007). A program for experimental syntax. Ph.D. thesis, University of Maryland.

Sprouse, J. and Lau, E. (2013). Syntax and the brain. In *The Handbook of Generative Syntax* (ed. M. den Dikken), pp. 971–1005. Cambridge University Press.

Starke, M. (2001). Move dissolves into merge. Ph.D. thesis, Université de Genève.

(2010). Nanosyntax: A short primer to a new approach to language. *Nordlyd*, **36**, 1–6.

Stepanov, A. (2001). Cyclic domains in syntactic theory. Ph.D. thesis, University of Connecticut.

(2007). The end of CED? Minimalism and extraction domains. *Syntax*, **10**(1), 80–126.

Stepanov, A. and Stateva, P. (2006). Successive cyclicity as residual wh-scope marking. *Lingua*, **116**(12), 2107–2153.

Stuhlmueller, A., Tenenbaum, J. B., and Goodman, N. D. (2010). Learning structured generative concepts. *Proceedings of the Thirty-Second Annual Conference of the Cognitive Science Society*.

Sugisaki, K. (2011). Preposition stranding: Its parametric variation and acquisition. Paper presented at the Workshop on Verbal Elasticity, Bellaterra, October 2011.

Svenonius, P. (2007). Adpositions, particles and the arguments they introduce. In *Argument Structure* (ed. K. S. T. Bhattacharya and E. Reuland), pp. 63–103. Amsterdam: John Benjamins.

(2012a). Lexical insertion in two stages. MS, Universitetet i Tromsø.

(2012b). Merge, project, and bundle. MS, Universitetet i Tromsø.

(2012c). Spanning. MS, Universitetet i Tromsø.

Svenonius, P., Ramchand, G., Taraldsen, T., and Starke, M. (eds.) (2009). *Nordlyd 36. Special issue on Nanosyntax*. Universitetet i Tromsø.

Szabolcsi, A. (1984). The possessor that ran away from home. *Linguistic Review*, **3**, 89–102.

(2006). Strong vs. weak islands. In *The Blackwell Companion to Syntax* (ed. M. Everaert and H. van Riemsdijk), vol. IV, pp. 479–531. Malden, MA: Blackwell.

Takahashi, S. (2010). The hidden side of clausal complements. *Natural Language and Linguistic Theory*, **28**, 343–380.

Takahashi, S. and Hulsey, S. (2009). Wholesale late merger: Beyond the A/$\overline{A}$-distinction. *Linguistic Inquiry*, **40**, 387–426.

Tattersall, I. (1998). *Becoming Human: Evolution and Human Uniqueness*. Oxford University Press.

Taylor, H. (2009). The complementizer *the*. In *Little Words* (ed. R. Leow, H. Campos, and D. Lardiere), pp. 87–98. Washington, DC: Georgetown University Press.

Tokizaki, H. (2010). Recursive compounds and wordstress location. Paper presented at On Linguistic Interfaces II, University of Ulster, Belfast, Ireland.

(2011). Stress location and the acquisition of morphosyntactic parameters. WCCFL 28 Online Proceedings. http://sites.google.com/site/wccfl28pro/tokizaki.

Toyoshima, T. (2011). Parameterized linearization of multidominance bare phrase structures. Presented at the Parallel Domains workshop, University of Southern California.

Truswell, R. (2007). Locality of Wh-movement and the individuation of events. Ph.D. thesis, University College London.

Tsai, W. D. (2010). High applicatives are not high enough: A cartographic solution. MS, Chinese University of Hong Kong.

Tungseth, M. E. (2008). *Verbal Prepositions and Argument Structure: Path, Place and Possession in Norwegian*. Amsterdam: John Benjamins.

Turing, A. M. (1936). On computable numbers, with an application to the *Entscheidungsproblem*. *Proceedings of the London Mathematical Society*, **2**(42), 230–265.

(1952). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London Series B*, **237**, 37–72.

Turing, A. M. and Wardlaw, C. (1953 [1992]). A diffusion reaction theory of morphogenesis in plants. In *Collected Works of A. M. Turing: Morphogenesis* (ed. P. T. Saunders), pp. 37–118. Amsterdam: North Holland.

Uriagereka, J. (1988). On government. Ph.D. thesis, University of Connecticut.

(1998). *Rhyme and Reason: An Introduction to Minimalist Syntax*. Cambridge, MA: MIT Press.

(1999). Multiple spell-out. In *Working Minimalism* (ed. S. D. Epstein and N. Hornstein), pp. 251–282. Cambridge, MA: MIT Press.

(2008). *Syntactic Anchors: On Semantic Structuring*. Cambridge University Press.

(2012). *Spell-Out and the Minimalist Program*. Oxford University Press.

van Craenenbroeck, J. and van Koppen, M. (2002). The locality of agreement and the CP-domain. Presented at 25th GLOW Colloquium.

van Riemsdijk, H. (2008). Identity avoidance: OCP-effects in Swiss Relatives. In *Foundational Issues in Linguistics* (ed. R. Freidin, C. Otero, and M.-L. Zubizarreta), pp. 227–250. Cambridge, MA: MIT Press.

Watanabe, A. (1995). Conceptual basis of cyclicity. In *MIT Working Papers in Linguistics 27*, pp. 269–291.

(2009). Vague quantity, numerals, and natural numbers. *Syntax*, **13**(1), 37–77.

Webelhuth, G. (1992). *Principles and Parameters of Syntactic Saturation*. Oxford University Press.

West-Eberhard, M. J. (2003). *Developmental Plasticity and Evolution*. Oxford University Press.

Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua*, **106**, 23–79.

Williams, A. (2009). Themes, cumulativity, and resultatives: Comments on Kratzer 2003. *Linguistic Inquiry*, **40**(4), 686–700.

Williams, N. A., Close, J. P., Giouzeli, M., and Crow, T. J. (2006). Accelerated evolution of protocadherin11X/Y: A candidate gene-pair for cerebral asymmetry and language. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, **141**(6), 623–633.

Wiltschko, M. (2009). The composition of INFL. An exploration of tense, tenseless languages and tenseless constructions. MS, University of British Columbia.

Yang, C. (2002). *Knowledge and Learning in Natural Language*. Oxford University Press.

(2004). Universal Grammar, statistics or both? *Trends in Cognitive Sciences*, **8**(10), 451–456.

(2005). On productivity. *Linguistic Variation Yearbook*, **5**(1), 265–302.

(2010). Three factors in language acquisition. *Lingua*, **120**, 1160–1177.

Yip, M.J. (2006). The search for phonology in other species. *Trends in Cognitive Sciences*, **10**(10), 442–446.

Zanuttini, R. (2010). A syntactic analysis of interpretive restrictions on imperative, promissive, and exhortative subjects. MS, Yale University.

Index

Abels, Klaus, 71, 84, 85
Acquaviva, Paolo, 29
Acedo-Matellán, Víctor, 166
adequacy
  descriptive, 1, 4
  explanatory, 1–164
Agree, 6, 10, 13, 30, 39, 50, 51, 53, 54, 56, 65, 72–74, 76, 78–79, 112, 115, 123, 133, 139, 141, 153, 155, 167–168, 170, 172
Adger, David, 4, 5, 11–14, 23, 45, 68, 83, 115, 168
Arregi, Karlos, 76
Arsenijević, Boban, 40, 57, 58, 85, 90, 108

Baker, Marc C., 17, 49, 50, 54, 75, 86, 92, 93, 121–123, 135–139, 154, 158, 163, 167–168, 169
Balari, Sergio, 22, 24
Benítez-Burraco, Antonio, 20, 126, 148, 163
Bentzen, Kristine, 84, 85
Berwick, Robert C., 10, 33–35, 127, 128, 136, 137
Bhatt, Rajesh, 76
Biberauer, Theresa, 110, 140, 172, 173
biolinguistics, 1, 4, 13, 15, 16, 21, 22, 24, 26, 33, 35, 63 88, 95, 135, 149, 154, 163
birdsong, 33–35
Bloom, Paul, 17, 128
Bloomfield, Leonard, 10, 155
Bobaljik, Jonathan, 73, 78, 138, 167
Boeckx, Cedric, 5, 9, 15, 16, 18, 20–22, 26, 27, 32, 38, 40, 44, 51, 54–56, 65, 66, 70, 71, 78–85, 91, 93, 102, 103, 107, 118, 119, 121, 126, 127, 141–144, 146, 148–172
Borer, Hagit, 5, 9, 11, 13, 23, 30–32, 43, 48, 62, 92, 120, 124, 136–140
Borer–Chomsky conjecture, 5, 120, 124, 136–139
Bošković, Željko, 6, 42, 44, 55, 71, 73, 79, 93, 134, 135
Bosque, Ignacio, 69
bundle, 7–11, 13, 25, 28, 37, 39, 40, 43, 77, 111–112, 114–116, 138, 140
Bye, Patrik, 111–115

Caponigro, Ivano, 4, 57, 58, 154
Carnie, Andrew, 20
cartography, 6, 14, 15, 19, 20, 30, 47, 49, 52, 55–56, 59, 87, 88, 107, 108, 150
Carstens, Vicki, 12
Chomsky, Noam, 1–29, 31, 36–43, 45, 48, 50–51, 53, 55, 56, 69–81, 83, 94, 97, 98, 105, 109, 110, 115, 120–124, 127, 129–132, 135–141, 149, 150, 152–155, 157–163
Chomsky–Baker conjecture, 136, 137, 139
Cinque, Guglielmo, 12, 14, 15, 20, 30, 51, 55, 60–62, 107, 108, 125, 135
Cinque hierarchy, 12, 20, 51, 55, 56, 61, 107
cognitive science, 1, 2, 32, 150, 163, 164
Collins, Chris, 4, 5, 20, 38, 45, 50, 91, 167, 168
complex systems, 18, 163
compositionality, 4, 31, 33, 38, 39, 116, 165
Corbett, Greville, 2, 3, 7
van Craenenbroeck, Jeroen, 31, 88, 115

Darwin, Charles, 18, 52, 59, 101, 163
Dawkins, Richard, 17
De Belder, Marijke, 31, 115
Deacon, T. W., 18, 110, 126
Dehaene, Stanislas, 15
Di Sciullo, Anna Maria, 22, 149

Distributed Morphology, 13, 28, 30, 31, 73, 76, 88, 92, 97, 110–112, 113, 115–118, 124, 137, 155, 170
Duguine, Maia, 170

Embick, David, 21, 110, 112
Emonds, Joseph, 4, 39, 51, 152
Endress, Ansgar, 116, 143–144
Epstein, Samuel D., 5, 10, 42, 74, 78, 79, 93
Etxepare, Ricardo, 173
evo-devo, 20, 163
Extended Projection Principle, 2, 3, 6, 12, 19, 78, 79, 139, 172

Fasanella-Seligrat, Adriana, 116, 133
feature, 2–13, 16, 19–20, 22, 24, 26–32, 34–51, 55–81, 84, 85, 88–89, 91–93, 93, 97, 101–102, 107–109, 111–117, 120, 126, 138, 139, 141, 146, 148, 150, 152, 153, 155, 158, 167, 169–170, 172
Felix, Sascha, 4
Final-over-Final Constraint, 172–173
Fitch, W. Tecumseh, 11, 21, 22, 157
Fodor, Janet Dean, 128, 141
Fodor, Jerry A., 17, 127
Fortuny, Jordi, 14, 15, 116, 138
Frampton, John, 5
Freidin, Robert A., 38
Fujita, Koji, 16
Fukui, Naoki, 3, 53, 92, 137

Gallego, Ángel J., 39, 49, 69, 75, 78, 83, 86, 93, 139
generative grammar, 2, 15, 17, 21, 26, 38, 82, 83, 92, 123, 129, 149, 151, 155, 169
Georgi, Doreen, 53
Gervain, Judit, 116, 143, 144
Gould, Stephen Jay, 17
Government & Binding, 6, 13, 15–17, 19, 21, 86, 119, 121–123
Grohmann, Kleanthes K., 40, 71, 85, 86
Guardiano, Cristina, 123, 129
Gutmann, Sam, 5

Haddican, Bill, 173
Haegeman, Liliane, 6
Hale, Kenneth, 46, 92, 99, 117, 138, 140, 141, 150, 155
Halle, Morris, 30, 92, 111, 141
Harbour, Daniel, 14, 59, 62
Harley, Heidi, 28, 30, 31, 50, 111, 139
Haspelmath, Martin, 58, 135
Hauser, Marc D., 5, 11, 16, 17, 98, 101, 102, 157
Heine, Bernd, 48–147
Hinzen, Wolfram, 40, 80, 85, 90, 93, 98–100, 107–109, 137, 147, 149
Holmberg, Anders, 131–133, 141, 171
Homo sapiens, 16, 18, 101, 125
Hornstein, Norbert, 4, 11, 16, 19, 26, 72, 83, 85, 88, 91, 103, 105

I-language, 17, 63, 94, 95, 123, 135, 146, 147
Irurtzun, Aritz, 57, 88, 105

Jackendoff, Ray, 9, 11, 16, 17, 23, 27, 86, 101, 147, 149, 150, 154, 155
Jenks, Peter, 54, 55, 96, 171
Jeong, Youngmi, 55, 168

Katz, Jonah, 29
Kauffman, Stuart, 21
Kayne, Richard S., 4, 44–50, 52, 53, 57, 66, 71, 75, 104, 113–115, 117, 118, 125, 133–135, 139, 169
Kibort, Anna, 2, 3, 7
Kinsella, Anna R., 16
Kinzler, Katherine D., 15
Keyser, Samuel Jay, 46, 92, 99, 117, 141, 150, 155
Koster, Jan, 15, 20, 24, 63, 130, 150
Kuteva, Tania, 48, 49, 54, 61

Lakoff, George, 2, 150
language faculty, 1, 2, 4, 11, 14, 16–18, 21, 22, 25, 26, 33, 47, 52, 62, 63, 80, 87, 90, 97, 98, 122, 123, 129, 135, 143, 152, 157, 168, 169
  broad, 5, 14, 16, 98
  narrow, 14, 16, 24, 101
Lasnik, Howard, 2, 6, 13, 38, 81, 127, 128, 153, 154, 168
Lau, Ellen F., 7
Legate, Julie Anne, 39, 142–145, 168
Leivada, Evelina, 127, 141, 143, 171, 172
Leu, Thomas, 49, 50, 54
Lewontin, Richard C., 17, 158
lexicocentrism, 2–5, 7, 12, 13, 15, 16, 19, 23–28, 30, 35, 39, 40, 43, 46, 49, 62, 72, 76, 79, 83, 85, 86, 88, 92, 93, 95,

99, 108–111, 114–117, 120, 137, 147, 148, 154, 164, 171
lexicon, 2–6, 9–11, 13, 20, 23–24, 33, 37, 43, 50, 51, 73, 77, 86, 88, 90, 92, 109, 113, 120–121, 124, 129, 136–140, 150–156, 156, 164, 169, 171
narrow, 26–31, 35, 41, 47, 70, 73, 98, 137
Linear Correspondence Axiom, 45, 118
Lohndal, Terje, 49, 103, 133
Longa, Víctor, 20, 52, 126, 129, 163
Longobardi, Giuseppe, 3, 16, 53, 96, 123, 129, 139, 141
Lorenzo, Guillermo, 22, 126, 129

Manzini, M. Rita, 31, 50, 57
Marantz, Alec, 3, 13, 23, 26, 30, 31, 43, 48, 73, 92, 97, 111, 149, 150, 152, 157
Marcus, Gary, 22, 158
Marler, Peter, 33–34,
Martín, Txuss, 50, 51, 53, 54
Martínez-Álvarez, Anna, 148
Mateu, Jaume, 46, 49
Mehler, Jacques, 116, 143, 144
Merge, 4–10, 13, 14, 23, 25–30, 34–38, 40, 42–46, 48, 51, 52, 55, 57, 59, 62–66, 69–71, 74, 75, 79–81 85, 89–94, 98, 102, 103, 105, 107, 113, 115–116, 118, 119, 123, 131, 139, 148, 167
internal, 65, 69–71, 73, 77, 80, 81, 85, 105, 106, 117
minimalism, 2, 4–12, 15–17, 19–23, 25–27, 31–32, 37, 55–56, 59, 62, 70, 72–74, 76, 81–86, 92, 98, 99, 101, 107, 119–121, 123, 130–132, 136, 146, 150, 154, 157, 171
Miyagawa, Shigeru, 30, 72
Moro, Andrea, 36, 37, 40, 45, 70, 92, 167
Müller, Gereon, 24, 42, 53, 81, 84, 93
Muysken, Pieter, 3, 7, 29, 36

nanosyntax, 13, 20, 111, 116, 150, 151, 170
Narita, Hiroki, 16, 42, 45, 66, 82
natural selection, 17, 18, 52, 95, 102
Nevins, Andrew, 76
Newmeyer, Frederick J., 6, 125, 130, 132, 133, 138
Noyer, Rolf, 110, 111
Nunes, Jairo, 70

Ott, Dennis, 32, 79, 80, 98, 133
Ouali, Hamid, 39

Panagiotidis, Phoevos, 29
Pesetsky, David, 10, 12, 29, 47, 73
Piattelli-Palmarini, Massimo, 9, 18, 21, 146, 148, 150
Picallo, Carme, 51
Pietroski, Paul, 31–33, 88, 99–101, 103–105, 105, 109, 166
Pinker, Steven, 11, 16, 17, 128
Plato's problem, 9, 119, 128, 133–135, 146
Platzack, Christer, 133
Poeppel, David, 21–23, 26, 35, 147
Polinsky, Maria, 4, 57, 58, 154
Principles & Parameters, 5, 10, 119–122, 127, 129–131, 133, 136, 140, 141, 144, 146
principles, 7, 10, 15, 16, 19, 21, 62, 87, 112, 113, 115, 119–120, 122, 123, 130–132, 136–172
parameters (*see also* variation), 2, 6, 10, 17, 32, 55, 94, 96, 112, 113, 118–141, 142–144, 146, 154, 156, 165, 173
Projection Principle, 6, 26, 138
Pylkkänen, Liina, 50, 51, 57, 138

Ramchand, Gillian C., 49, 107, 136
Raposo, Eduardo, 65
recursion, 12, 22, 24, 29, 35, 85, 90
Relativized Minimality, 40, 58, 80, 168
Reuland, Eric, 3, 21
Richards, Marc, 16, 37, 39, 56, 63, 65, 75, 76, 78, 110
Richards, Norvin, 39–42, 44, 63–66, 70, 82, 83, 89, 97, 166, 167
van Riemsdijk, Henk, 3, 7, 29, 36, 40, 131, 168
Rizzi, Luigi, 10, 14, 20, 51, 55–57, 60–62, 83, 105, 108, 125, 127, 130, 133, 139, 170
Roberts, Ian, 78, 120, 121, 125, 131, 141, 171–172
Ross, John R., 67

Saito, Mamoru, 13, 55
Sakas, Gregory William, 127, 128
Salanova, Andrés, 4
Samuels, Bridget, 83, 97, 111, 118, 141
Savoia, Leonardo M., 31, 50, 57
Scheer, Tobias, 38
Seely, T. Daniel, 5, 42, 74, 79, 93
Sheehan, Michelle, 107, 108, 125, 126

Shlonsky, Ur, 53, 55, 56, 83, 133
Sigurdsson, Halldór Ármann, 31, 88, 138
Šimík, Radek, 7
Snyder, William, 47, 82, 135, 139, 165, 167–169
Speas, Margaret, 3, 88, 105, 107
Spelke, Elizabeth S., 15, 46, 101
Sprouse, Jon, 7, 82
Starke, Michal, 20, 30, 60, 67, 80, 111, 151
Strong Uniformity Thesis, 119, 136
Svenonius, Peter, 11–13, 48, 49, 107, 111–116, 136, 166

Tattersall, Ian, 18, 102
third factor, 15, 28, 40, 62, 87, 89, 94, 95, 115, 120, 142, 143, 163, 172
Thráinsson, Höskuldur, 138
Torrego, Esther, 10, 12, 29, 73
Turing, Alan, 28, 90, 94, 95, 163, 164

Unify, 9, 11, 23
Universal Grammar, 1, 4, 5, 7, 10, 13, 16–17, 22, 26, 36, 38, 47, 58, 61, 62, 94, 107, 110, 121–123, 125–129, 131, 132, 143, 146, 152, 159, 167, 171
Ura, Hiroyuki, 20
Uriagereka, Juan, 16, 18, 39, 66, 82, 88, 98, 104, 105, 133, 137, 148, 149, 158

variation, 2, 10, 13, 32, 33, 61, 82, 96, 109, 119–121, 123–124, 126, 128–131, 136–138, 140–141, 143–146, 166–173
  parametric, 5, 9–11, 45, 50, 121, 137
  semantic, 96, 137

Watanabe, Akira, 15
Walkow, Martin, 76
Wiltschko, Martina, 30

Yang, Charles, 16, 122, 127, 128, 141–145, 149

Zushi, Mihoko, 53