UNIVERSITY OF CAMBRIDGE

On the Formal Flexibility of Syntactic Categories

Chenchen Song

Supervisors: Prof. Ian Roberts
Dr. Theresa Biberauer

Theoretical and Applied Linguistics
Faculty of Modern and Medieval Languages
University of Cambridge

This thesis is submitted for the degree of
Doctor of Philosophy

Gonville & Caius College July 2019

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Chenchen Song
July 2019

On the Formal Flexibility of Syntactic Categories

Chenchen Song

This dissertation explores the formal flexibility of syntactic categories. The main proposal is that Universal Grammar (UG) only provides templatic guidance for syntactic category formation and organization but leaves many other issues open, including issues internal to a single category and issues at the intercategorial, system level: these points that UG "does not care about" turn out to enrich the categorial ontology of human language in important ways.

The dissertation consists of seven chapters. After a general introduction in **Chapter 1**, I lay out some foundational issues regarding features and categories in **Chapter 2** and delineate a featural metalanguage comprising four components: specification, valuation, typing, and granularity. Based on that I put forward a templatic definition for syntactic categories, which unifies the combinatorial and taxonomic perspectives under the notion *mergeme*. Then, a detailed overview of the "categorial universe" I work with is presented, which shows that the syntactic category system (SCS) is an intricate web structured by five layers of abstraction divided into three broad levels of concern: the individual level (layers 1–2), the global level (layers 3–4), and the supra-global level (layer 5). In the subsequent chapters I explore the template-flexibility pairs at each abstraction layer, with **Chapters 3–4** focusing on the first layer, **Chapter 5** on the second layer, and **Chapter 6** on the third and fourth layers; the fifth layer is not in the scope of this dissertation.

Chapter 3 examines a special type of category defined by an underspecified mergeme, the *defective category*, which behaves like a "chameleon" in that it gets assimilated into whatever nondefective category it merges with. This characteristic makes it potentially useful in analyzing certain adjunction structures, and I explore this potential by two case studies, one focusing on modifier-head compounds and the other on sentence-final particles. **Chapter 4** examines another special type of category defined by the absence of a mergeme, the *Root category*. Deductive reasoning leads me to propose a *generalized root syntax*, according to which roots are not confined to lexical categorial environments but may legally merge with and hence "support" any non-Root category. I demonstrate the empirical consequences of this theory by a comprehensive study of the half-lexical–half-functional vocabulary items in Chinese.

Chapter 5 ascends to the second abstraction layer and raises the question of whether the categorial sequences (or projection hierarchies) in human language are necessarily totally ordered, as certain analytical devices (e.g., "flavored" categories) can only be theoretically maintained if we also allow categorial sequences to be partially ordered. After a diachronic study of the flavored verbalizer v_{BE} (stative) in Chinese resultative compounds, I conclude that while "flavoring" is indeed a possible type of flexibility in the SCS, it is the deviation rather than the norm due to non-UG or "third" factors and hence should be cautiously used in syntactic analyses.

Chapter 6 ascends even higher on the ladder of abstraction and examines the global inter-

connection in the SCS ontology with the aid of mathematical Category theory. I formalize the functional parallelism across major parts of speech and the inheritance-based relations across granularity levels as Category-theoretic structures, which reveal further and more abstract templates and flexibility types in the SCS. A crucial mathematical concept in the formalization is *epi-Adjunction*. Finally, in **Chapter 7** I summarize the main results of this dissertation and briefly discuss some potential directions of future research.

Acknowledgments

I would like to thank everyone who has helped me during my PhD studies in one way or another. First of all, I am truly grateful to my supervisors, Ian Roberts and Theresa Biberauer, for their patient guidance and continuous support. I had arrived at Cambridge knowing little about formal syntax but now can understand the most advanced works in the field. My academic growth is the result of their scholarly enlightenment, and I cannot thank them enough for this intangible gift. Ian gave me ample freedom since day one, allowing me to explore my interests at my own pace but also pointing me in the right directions, which were not necessarily in his own research areas. He often knew what I would be interested in even before I knew it myself, which in my eyes was like magic. I will never cease admiring his erudition and achievements.

I have always admired Theresa's knowledge and enthusiasm in linguistics too, so when she offered to be my co-supervisor at the end of my first year, I felt really lucky. I am deeply indebted to her for her detailed feedback on my work, her encouragement when I accidentally stepped into a whole new field, her practical advice on various (semi)academic matters, and her always timely help with my paperwork. Being an international student often means extra difficulty, and her administrative aid has made my life *much* easier. Both Ian and Theresa have hugely influenced my way of thinking, as can be seen in multiple places in this dissertation, but what I especially appreciate is that they have also helped me produce a piece of work that I can confidently call my own. Their open-mindedness to and protection of nonmainstream ideas is something I will pass on to my own students in the future.

I have been fortunate enough to work or converse with a lot of experts in and beyond the linguistics community at Cambridge and elsewhere, which has been an important source of inspiration for me to explore further realms in the intellectual world. Many thanks to Víctor Acedo-Matellán, who had been a great mentor-advisor while being here and continued helping me after moving to The Other Place; to Andrew Pitts and his Category Theory course in the Department of Computer Science and Technology, which was the first math course I had seriously taken years after high school, as well as John Baez and his Applied Category Theory course in the Azimuth forum, where my layman questions were patiently answered; to Anders Holmberg, Igor Arrieta, Erin Pretorius, Cherry Lam, and others who kindly commented on my written work; to Jenneke van der Wal, Michelle Sheehan, Craig Sailor, and other SyntaxLab goers who gave me feedback at various points of my research; to Jim Baker, András Bárány, Jamie Douglas, Anna

Hollingsworth, Daniela Gerz, Hanna Ajer, Li Nguyen, and other PhD colleagues with whom I have had thought-provoking chats; and also to the many friendly people I have luckily had the chance to meet at various conferences—I look forward to seeing you again!

I have been able to fully concentrate on my PhD research thanks to the generous funding from Cambridge Trust and China Scholarship Council. I should particularly thank Cambridge Trust for extending my scholarship into my fourth year, without which I would not have been able to write up this dissertation. I would also like to thank Gonville & Caius College for giving me financial help when I was in need. In addition, both Gonville & Caius College and the Faculty of Modern and Medieval Languages have offered me student grants, which have allowed me to present my work at conferences and attend language courses.

Life is not all about linguistics, and I should like to extend my thanks to all my dear friends who have made my nonacademic life at Cambridge enjoyable. Among those I hang out with, Wiktor Olszowy is the most reliable one; sometimes he plays loud music at midnight but at least his taste in music is fine. Beñat Mencia Uranga is unlike Wiktor but has the funniest style (despite what Wiktor says in his PhD Acknowledgments) and arranges the best meal groups (with a lot of sparkling conversations). The Caius Movie Appreciation Society, the beginning of which I had witnessed, is probably the best Cambridge society I have been to. There I have watched (or occasionally napped through) many movies that I would probably never watch elsewhere and met a bunch of lovely people, including but not limited to Michael Kosicki (who makes good pie), Michele Sanguanini (who makes good tiramisu), Alex Shattock (who has the best party costumes), Christine Yu (who is the most energetic), and Daniela Peris (who gave me her *Beedle the Bard*). I would also like to thank all my old friends, in China and elsewhere, who have shared my joy and heard my complaints, given me advice and helped me grow, and never got annoyed by my monthslong silence. Mengmi Lü, Yuning Chen, Ruikang Zhang, and Xiaowen Xiong: I wish you all the best in your future life and career!

Finally and most importantly, I want to thank my parents and the rest of my family, who have always been there for me whether I am home or abroad, happy or sad, healthy or sick, and I am forever grateful for their love, understanding, and support. I also feel indebted to my bike, which saved me from a car accident by sacrificing itself, and to my laptop, which never acted up during my writing marathon. I solemnly swear that the above acknowledgments are all sincere.

Contents

List of figures xv

List of tables xvii

Abbreviations xix

1 Introduction 1

2 Foundational issues: Features and categories 5

2.1 Introduction 5

2.2 The featural metalanguage 8

2.2.1 Classification 9

2.2.1.1 Valency-based 9

2.2.1.2 Pollency-based 11

2.2.2 Value and notation 12

2.2.3 Ontological organization 15

2.2.3.1 Feature typing 15

2.2.3.2 Feature classing 19

2.2.4 Granularity 21

2.2.5 Interim summary 23

2.3 Syntactic category 23

2.3.1 Definition 24

2.3.1.1 Taxonomic vs. combinatorial perspectives 24

2.3.1.2 The mergeme 25

2.3.2 Template 30
2.3.2.1 Granularity levels 30
2.3.2.2 More about the mergeme 32
2.4 Flexibility 34
2.4.1 Individual level 35
2.4.1.1 First abstraction layer 35
2.4.1.2 Second abstraction layer 37
2.4.2 Global level: Third and fourth abstraction layers 39
2.4.3 Supraglobal level: Fifth abstraction layer 40

3 Defective category and feature-based adjunction 43
3.1 Defective category theory 43
3.1.1 Derivational schema 44
3.1.2 Approaches to adjunction 46
3.1.2.1 Axiomatic approach 46
3.1.2.2 Derivational approach 48
3.1.2.3 Lexicalist approach 52
3.1.2.4 Interim summary 54
3.2 Modifier-head compounds 55
3.2.1 Derivation 56
3.2.2 Composition 58
3.2.2.1 Possibility I: Roots denote type-generic predicates 59
3.2.2.2 Possibility II: Roots denote type-open predicates 62
3.2.2.3 Possibility III: Roots lack model-theoretic denotations 63
3.2.2.4 Interim summary 65
3.2.3 Implications 66
3.2.3.1 Modifier-head compounding is universal 66
3.2.3.2 Modifier-head compounding has no intrinsic categorial bias 67
3.2.3.3 Modifier-head compounding is parameterizable 70
3.2.3.4 Modifier-head compounding blocks head movement 72

3.2.4 Remaining questions 80
3.3 Sentence-final particles 82
3.3.1 Introduction 82
3.3.2 Similarities with modifier-head compounds 85
3.3.3 A defective-category-based analysis 91
3.3.4 Further implications 95
3.4 Chapter summary 97

4 Root category and semifunctionality 99
4.1 Generalized root syntax 99
4.1.1 Root-supported head 99
4.1.2 Layered derivation 104
4.1.3 Implications 106
4.1.3.1 Vocabulary item typology 106
4.1.3.2 PF-oriented head movement 107
4.2 Semifunctional items in Chinese 109
4.2.1 Verbal/sentential domain 109
4.2.1.1 “Light verbs” 110
4.2.1.2 Voice elements 112
4.2.1.3 Aspectual elements 116
4.2.1.4 Modal elements 118
4.2.1.5 Sentence-final particles 121
4.2.2 Nominal domain 123
4.2.2.1 Classifiers 123
4.2.2.2 Pronouns 126
4.2.3 Prepositional domain 129
4.2.4 Acategorial domain 131
4.2.5 Interim summary 134
4.3 Further implications 134
4.3.1 Root support and analyticity 134

4.3.1.1 Deducing the analyticity typology 134
4.3.1.2 Criterion of analyticity 136
4.3.1.3 Is there an analyticity parameter? 137
4.3.2 Root support and grammaticalization 140
4.3.2.1 Grammaticalization routes 140
4.3.2.2 Multifunctionality 141
4.4 Chapter summary 144

5 Incomparable categories and category flavoring 145
5.1 Climbing up the ladder of abstraction 145
5.1.1 The categorial sequence poset 146
5.1.2 Incomparable categories 147
5.1.3 Characteristics of incomparable categories 149
5.2 Category flavoring 151
5.2.1 Flavored little *x*s 151
5.2.2 Valued categories vs. flavored categories 153
5.2.3 Further implications 156
5.2.3.1 Split-X categories and the limit of cartography 156
5.2.3.2 Pseudo-antidecompositionality and the limit of flexibility . . 159
5.3 Revisiting Chinese resultative compounds 161
5.3.1 A neo-emergentist view on category flavoring 162
5.3.2 Previous studies 163
5.3.2.1 A note on glossing 163
5.3.2.2 The dummy verbalizer pitfall 164
5.3.2.3 Is the result slot adjectival? 166
5.3.2.4 Interim summary 172
5.3.3 The history of resultative compounds 173
5.3.4 Distinctness and stative verbalizer 178
5.3.5 Ramifications of a new categorizer 182
5.4 Chapter summary 187

6 Global interconnection: A Category-theoretic sketch 189

6.1 Ascending to the global level . . . 189

6.1.1 Cross-major-part-of-speech parallelism . . . 190

6.1.2 Inter-granularity-level relations . . . 192

6.2 Why Category theory? . . . 195

6.3 A Category theory primer . . . 197

6.3.1 Category . . . 197

6.3.2 Functor and natural transformation . . . 202

6.3.3 Adjunction and Category comparison . . . 204

6.3.4 Yoneda embedding . . . 210

6.4 Cross-major-part-of-speech parallelism . . . 211

6.4.1 No direct connection . . . 211

6.4.2 Connection via universal spine . . . 215

6.4.2.1 Epi-Adjunction . . . 215

6.4.2.2 Flower-shaped parallelism . . . 218

6.5 Inter-granularity-level relations . . . 221

6.5.1 Isomorphic granularity levels . . . 221

6.5.2 Augmented granularity level isomorphism . . . 228

6.5.3 Multiple universal spines . . . 230

6.6 Revisiting categorial sequence indexing . . . 232

6.7 Chapter summary . . . 235

7 Conclusion 237

Appendix A Examples of modifier-head compound verbs in current English 243

Appendix B Historical Chinese documents 245

Appendix C Order-theoretic notions 247

Appendix D Size-related notions in Category theory 249

Appendix E Special Categories and Functors **251**

Bibliography **253**

List of figures

2.1 Subclassification of nonprivative features 10

2.2 Boolean features in algorithm design 11

2.3 Sort-type hierarchies in syntax and phonology 17

2.4 Two proposed feature geometries 20

2.5 Three classes of syntactic-category-defining properties 25

2.6 Derivation in propositional logic and minimalist syntax 27

2.7 Classification of syntactic categories 37

2.8 A c-sequence with two flavored categories 39

3.1 Derivational schema for the defective category 44

3.2 Modifier-head compounding as a function 56

3.3 Derivational path of *blackboard* 56

3.4 Derivational path of Mandarin *nǐhǎo a* 'Hello!' 92

4.1 Derivational schema for root categorization 101

4.2 Root-containing configurations 104

4.3 Root-supported heads in proof tree format (cf. Figure 4.2b) 105

5.1 Three patterns of category subtyping 150

5.2 A vivid depiction of fissioning 153

5.3 Three levels of category subtyping in Chomskyan syntax 157

5.4 Split-*v* categories 157

5.5 Flavored Infls 158

5.6 Two DM analyses of Chinese resultative compounds 164

5.7 Establishment of the resultative compound construction 178
5.8 V-V coordination and V-V subordination 178
5.9 Two functional zones in Early Modern Chinese 185
5.10 Potential shapes for a bicategorizer c-sequence 186

6.1 Division of categorial space for v 192
6.2 Order-theoretic organizations of the GLS 194
6.3 Commutative diagrams for associativity and unit law 198
6.4 A real-life Category 199
6.5 Categorical parallelism across multiple MPOSs 220
6.6 The poset Category **GLS** of granularity levels ordered by inheritance 221
6.7 The Category $\mathbf{Cat}_{\mathrm{GLS}}$ of granularity-level Categories 224
6.8 $\mathbf{Cat}_{\mathrm{GLS}}$ extended with incomparable granularity levels 224
6.9 A Categorical skeleton of $\mathbf{Cat}_{\mathrm{GLS}}$ 225
6.10 Biberauer & Roberts' (2015) syntactic category division hierarchy 226
6.11 A corner in $\mathbf{Cat}_{\mathrm{GLS}}$ (cf. Figure 6.8) 226
6.12 A corner in $\mathbf{Cat}_{\mathrm{GLS}}$ (complete) 227
6.13 An inheritance hierarchy for universal spines 231
6.14 Category division hierarchy overlaid with inheritance Functors 231

List of tables

2.1 Templates and flexibility types in the syntactic category system 41

3.1 Comparison of Epstein et al. (2012) and Oseki (2015) 51

3.2 Three approaches to adjunction in Chomskyan syntax 55

3.3 Three possibilities of root-categorizer composition 66

3.4 Verbal modifier-head compounds in English 69

3.5 Verbal modifier-head compounds in Chomskyan syntax 70

3.6 Translations of English compound verbs in French and Spanish 75

3.7 Correlation between verb movement and modifier-head verb compounding . . 80

3.8 Decomposition of Chinese sentence-final particles 84

3.9 Basic sentence-final particles in Mandarin and their functions 85

3.10 Etymologies of basic consonant-based sentence-final particles in Mandarin . . 85

4.1 Cross-classification of vocabulary items 106

4.2 Overt DO light verbs in Mandarin 110

4.3 Voice auxiliaries in Mandarin 113

4.4 Aspectual elements in Mandarin 116

4.5 Modal auxiliaries in Mandarin 119

4.6 Classifiers in Mandarin 124

4.7 Pronouns in Mandarin 126

4.8 Prepositions in Mandarin 132

4.9 Conjunctions in Mandarin 133

4.10 Huang's (2015) analytic properties 137

4.11 Grammaticalization of Mandarin semifunctional items 142
4.12 Grammaticalization of Mandarin multifunctional items 142

5.1 Templates and flexibility types in the syntactic category system (= Table 2.1) . . 145
5.2 An expanded classification of result slot items 169
5.3 Readings of result slot items 172
5.4 Periodization of Chinese 173
5.5 Historical development of the resultative compound construction 174

6.1 Templates and flexibility types in the syntactic category system (= Table 2.1) . . 189

7.1 Templates and flexibility types in the syntactic category system (= Table 2.1) . . 241

Abbreviations

Arch	archicategory
ATT	attribute
BCC	Borer-Chomsky conjecture
○	undivided categorial space
Cart	cartography
CAT	categorial
Cat	defective category
CCS	conceptual category system
CFC	core functional category
CL	classifier
CoL	change-of-location
COMP	completive
Comp	complement
CONT	continuative form
COP	copula
CoS	change-of-state
CRS	currently relevant state
DISP	disposal
DM	distributed morphology
EMC	Early Modern Chinese

EP	extended projection
EXP	experiential
FE	feature economy
FTC	feature-type correspondence
GB	government and binding
GLS	granularity level space
GPSG	generalized phrase structure grammar
GRAN	granularity
GRS	generalized root syntax
HPSG	head-driven phrase structure grammar
IG	input generalization
LA	lexical array
LI	lexical item
LK	linking element
LV	light verb
MHC	modifier-head compound
MidC	Middle Chinese
MMM	maximize minimal means
ModC	Modern Chinese
MODI	modifier marker
MPOS	major part of speech
m-root	morphological root
NM	nonmergeme
NML	nominalizing element
OC	Old Chinese

OCP	obligatory contour principle
PAD	pseudo-antidecompositionality/pseudo-antidecompositional
PERF	perfective
Ph	phase
PIC	phase impenetrability condition
PLD	primary linguistic data
POS	position
PROG	progressive
Q	quantifier
RSI	result slot item
SCDP	syntactic-category-defining property
SCS	syntactic category system
SDA	successive division algorithm
SFP	sentence-final particle
SFTC	simplified feature-type correspondence
SO	syntactic object
s-root	syntactic root
TAM	tense-aspect-mood
TAX	taxonomic
UG	universal grammar
VAL	value
XS	exoskeletal syntax

Chapter 1

Introduction

Categories are the cornerstone of human language. It is fair to say that there would be no language if the human mind were unable to form and manipulate categories, for each phoneme or concept entails a particular mental category, and so does each sound-meaning pair. Cohen & Lefebvre begin their coedited *Handbook of Categorization in Cognitive Science* with the comment below:

> Categorization is the mental operation by which the brain classifies objects and events. This operation is the basis for the construction of our knowledge of the world. It is the most basic phenomenon of cognition, and consequently the most fundamental problem of cognitive science. (Cohen & Lefebvre 2005: 2)

In a similar vein, Ellis (1993) asserts in the preface of *Language, Thought, and Logic* that "categorization (not syntax) is the most fundamental aspect and process of language" and that "neither anything else in language nor indeed its purpose can be properly understood until the nature of categorization has been grasped." The importance of categories and categorization is a matter of general consensus among cognitive scientists, linguists, philosophers, and possibly also experts from other fields who are fascinated by language, a spectacular natural phenomenon and the hallmark of intelligence.

Against this backdrop, this dissertation investigates certain aspects of categories in natural language syntax. Unlike Ellis, I will not attempt to convince my readers that categories are more fundamental than syntax in allowing human language to exist: categories do antecede syntax, but they also antecede language as we know it. As Smith et al. (2016: 1) point out, "categorization has conferred fitness advantages on vertebrates for hundreds of millions of years."[1] Furthermore, if Berwick & Chomsky (2016) are right in suggesting that the most significant event in the evolution of language is the genesis of Merge (i.e., the basic combinatorial operation for hierarchical structure building), then syntax in its minimalist conception defines language in ways that categories cannot. As Berwick concludes in the final passage of *The Biolinguistic Enterprise*:

[1] See references cited in Smith et al. (2016) for evidence that nonhuman animals do possess conceptual categories.

> Chimpanzees, and possibly other primates, got conceptual atoms. Birds got rhythm. But only people combined both, getting undeletable features, *and* Merge. And from this, came language: Merge is all you need. (Berwick 2011: 491)

Berwick's mentioning of "undeletable feature" (i.e., Chomsky's 2007 edge feature) is interesting in the current context. We know that Merge operates on syntactic categories (on the simplest assumption it just puts them into sets, which are themselves again categories); but if unbounded Merge relies on a particular type of feature (i.e. edge feature) as Berwick and Chomsky suggest, and if features are properties of categories, then in a sense categories are not only the operands of Merge but also an indispensable part of (or supplement to) its definition. In other words, *Merge* and *syntactic category* define each other: what can merge is a syntactic category and what is a syntactic category can merge. Hence, Ellis' assertion is not necessarily an overstatement after all.

In the upcoming chapters I study the nature of syntactic categories from a particular angle: the formal flexibility in their formation. My inquiry is motivated by two big-picture questions:

1. What does an abstract ontology of syntactic categories look like?

2. To what extent is syntactic category formation instructed by UG?

To state my answers up front, I hold that the syntactic category ontology is organized as an information web with interlocking layers of abstraction. The bits and bytes in this information web are features and categories, and their layered organization presents us with a microscopic and a bird's-eye view simultaneously. Specifically, this ontology is so well-connected that whichever corner in it we start from, we can easily reach any other point in it following a finite number of steps (conceived as functional mappings). I assume that the topology of this network is part of UG, in the form of a small number of highly general templates, each at an abstraction layer. On the other hand, there are also aspects in the ontology that UG cares less about, and it is in those places that what I call "formal flexibility of syntactic categories" arise. In other words, the term *formal flexibility* in my title does not refer to language-specific categorial types or crosslinguistic categorial variation; rather, it refers to the potential shapes the abstract notion termed *syntactic category* can take. As my inquiry discovers, very important aspects of human language grammar arise not under the instruction of UG, but out of its disregard.

The main goal of this dissertation is a highly conceptual one. It to some extent demonstrates that we can study the nature of categories without inspecting any (language-)specific category. This situation incidentally converges with that in a branch of mathematics—Category[2] theory—which I will incorporate into one of my chapters (Chapter 6). That said, since linguistics is still largely an empirical science, I will also endeavor to map my conceptual predictions onto (cross)linguistic facts. An overall picture revealed by this endeavor is that predictions made based on the abstract nature of categories tend to be big and quotidian—sometimes so quotidian

[2] I capitalize this mathematical term to distinguish it from the homonymous linguistic term.

that linguists seldom view them as "predictions" (e.g., a major implication in my Chapter 3 is that adjunction is universal). Nevertheless, I deem it a step forward to be able to reduce such "irreducible" properties of human language to more basic terms.

Following this general introduction, I begin my inquiry with an examination of the two foundational notions of this dissertation—feature and category—in Chapter 2, by the end of which the basic scaffolding of my working ontology will be in place. A more detailed chapter-by-chapter outline will also be provided there after the necessary terminology is introduced. In Chapters 2–6, I take closer looks at the above-mentioned abstraction layers one by one from the bottom up, specifying the UG template and formal flexibility at each layer. Chapters 2 and 6 are purely conceptual, while Chapters 3–5 will be accompanied by empirical case studies. Finally, I summarize the main results of my study in Chapter 7, together with a few concluding remarks.

Chapter 2

Foundational issues: Features and categories

2.1 Introduction

This dissertation is about syntactic categories, so a natural first question is, What are they? There may be a variety of answers to this question based on different assumptions and suited to different purposes, but I want to begin with a most generic perspective, one that puts syntactic categories back in the universe of categories (linguistic or not). Setting out with such a bird's-eye view will help us keep the big picture in mind as our discussion moves on to more specific domains, so the first slogan of this chapter is:

Slogan 1. Syntactic categories should conform to the general definition for categories; that is, any category from any domain.

According to Harnad (2005: 28, 40), there are two ways to define a general category: (i) (extensional) by explicitly listing its members; (ii) (intensional) by specifying a set of invariant properties, or features, as the necessary and sufficient conditions for its application. For example, the extension of the category Blue is the set of all blue objects in the universe of discourse, while its intension is the property of being blue (Carnap 1956: 23), or in featural terms, [BLUE].[1] The intensional definition can be converted to an extensional one by set comprehension: $\{x \mid \text{Blue}(x)\}$.[2] Between the two definitions, the intensional one (known as the classical theory of categorization) has been predominant in a number of disciplines including linguistics (Cohen & Lefebvre 2005). Following it we can view a syntactic category simply as a set of features, which together determine a set of individuals in the universe of syntactic objects (SOs).

The above perspective is essentially a taxonomic one, which presupposes that categories ex-

[1] Throughout this dissertation I notate category labels with capital initial letters (e.g., Voice, Asp) or by acronyms (e.g., SFP) and write features in small capitals (e.g., EPP) except φ and the placeholder F. I follow the common practice of enclosing (possibly singleton) feature sets in square brackets; e.g., $[F_1, F_2, F_3]$.

[2] A feature term like BLUE is usually taken to denote a first-order predicate, though as we will see in §2.2, there are still other semantic models to interpret features.

ist for the purpose of classifying objects; this is also the everyday sense of the word *category*. Notions about syntactic categories from the taxonomic perspective include nominal/verbal category, lexical/functional category, and so on. However, syntactic categories arguably also have some unique characteristics that distinguish them from just any category (e.g., Chicken). Most importantly for my purpose, they also have a combinatorial side, whereby (taxonomized) SOs can be combined into larger SOs (e.g., $\textit{the}_{\mathrm{D}}$ + [$\textit{blue}_{\mathrm{A}}$ + $\textit{T-shirt}_{\mathrm{N}}$] = $\textit{the}_{\mathrm{D}}$ + *blue* $\textit{T-shirt}_{\mathrm{NP}}$ = *the blue* $\textit{T-shirt}_{\mathrm{DP}}$)[3] or permuted to render extra interpretive effects (e.g., *I saw* $\textit{John}_{\mathrm{TP}}$ → $\textit{John}_{\mathrm{DP}}$ + *I saw* $\cancel{\textit{John}}_{\mathrm{TP}}$ = *John, I* $\textit{saw}_{\mathrm{FocP}}$).[4] Such combinatorial possibilities are an unusual trait in the universe of categories and may even be unique to human language (Berwick & Chomsky 2016); therefore, they should receive at least as much attention as the taxonomic perspective in an adequate theory of syntactic categories. Notions about syntactic categories from the combinatorial perspective include head/phrasal category, minimal/maximal category, probe/goal, and so on. Unlike taxonomic categories, whether or not combinatorial categories should also be defined by features is less clear (and less often addressed). In the GB/minimalism tradition, the combinatorial aspect of syntactic categories is often directly encoded in phrase structuring—as in labels like V^0, N′, TP, $\mathrm{D}^{\mathrm{min}}$—or indirectly via other (noncombinatorial) features; for example, a probe category is a category that bears an unvalued feature (which is itself taxonomically oriented), and there is generally no such feature as [PHRASE] or [PROBE]. That said, we do occasionally see combinatorially oriented features, such as Chomsky's (2007) "edge feature,"[5] which encodes the fundamental combinatorial capacity of SOs; the GPSG bar-level feature [BAR: 0/1/2], which encodes the GB projection levels X/X′/XP (Gazdar et al. 1985: 25); and perhaps also selection features in general.[6] The above two perspectives on syntactic categories (i.e., taxonomic and combinatorial) will be further discussed and eventually unified in §2.3.

There are ongoing debates over the definition of syntactic categories, especially on the taxonomic side. While the feature-based approach is generally accepted, not so much agreement has been reached on further questions like what or how many features should be used to define a syntactic category. Take the familiar lexical categories for example. Chomsky (1981: 48) uses two cross-classifying bivalent features [±N] and [±V], respectively interpreted as "substantive" and "predicate," to define four categories (N, V, A, P); Baker (2003: 21) uses two possibly underspecified bivalent features [±N] ("has a referential index") and [±V] ("has a specifier") to define three categories (N, V, A); Panagiotidis (2015: 174) uses two privative features [N] ("sortal") and [V] ("temporal/sub-eventive"), which are themselves values of another feature [PERSPECTIVE], to

[3] In Lambek's (2008) words, syntactic categories have an algebraic nature.

[4] These are just external and internal Merge in minimalist terms.

[5] Follow-up discussions on this feature are sparse and disparate; see, e.g., Boeckx (2011) and Adger & Svenonius (2011) for quite different ideas. As we will see in §2.3, Chomsky's insight has a natural place in the approach to syntactic categories I pursue. Besides, the term *edge feature* has two separate senses in Chomsky's work: Chomsky (2005: 18) uses it as a synonym of EPP feature, while Chomsky (2007: 11) uses it to indicate that an element can be merged. Unless otherwise clarified, when mentioning edge feature I mean its second sense (i.e., mergeability).

[6] I will leave selection features aside due to their uncertain (and disputed) theoretical status (see §2.3.1 for an additional comment). This restriction of attention does not affect the development of my thesis.

define two categories (N, V);[7] and Biberauer & Roberts (2015: 299) use a single bivalent feature [±V][8] with no fixed interpretation to define two broad categories N and V—and the list of alternative proposals continues. By comparison, there is less definitional controversy over functional categories, probably because their postulation is usually based on a priori theoretical consideration; however, things are no more settled if only one asks. To illustrate, cartography (Rizzi 1997, Cinque 1999) and its descendants (especially nanosyntax; Caha 2009, Starke 2009, 2011) ground numerous functional categories on atomic features with semantic content (e.g., Asp = [+ASP], Top = [+TOP]), while Biberauer & Roberts (2015) define functional categories by an increasing number of decreasingly scoped contrastive features (e.g., C = [+V, +C], T = [+V, +C, +T]) that may or may not have semantic content.[9]

What is revealed from the above comparison is not only heterogeneous theorization but also highly nonunanimous terminology; for instance, the symbols N and V are given different interpretations by different authors, which means that they as ground terms are ambiguous and overloaded. Equally ambiguous are fundamental terms like *feature* and *value* and quasi-mathematical symbols like +. While theoretical debates are a normal part of scientific research, terminological ambiguity can be detrimental. For one thing, no field of inquiry can become a "very definite science" when its terminology and concepts are badly defined (Lundberg 1929: 59); for another, neglect of terminological problems may lead to more substantive trouble.[10] With this in mind, my aim in this chapter is not to end any debate or make any conclusive decision but to carefully characterize the concepts and terms I will rely on to reason about syntactic categories. By doing so I hope to enhance the precision of my theory and lay the groundwork for later chapters.

After this introduction, §2.2 is a self-contained examination of the featural metalanguage in linguistic theories. It is not directly related to the empirical chapters but included for foundational purposes, as it gives me the opportunity to define terms like *nonequipollent feature*, *attributive value*, *sort-type*, and *granularity level* collectively in a coherent narrative rather than one by one in isolation. Next, §2.3 spells out my approach to syntactic categories, and §2.4 describes the categorial universe I work with, which in a nutshell is a multilayered web organizing individual categories into sequences, sets of sequences, and so forth. Each layer of organization (which I will call an *abstraction layer*) has its own UG-given template, and where UG does not care shows considerable yet principled flexibility. Exploring the template-flexibility pairs in the *syntactic category system* (SCS) is the motivation for and theme of this dissertation. To demonstrate, the template at the abstraction layer of single categories is given in (1).

(1) syntactic category$_{\text{LANG, GRAN}}$:= [⟨CAT: `cat`, POS: `nat`⟩, NONMERGEME FEATURES]

[7] Note the blurred feature-value boundary, which will be discussed in §2.2.2.

[8] Douglas (2018: 28) calls this an "archifeature" by analogy with Prague school archiphonemes; see also §2.2.1.2.

[9] Biberauer & Roberts' approach is similar to Dresher's (2009) contrastivist theory on phonological categories and opens up new possibilities in understanding acquisition and change.

[10] A well-known example is in the history of set theory, where loosely defined terms like *set* and *predicate* eventually led to "crippling logical paradoxes" (Shao 2012: 1–2).

For instance, the category T in a specific variety of English (call it ENG$^\flat$) at the granularity level called CFC—which is characterized by the core functional categories in the sense of Chomsky (2000)—can be defined as in (2).

(2) $\mathrm{T}_{\mathrm{ENG}^\flat,\,\mathrm{CFC}} \coloneqq [\langle \textsc{cat}\text{: } \mathbf{v}, \textsc{pos}\text{: } 3\rangle, \textsc{nom}, \textsc{epp}, \ldots]$

The ⟨CAT: `cat`, POS: `nat`⟩ part of the template in (1)—which I will call the *mergeme*—is an ordered pair specifying the category's scope-based[11] first-Merge position (which is either unique or none). It largely corresponds to the conventional "categorial feature" but differs in important ways that will be discussed in §2.3; I will generally not use "categorial feature" in my formal statements to avoid the accumulated ambiguity it carries and the heterogeneous presuppositions it may invoke. On the other hand, labels like `cat` and `nat` (in typewriter font) are sort-types for attributive feature values, and subscripts like ENG$^\flat$ and CFC indicate the language and granularity level the given category is part of. So, the definition in (2) can be read as "the scope-based category T in the English variety ENG$^\flat$ at the CFC granularity level is the third **v**-sequence category,[12] which also bears (among others) a nominative case feature and an EPP feature." Definitions for all terms used in this passage can be found in §2.2. Finally, the nonmergeme part of (1) accommodates other features that may help define a syntactic category. Since these are not my focus, I will often just use a NONMERGEME (NM) flag or ellipsis points to indicate their existence. I will also leave out the language or granularity level tags when the relevant information is clear from the immediate context.

2.2 The featural metalanguage

Features are the "standard currency" in linguistics (Corbett 2010: 17), yet their usage is far from being standardized, partly because different theories have different descriptive needs which lead to different conventions, and partly because there is much ambiguity and little uniformity in the terminology being used. Svenonius (2019: 8) mentions that for feature systems developed as descriptive tools "arbitrary decisions can be made about their formal properties." The overall justifiability of this comment left aside, arbitrariness does not equal vagueness; on the contrary, clarity is of utmost importance in the use of features (Corbett 2012: 31). So, as feature users we should aim for clear definitions and explicit declarations whenever possible.

In this section I examine some key aspects of existing feature systems, including their classification (§2.2.1), notation (§2.2.2), and organization (§2.2.3); I finish with a remark on granularity (§2.2.4). By the end of this section, there will emerge a four-component semiformal metalanguage, which may help us characterize and compare feature systems along four dimensions:

[11] By *scope-based* I mean a combinatorial scope expressible by derivational relations like c-command. This is different from (and orthogonal to) the contrastive scope of taxonomic features in Dresher (2009).

[12] *Third* is just a shorthand for a more precise and nonnumerical representation (see §2.3).

(3) a. **Specification:** What specification states does the feature system make available? Do they all have equal theoretical status?

b. **Valuation:** Does the feature system reference values? If yes, how are they conceived?

c. **Typing:** Does the feature system have a type(-like) discipline? If yes, in what sense and to what extent?

d. **Granularity:** How fine-grained an analytical level is the feature system designed for? Can it accommodate other granularity levels too?

The feature system I will adopt in §2.3 makes available multivalent and bivalent specifications as well as two theoretically special incomplete specification states; it references two value formats, attributive and coefficient, respectively for multivalent and bivalent features; it explicitly types feature specifications into logico-mathematical orders and taxonomic sorts; and it directly incorporates granularity as part of the featural description. Considering the standard currency status of features, this section is not limited to the GB/minimalism perspective but gathers insights from multiple sources, including phonology, unification-based grammars, and so on. These apparently disparate feature systems subtly converge and together fit into a more complete picture.

2.2.1 Classification

The first key aspect of a feature system is its classification based on the number and nature of specification states it makes available for each feature. A *specification state* of a feature is a form it may occur in. The number of specification states corresponds to a *valency-based* classification, and the nature of specification states corresponds to a *pollency-based* classification.

2.2.1.1 Valency-based

Three classes of features are generally distinguished based on their valency (i.e., value set size), as in (4). I defer a formal definition of *value* to §2.2.2 and here simply treat a value as an extra piece of information attached to a base specification (so each value defines a specification state).[13]

(4) a. *Privative* features have no value-defined specification state.

b. *Bivalent* features have exactly two value-defined specification states.

c. *Multivalent* features have more than two value-defined specification states.

For example, the feature [BLUE] is privative if this is the only form it may occur in; namely, an object in the universe of discourse is either specified as [BLUE] or not; by comparison, [BLUE:

[13] The converse is not true; i.e., not every specification state is defined by a value. Failure to distinguish the two notions may lead to contradictory statements like "X is unvalued and has an empty set *value*" and "Y lacks a value and has as *value* lack-of-value." The italic *value* in both statements should be replaced by *specification state*.

YES/NO] is bivalent since it has two value-defined forms, and [BLUE: YES/NO/HALF/LIGHT/DARK] is multivalent since it has five. We can define privative, bivalent, and multivalent feature systems as systems consistently using privative, bivalent, or multivalent features, though in practice many feature systems have mixed valency. This is innocuous as long as features of different valency are used for separate purposes. Thus, in Adger & Svenonius' (2011) minimalist feature system grammatical features are multivalent (e.g., [PER: 1ST/2ND/3RD]) while categorial features are bivalent (e.g., [D: ±]). Similarly, Clements (2003) employs privative and bivalent phonological features respectively for articulatory places and manners (e.g., [LABIAL] and [±VOICED]).[14]

Nonprivative features may be further subclassified based on their value set pattern. Consider the following bivalent features (written in the Jakobsonian style[15] for expository convenience) for example: (i) [BLUE/NONBLUE], (ii) [BLACK/WHITE], (iii) [GRAY/SILVER], (iv) [GRAY/BLUE], and (v) [GRAY/CIRCLE]. Among the five features, (i) encodes a boolean switch, (ii) is not boolean but merely encodes a bipolar opposition, (iii) is not bipolar but merely scale-based, (iv) is neither bipolar nor scale-based though still distinctive, and (v) is not even distinctive (for there can be gray circles) but merely distinct. The five patterns correspond to five subclasses of bivalence as structured in Figure 2.1 and exemplified in (5). (The subclassification of multivalence follows the same pattern except that it stops at the scale-based vs. non-scale-based level.)

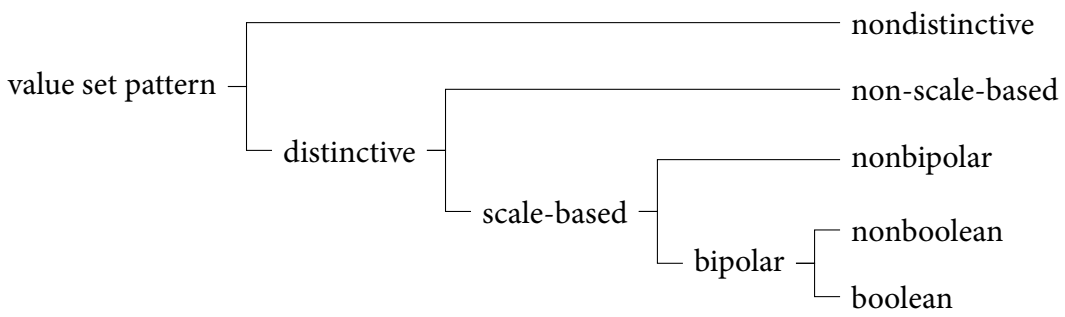


Figure 2.1 Subclassification of nonprivative features

(5) a. Jakobsonian "joint" features like [VOCALIC/CONSONANTAL] (Jakobson & Lotz 1949: 152) are nondistinctively bivalent; they are nondistinctive since their components may be joined (e.g., liquid phonemes are vocalic *and* consonantal).

b. Morphosyntactic features like [MASC/FEM] are usually distinctive but non-scale-based.

c. Continuously multivalent features are rarely used outside phonetically based phonology; an example is Flemming's (2004) [F_1: 1, F_2: 6, F_3: 3] (for IPA [i]).

d. Nonjoint phonological features like [NASAL/ORAL] are usually scale-based and bipolar.

[14] The mixed valency would be problematic (i.e., make the system inconsistent) if there were also multivalent categorial features in Adger & Svenonius' system or bivalent place features in Clements' system.

[15] In Jakobsonian phonology features are named by directly listing their values.

e. Bipolar features can often be rewritten in the boolean format; for example, [NASAL/ORAL] → [NASAL/NONNASAL].

Among the five subclasses, the nonboolean/boolean contrast is mostly one of analytical perspective. In some situations the boolean perspective has additional advantages; for example, in Dresher's (2009: 16) successive division algorithm (SDA) as illustrated in Figure 2.2, [NASAL] is only applied if [VOICED] is true; hence, the boolean values +/− (or 1/0 or TRUE/FALSE) serve a control flow purpose and efficiently establish a contrastive-scope-based ordering in the feature inventory (here [NASAL] < [VOICED]).[16]

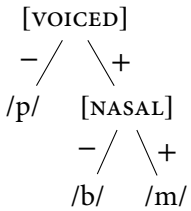


Figure 2.2 Boolean features in algorithm design

2.2.1.2 Pollency-based

A feature may have more specification states than those defined by its values. Any feature may have an *unspecified* state ([ ]), and nonprivative features may additionally have an *underspecified* state ([F:_]).[17] Such incomplete specification states have long been deployed in phonology to establish nonequipollent or more exactly markedness-based contrasts, such as the contrast between the presence and absence of a privative feature (à la Prague school) and that between the +/− and 0 values of a bivalent feature (Kiparsky 1982). Pollency-based feature classification essentially concerns the comparative theoretical status of a feature's specification states, and in this sense its application is not confined to phonology. In syntax, for example, Chomsky (2000, 2001) uses unvalued features to trigger Agree, and GPSG/HPSG recognize the empty set as a possible feature structure, called an "archicategory" (a Praguian term) in Gazdar et al. (1985: 26) and a "top"[18] feature structure that "subsumes all other feature structures" in Pollard & Sag (1987: 30, 37). Such a value-less feature structure may seem trivial, but from another perspective being empty also means being free in unification. Underspecification-induced combinatorial freedom is precisely what underlies the defective category theory to be presented in Chapter 3.

Since the unvalued features in minimalism and the empty feature structure in GPSG/HPSG are theoretically special, the feature systems in both frameworks are nonequipollent, though that

[16] As Theresa Biberauer (p.c.) points out, the boolean format also allows us to do more with a single feature; e.g., both NASAL and ORAL may be covered by manipulating NASAL.

[17] Neither unspecified nor underspecified is the same as a negative value. Adger (2010: 192) recognizes "lack of value" as an "empty set value," and Harbour (2013) similarly argues that "not plus" is not equivalent to "not there."

[18] In the sense that the empty set is the top element in Pollard & Sag's (1987) feature structure semilattice.

has nothing to do with markedness. In fact, to the extent that incomplete specifications are used, most existing feature systems are nonequipollent at the system level, and equipollence only exists in special, restricted contexts; for example, in Chomsky's [±N, ±V]-based definition of lexical categories, since both features must always be valued and all value combinations are exploited,[19] no specification state is more special than others and therefore this tiny subsystem is equipollent.

2.2.2 Value and notation

The second key aspect of a feature system concerns how it defines and represents values. The actual usage of feature values involves more complexity beyond the classification in §2.2.1.1. First, while values arguably are only defined for nonprivative features, in practice logically privative features are often given quasi-valued presentations too, perhaps due to the prevalence of value-oriented thinking. For example, the EPP feature is often introduced as being present on some categories and absent on others, but this contrast tends to be denoted by [±EPP].[20] This unfortunately overloads the ± symbolism, because in ordinary bivalent features like [±NASAL], ± represents a positive/negative (rather than present/absent) contrast.[21] Another example of quasi-valued features is the vacuous use of +, as in [+MOOD], [+TOPIC], and the like, where + merely "flag[s] something as a feature" in Adger & Svenonius' (2011: 51) words, as the relevant minus values are not in use (at least not in the same theoretical context). I call logically privative but notationally quasi-bivalent features like [±EPP] and [+MOOD] *bivalently presented privative features* to distinguish them from bona fide bivalent features like [±NASAL].[22]

Second, the feature-value boundary is often blurred due to different notational conventions and theoretical assumptions. For example, in Panagiotidis' (2015) approach to lexical categories, the categorial features [N] and [V] are privative on the one hand and values of another feature [PERSPECTIVE] on the other hand. Similarly, given two specifications like [+PST] and [TNS: PST], PST is usually called a "feature" in the former but a "value" in the latter. What is reflected in the terminological variation is the relative nature of feature and value, as stated in (6).

(6) In a bipartite feature specification with components a and b (formally an ordered pair $\langle a, b\rangle$), if the information in a is extended to a more specific state by the information in b, then call a a *feature* and b its *value*.

As such, the terms feature and value are only informative when a system's valency-based classification is specified; conversely, we may also deduce a system's valency setting by examining its

[19] Kay (1966: 22) describes this scenario as a "perfect paradigm," which are rare in naturally evolved symbol systems due to their "zero redundancy." Dresher (2009: 29) has a similar remark on the inadequacy of feature space diagrams in categorizing natural language phonemes.

[20] As a recent concrete example, Levin et al. (2018) propose two types of transitive v^0, one with EPP ([+EPP]) and the other without it ([−EPP]).

[21] Biberauer et al.'s (2014) reformulation of EPP as a movement diacritic (denoted by ^) is free from such ambiguity.

[22] Harbour (2011: 583–584) similarly calls a [±F] that encodes a single two-way distinction "plus∼minus privativity."

usage of these terms. So, strictly speaking, if Panagiotidis (2015) recognizes the relation between [PERSPECTIVE] and [N/V] as one of feature and value, then his system is not really privative but only locally privative, in contexts where [PERSPECTIVE] is not considered.[23]

Due to complications like bivalently presented privative feature and the often blurred feature-value boundary, the notion of value and its underlying logic deserve more careful investigation. The use of feature values in the literature is lax in both form and meaning. In form, [+PST] and [TNS: PST] represent two popular formats for feature specifications: [αF] and [F: α]. While the two formats are mutually convertible (e.g., [+PST] ↔ [PST: +]), they are preferred for different scenarios. Values typically written in the [αF] format include the boolean ±, the underspecified 0, and scale-based values like 5 in [5ROUND] (Clark et al. 2007: 384); I call these *coefficient values* (following Chomsky & Halle 1968: 65 inter alia). A key characteristic of coefficient values is that they cannot stand alone but must be attached to something else (e.g., /b/ = [+] is ill-formed because + lacks a host). On the other hand, the [F: α] format is typically used for attribute-value pairs; I call values in this format *attributive values*. Attributive values can stand alone, and in systems where attributes are conventionally hidden from featural descriptions, they effectively become the "features." Occasionally we also see quasi-coefficient values appearing in the [F: α] format, as in GPSG [AUX: ±] and [BAR: 0/1/2]. According to Bird (1995: 159), the values in such cases are abbreviations for fuller attributive forms; thus, [SPREAD: ±] = [SPREAD: ±SPREAD]. So, while the value slot in [F: α] is by design attributive, part of the attributive value may be coefficient. Besides, in systems that hide attributes, if the attributive values qua "features" are allowed to have a coefficient part, then that part becomes the "value" (e.g., + in [(NUM:)+SG], with the parenthesized attribute omitted); if no coefficient part is allowed, then the stand-alone attributive values become "privative features" (e.g., N/V in Panagiotidis' [(PERSPECTIVE:)N/V]).

In meaning, the denotation of a coefficient value α_{coef} (in an appropriate semantic model) maps its base's denotation d to a modified denotation d_α of the same type (whatever that type is).[24] For example, + and − respectively denote the identity and the negation function over their host's denotation (usually conceived as a first-order predicate), as in (7).

(7) a. $[\![+]\!]([\![\text{VOICED}]\!]) = [\![\text{VOICED}]\!] = \lambda x\,.\,\text{Voiced}(x)$

b. $[\![-]\!]([\![\text{VOICED}]\!]) = \neg[\![\text{VOICED}]\!] = \lambda x\,.\,\text{Nonvoiced}(x)$

So a coefficiently valued feature like [+VOICED] denotes a modified predicate, and a description like /b/ = [+VOICED] abbreviates the evaluated function $[\![\text{VOICED}]\!](/\text{b}/) = 1$ ("it is true that /b/ is voiced"). By comparison, attributive values usually just denote first-order predicates (e.g., $[\![(\text{NUM:})\text{SG}]\!] = \lambda x\,.\,\text{Sg}(x)$), which is why they can stand alone. Since stand-alone attributive val-

[23] This valency mixture is problematic in the sense of §2.2.1.1, as both [N]/[V] (privative) and [PERSPECTIVE: N/V] (multivalent) are used for the same (lexical-category-defining) purpose.

[24] More formally, a coefficient value denotes a *generic* (or *polymorphic*) *function* whose application is not dependent on argument type; this may be expressed in second-order typed lambda calculus ($\lambda 2$ in Barendregt's 1991 lambda cube) as $[\![\alpha_{\text{coef}}]\!] = \lambda\tau\!:\!*\,.\,\lambda d\!:\!\tau\,.\,d_\alpha\!:\!\Pi\tau\!:\!*\,.\,\tau \to \tau$ (read "a function mapping d of any type τ to d_α of type τ").

ues are just called "features," the first-order predicate interpretation of attributive values in effect blurs the boundary between features and (taxonomic) categories, as in (8), for the latter are also commonly interpreted as first-order predicates.[25]

(8) a. $[\![\text{VOCALIC}]\!] = \lambda x\,.\,\text{Vocalic}(x)$ (a feature)

b. $[\![\text{Vowel}]\!] = \lambda x\,.\,\text{Vowel}(x)$ (a category)

Note that the first-order predicate model for the featural metalanguage is popular but not unanimous. Take the +/− symbols for example. Apart from the predicate modifier interpretation, they are used in the literature in at least two other ways:

(9) a. Unary functions from predicate pairs to pair components, such as
$[\![+[\text{VOC/CONS}]]\!] = [\![+]\!](\langle[\![\text{VOC}]\!], [\![\text{CONS}]\!]\rangle) = [\![\text{VOC}]\!]$,
$[\![-[\text{VOC/CONS}]]\!] = [\![-]\!](\langle[\![\text{VOC}]\!], [\![\text{CONS}]\!]\rangle) = [\![\text{CONS}]\!]$,
$[\![\pm[\text{VOC/CONS}]]\!] = [\![\pm]\!](\langle[\![\text{VOC}]\!], [\![\text{CONS}]\!]\rangle) = [\![\text{VOC}]\!] \wedge [\![\text{CONS}]\!]$.
[in Jakobsonian phonology; see Jakobson & Lotz 1949 and Jakobson et al. 1952 inter alia]

b. Binary functions from lattice structures to lattice structures, such as
$[\![+\text{AUTHOR}(\pi)]\!] = [\![\text{AUTHOR}]\!] \oplus [\![\pi]\!] = \{a \sqcup b : a \in \mathscr{L}_{au}, b \in \mathscr{L}_{\pi}\}$,
$[\![-\text{AUTHOR}(\pi)]\!] = [\![\text{AUTHOR}]\!] \ominus [\![\pi]\!] = \{b \setminus max(\mathscr{L}_{au}) : b \in \mathscr{L}_{\pi}\}$.
[based on Harbour 2016: 75, for person features]

In (9a), the base specifications VOC and CONS still denote first-order predicates, but +/− no longer denote identity/negation functions; they denote the projection functions of ordered pairs instead (which map pairs to their components). On the other hand, (9b) is based on a completely different model, where base specifications do not denote predicates but denote lattices. Specifically, [AUTHOR] and $[\pi]$ respectively denote the "author lattice" ($\mathscr{L}_{au}$) and the "social ontology lattice" ($\mathscr{L}_{\pi}$), while + and − denote the lattice-theoretic operations "disjoint addition" and "cumulative subtraction." The combined symbol ± only has a principled interpretation in the Jakobsonian model; it does not occur in Harbour's model and is merely ornamental in the first-order predicate model (by flagging something as a feature).[26]

Corbett (2012: 1, 31) points out that superficially similar notations sometimes hide differences in the "underlying logic" and the "substantive semantics" of features, and that different conventions can lead to confusion "lurking behind the apparent formal tidiness." Exactly the same moral can be concluded from our discussion in this section. In the feature system adopted in this dissertation, I reserve the [αF] format for coefficient values and the [F: α] format for attributive values, and stick to the first-order predicate model unless otherwise declared.[27]

[25] The blurred feature-category boundary in turn opens the gate to further componential analyses for features (by treating them as categories), which is an atom-breaking process that pushes the analysis to a higher granularity.

[26] This flagging ± is common in Chomsky's work, as in [±TENSE], [±WH], [±ANAPHOR], etc. (Chomsky 1981, 1995).

[27] This does not mean that the first-order predicate model is necessarily better. I merely use it because it is familiar and suffices for the purpose of my study. A proper comparison and evaluation of different semantic models for

2.2.3 Ontological organization

The third key aspect of a feature system concerns whether and how it organizes features ontologically. In this section I investigate two feature-organizing methods, *feature typing* and *feature classing*, and conclude that organizational decisions are always made in feature-based theories, whether explicitly or implicitly. When value-oriented thinking and type/class-oriented thinking coexist, however, extra caution is needed to keep terminological accuracy and conceptual clarity.

2.2.3.1 Feature typing

The term *type* as in *type theory* or *type system* has developed various meanings in a number of interrelated subjects (e.g., logic, computer science, linguistics) since it first appeared in Russell (1903), not all of which have rigorous definitions or clear-cut boundaries.[28] Below I distinguish two senses of *type* relevant to linguistic feature systems; call them *order-type* and *sort-type*.

Order-type is the original sense of *type* used in mathematical logic, which is a tool to avoid set-theoretic paradoxes. Specifically, mathematical objects in a homogeneous universe of discourse are stratified into a hierarchy: individuals are assigned type 0, objects (e.g., classes, functions) built on individuals are assigned type 1, objects built on type-1 objects are assigned type 2, and so forth (Collins 2005: 1). As the name suggests, order-type is closely related to logical order (Gödel 1944: 135, Russell & Whitehead 1997: 51–52); for instance, first-order predicates are of type 1 since they are built on individual variables, second-order predicates are of type 2 since they are built on first-order-predicate variables, and so forth. In general, higher-order logic is considered a synonym for simple type theory (Farmer 2008: 268).

By comparison, sort-type is a later-developed sense of *type* via computer science and related fields like artificial intelligence, where types are used to classify computational constructs and thereby prevent execution errors (Cardelli 2004: 1) or to build ontologies for knowledge bases (Meinke & Tucker 1993). Simply put, a sort-type names a collection of values with shared properties, such as `int` for integers and `char` for characters (Mitchell 2003: 129). Sort-types do not live in a homogeneous universe but live in a heterogeneous one; they partition the universe of discourse into disjoint subuniverses called "sorts" or "types."[29] Sort-types can also be organized into hierarchies (especially in knowledge base applications), though these are fundamentally different from the order-type hierarchy; sort-type hierarchies are typically lattice-theoretic structures based on the supertype-subtype relation instead (i.e., set inclusion in the universe).

Order-types are relevant to feature systems because feature specifications qua terms in a met-

the featural metalanguage merits a project of its own (see Harbour 2016: §1.1 for an insightful remark).

[28] Ganesalingam (2013: 114): "Due to [its] complex history, 'type' has come to denote several distinct but related notions, which have not been clearly separated in the literature."

[29] The latter can be confusing when the context does not tell which sense of *type* is being used, for objects of different sort-types may have the same order-type; e.g., an integer (`int`) and a character (`char`) both have order-type 0, since they are both individuals in the universe of discourse (that the universe is heterogeneous does not matter).

alanguage are formally interpreted in some mathematical structure. Recall the three models illustrated in §2.2.2, (7)–(9), namely the first-order predicate model, the Jakobsonian model, and Harbour's lattice-theoretic model. All these models can be order-typed. Suppose we have a restricted universe of discourse consisting of all linguistic constructs[30] that may be featurally described, then the first-order predicate model defines at least two order-types: first-order predicates (type 1) for stand-alone specifications (e.g., privative features, attributive values) and predicate modifiers qua second-order functions (type 2) for coefficient values (e.g., +/−).[31] Similarly, we can define three order-types for the Jakobsonian model: first-order predicates (type 1) for atomic specifications (e.g., VOC, CONS), products of first-order predicates (type 2) for joint specifications (e.g., [VOC/CONS]), and projection functions (type 3) for +/− (which map products of first-order predicates to their components). Finally, Harbour's (2016) model requires a differently restricted universe, one where the individuals are various persons (e.g., speaker, hearer) rather than linguistic constructs (e.g., signs, phonemes); but order-typing proceeds in the same way, with at least two types distinguished: lattice-theoretic structures (type n) for person features and actions on those structures (type $n + 1$) for +/−.[32] In short, order-typing is a fundamental aspect of the featural metalanguage whichever model we adopt.[33]

Despite the fundamental role order-types play, however, they are seldom mentioned in linguistics practice, let alone systematically investigated. Instead, when *type* is referenced in feature systems, the intended sense is mostly (if not always) sort-type. This is most clearly reflected in HPSG and its offshoots (e.g., Shieber et al. 1983, Porter 1988, Copestake 2002), where *type* and *sort* are used interchangeably.[34] In HPSG types are employed to classify feature structures into ontological categories (Pollard & Sag 1994: 17), such as `cat` for categorial features (e.g., N, V, S) and `word` for word-level feature structures (e.g., [ORTH: dog, CAT: N]). This line of application is inspired by works in artificial intelligence. As Pollard & Sag (1987: 10) point out, their tools for working with feature structures are borrowed from "computational work on knowledge representation and data types"; in particular, their type discipline has been much influenced by Ait-Kaci's (1984) lattice-theoretic formalization of semantic networks (Pollard & Sag 1987: 50; see also Ait-Kaci 1984: 1). Note that the HPSG feature system, albeit standardly sort-typed, may also be order-typed. In fact, it just embellishes the first-order predicate model with sorts, and lin-

[30] More specifically, this subuniverse is restricted from the universe of all possible individuals.

[31] The relevance of second-order entities here is orthogonal to the debate about whether higher-order logic is appropriate for studying natural languages (Higginbotham 1998), as the featural metalanguage is not a natural language.

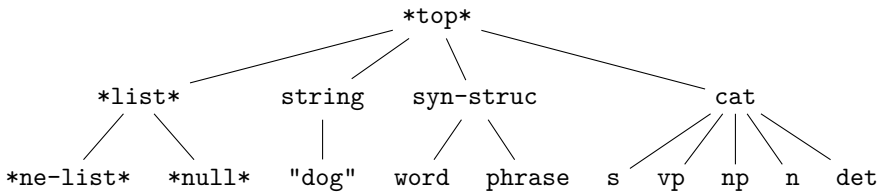
[32] I abstract away from the concrete value of n because there may be multiple intermediate order-types between the individuals and the lattices in Harbour's model.

[33] Adger & Svenonius (2011: 36) use *order* in a different way (see also Epstein et al. 2012: 265–266 and Svenonius 2019: 12). They distinguish "first-order" and "second-order" features, defining the latter as properties that "hold of tokens of features," such as the minus sign, syntactic diacritics (e.g., *u*, EPP), and valuation in general. Although they do not specify where this usage of *order* follows from, it is unlikely to be logically based, because in mathematical logic higher-order entities are built on lower-order ones (e.g., second-order predicates have first-order predicates as arguments), but the so-called second-order features are merely attached to (or specified for) first-order ones (e.g., in [*u*D], D is not part of *u*). Pending a clearer motivation for this usage of *order*, I remain agnostic about its place in the featural metalanguage.

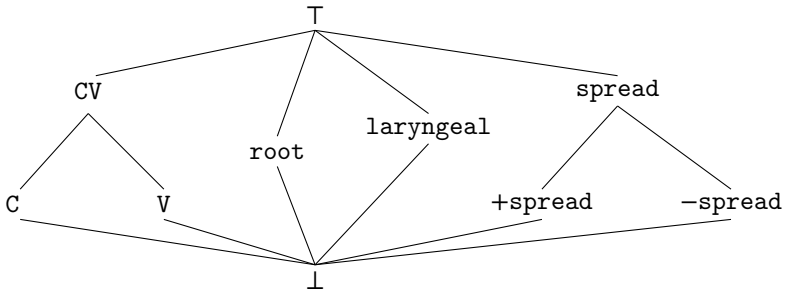
[34] Pollard & Sag (1987) systematically use "type," while Pollard & Sag (1994) mainly use "sort."

guistic constructs of different sort-types may have the same order-type (like the computational constructs in note 29, p. 15); for example, nouns, verbs, voiced phonemes, and spread phonemes are of four different sort-types (i.e., `n`, `v`, `voiced`, `spread`) but all have order-type 0, since they are all individual-level objects in the (heterogeneous) universe.

As aforementioned, sort-types can be put in supertype-subtype (or subsumption, symbolically $\sqsubseteq$) hierarchies. As such, they are very useful in organizing feature systems. In HPSG sort-type hierarchies are used for both syntax and phonology, as in Figure 2.3. (Such type hierarchies are often likened to biological taxonomies; see, e.g., Copestake 2002: 39–42.)



(a) Syntax (Copestake 2002: 38)[35]



(b) Phonology (Bird 1995: 56)

Figure 2.3 Sort-type hierarchies in syntax and phonology

Bird & Klein (1994: 548) give two methods to define a sort-type: (i) by appropriateness conditions (i.e., all and only the attributes the sort-type is specified for); (ii) by disjoint subtypes (i.e., based on the subsumption ordering over types).[36] Some sort-types can be declared by both methods, as in (10a), but some others (the atomic ones) can only be declared by subtype disjunction, as in (10b). The two methods are reminiscent of the intensional and extensional definitions for categories in §2.1, which is a further reflection of the similarity between sort-types and ontological categories.

[35] Copestake's diagram does not show the bottom element ($\perp$), but it is a standard HPSG assumption that incompatible sort-types have $\perp$ as their conjunction or meet (Pollard & Sag 1987: 43).

[36] More formally, the two methods correspond to two widely used composite data types, *record* and *variant* (Moshier & Rounds 1987: 157), which in turn are based on two basic (nonfunctional) data type constructions in denotational semantics, *Cartesian product* and *disjoint union* (Cardelli 1988: 145). Ait-Kaci (1984: 1) and Carpenter (1992: 1) explicitly recognize attribute-value feature structures as "record-like data structures."

(10) a. `sign` ≔ [PHON: `phon`, SYNSEM: `synsem`, DTRS: `list`] (nonatomic)
(a sign is defined by three features with separately typed values)
`sign` ≔ `morph` ∨ `stem` ∨ `word` ∨ `phrase`
(a sign is defined as either a morph or a stem or a word or a phrase)

b. `cat` ≔ `n` ∨ `v` ∨ `a` ∨ `p` ∨ `np` ... (atomic)
(based on Bird & Klein 1994)

There is a noticeable correspondence between the subsumption-based definition for atomic sort-types and the [ATT: VAL] notation for features; for example, (10b) can be straightforwardly converted to the feature [CAT: N/V/A/P/NP ...]. Insofar as the familiar morphosyntactic and phonological features are concerned, this correspondence is fairly robust, as in (11).

(11) a. `gen` ≔ `masc` ∨ `fem` ∨ `neu` ↔ [GEN: MASC/FEM/NEU]
`tns` ≔ `pres` ∨ `pst` ∨ `fut` ↔ [TNS: PRES/PST/FUT]

b. `spread` ≔ +`spread` ∨ −`spread` ↔ [SPREAD: +(SPREAD)/−(SPREAD)]

This state of affairs leads us to the following generalization:[37]

(12) Feature-type correspondence (FTC)
There is a nomenclative correspondence between an attributive value and its (most specific) sort-type; if the latter has a nongeneric immediate supertype,[38] a further correspondence exists between that supertype and the value's attribute (e.g., [CAT: N] ↔ `cat` ⊑ `n`, [TNS: PST] ↔ `tns` ⊑ `pst`).

When working in minimalism, where most features and values are nongeneric, we may simplify (12) into (13).

(13) Simplified feature-type correspondence (SFTC)
(In the minimalist feature system) There is a nomenclative correspondence between the [ATT: VAL] feature notation and the `supertype` ⊑ `subtype` subsumption relation.

The above feature-type homonymy is common and innocuous, as nothing prevents "the same name [from] being used [simultaneously] for a type and for a feature" (Copestake 2002: 47). In fact, it gives us a quick method to sort-type an [ATT: VAL]-based feature system.

[37] See Bird (1995: 157) and Pollard & Sag (1987: 197, note 6) for similar ideas.

[38] This clause is needed since generic sort-types (e.g., `string`, `int`) carry little contextual information and hence are inefficient as appropriateness conditions. For instance, according to Copestake (2002) `string` is the supertype for arbitrary strings (each considered a sort-type), but we would not want to label all string values (e.g., person names, car models, word orthographies) with the vague attribute STRING; more specific attributes like NAME, MODEL, and ORTH are desired instead.

2.2.3.2 Feature classing

Sort-type-based feature organization is mainly used in unification-based theories. In minimalism it is only seen in Adger & Svenonius' (2011: 35) "simple feature-typing system," where the attribute part of an [ATT: VAL] notation is treated as a "feature class" and the value part as a "feature." However, this formulation may cause unnecessary confusion due to the accumulated ambiguity in *feature* (see §2.2.2) and the back-and-forth switch between feature talk and value talk, as in statements like "one class of features … can take another class of *features* as their *values*" and "an attribute is a feature class … and a *value* is a *feature*" (Adger & Svenonius 2011: 38; emphases mine). As such, it is more efficient to reformulate Adger and Svenonius' idea directly in sort-type terms, which would form a subpart of the HPSG feature type hierarchy and hence better match the title "simple feature-typing system."

That said, classing is indeed a popular alternative method to ontologically organize features, especially in systems that do not use the [ATT: VAL] notation or do not use values at all. A typical example is the "feature geometry" approach defined in (14).[39]

(14) Features can be classified into taxonomic categories based on their grouping into higher-level functional units constituting natural classes. The hierarchical organization thus obtained is a *feature tree* or *geometry*, where individual features are organized under *class nodes* and class nodes dominated by a higher-level class node called the *root node* (Clements 1985: 228, 234, McCarthy 1988: 89, Clements & Hume 1995: 248, Hall 2007: 313).

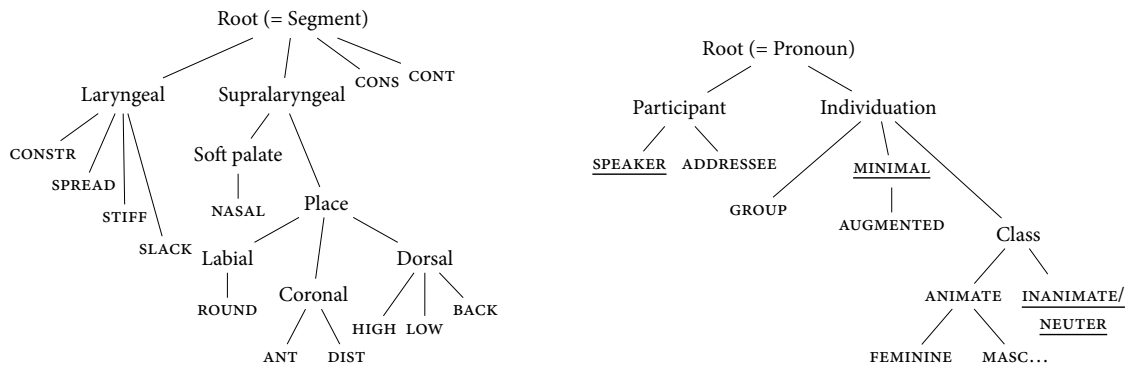
While feature geometry is mainly used in phonology, it has also been harnessed in (morpho)syntax, notably in Harley (1994) and Harley & Ritter (2002). Figure 2.4 displays two concrete proposals, one from phonology (adapted from Sagey 1986: 2) and the other from morphosyntax (adapted from Harley & Ritter 2002: 486).

Unlike Adger and Svenonius' system, which has mixed valency and uses the [F: α] format, the features in Sagey's (1986) system (Figure 2.4a) are all boolean and those in Harley & Ritter's (2002) system (Figure 2.4b) all privative. As such, the class-based feature organization does not cause terminological confusion in these systems. Since I mainly use the [F: α] format and (attributively) multivalent features, I opt for the type-based organization, though as I demonstrate below, the two strategies (typing and classing) are not necessarily incompatible.

First, the feature geometry model can be given an attribute-value transcription. Take Figure 2.4a for example. A phoneme segment can be viewed as a sort-type `seg` with four appropri-

[39] The formal underpinning of feature geometry is not exactly clear. For one thing, *geometry* is a misnomer (McCarthy 1988: 85, Kornai 2008: 32), as the definition of a feature geometry is unlike that of a standard geometry (i.e., a space together with a set of transformations and a group on that space; Brannan et al. 2012: 4). For another, despite the widely used tree diagrams, a feature geometry is not a (canonical) tree (Clements & Hume 1995: 249), for its terminals are placed on separate tiers, which are conceived as planes in Clements (1985) and Sagey (1986).

[40] The underlined features are the defaults in case of underspecification.



(a) Phonology **(b)** Morphosyntax[40]

Figure 2.4 Two proposed feature geometries

ateness condition attributes, two simple (i.e., CONS and CONT) and two complex (i.e., LARYNGEAL and SUPRALARYNGEAL). Similarly, in Figure 2.4b a pronoun can be viewed as a sort-type `pron` defined by the two attributes PARTICIPANT and INDIVIDUATION. I partly present the transcriptions in (15).[41] There are certain complications as we move on to the lower nodes—especially in Harley & Ritter's proposal, where the features are privative[42] and the internode relationship is inconsistent[43]—but a complete transcription is possible for both diagrams in Figure 2.4. I leave out the detail due to space limitation.

(15) a. $\mathtt{seg}\begin{bmatrix} \text{CONS:} & \mathtt{bool}_0 \\ \text{CONT:} & \mathtt{bool}_0 \\ \text{LARYNGEAL:} & \mathtt{lar} \\ \text{SUPRALARYNGEAL:} & \mathtt{supra} \end{bmatrix}$ b. $\mathtt{pron}\begin{bmatrix} \text{PARTICIPANT:} & \mathtt{part} \\ \text{INDIVIDUATION:} & \mathtt{ind} \end{bmatrix}$

Second, a feature geometry encodes not a single attribute-value structure but a set of possible structures, all of which are typological variants of the same sort-type. Take `agr` for example. In HPSG it is usually defined by the attributes PER, NUM, and GEN (sometimes also CASE). However, since some languages lack number and others lack gender, the type signature for `agr` in fact changes from language to language (e.g., [PER, NUM, GEN] or [PER, GEN] or [PER]). A feature geometry for agreement thus specifies the underlying set Φ = {PER, NUM, GEN} for all possible `agr`-attribute sets and rules out impossible sets; that is, the possibility space for the definition of `agr` is a subset of the powerset of Φ.

[41] See Gazdar & Pullum (1982: 42), Bird & Klein (1994: 462), Orgun (1996: Chapter 6), and Kracht (2007: Lecture 3) for similar practices. The bottom-left subscripts are sort-types. I use $\mathtt{bool}_0$ ($:= + \vee - \vee 0$) to incorporate underspecification, though in the feature geometry model nondistinctive features are simply omitted.

[42] Recall from §2.2.2 that privative features can occupy the value slot in [ATT: VAL], so we can legally use the [ATT: VAL] format to represent the class-feature relation in privative systems as long as the VAL slot is kept privative.

[43] e.g., some edges in Figure 2.4b encode the class-feature/subclass relation (e.g., SPEAKER $\in$ Participant, Class $\subset$ Individuation) while others encode existential implication (e.g., $\exists$AUGMENTED $\Rightarrow$ $\exists$MINIMAL). Besides, some sister nodes can cooccur (e.g., Participant and Individuation, SPEAKER and ADDRESSEE) while others cannot (e.g., the Class nodes). Such inconsistency makes Figure 2.4b hybrid in nature (see Harbour 2016: 192 for a similar remark).

In sum, feature classing and feature typing as organizational tools are cut out for different tasks: feature classing is more suitable for crosslinguistic generalizations, while feature typing is more suitable for methodical data descriptions when the typological parameters are fixed.

2.2.4 Granularity

The feature systems we have seen are essentially theories of componential analysis that break observables down to hypothetical underlying units. Hence, a methodological question is, How fine-grained should the units be? To use an analogy from physics, one can analyze matter based on either molecules and atoms, or protons, neutrons, and electrons, or quarks and leptons. Each analytical level magnifies the previous level to a higher granularity and reveals more detail about the matter being studied. Similarly, among featural analyses, some stop at a relatively coarse level, while others proceed to various more fine-grained levels. The granularity difference adds to the terminological confusion, for an atomic feature/value at a coarser granularity level may become a category at a finer level, which has its own intensional definition based on further features. Nevertheless, the underlying logic for this confusion is a benign one: there is no fixed boundary between categories and features; what we have is a series of granularity levels instead. In other words, just like the feature/value distinction, the category/feature distinction is also relative.

Take number for example. Unification-based theories usually adopt simple labels like SG and PL, assuming a granularity level where these are atomic; thus, a (partial) specification for *dog* may be [NUM: SG] or [+SG]. By contrast, some minimalist theories (e.g., Harley & Ritter 2002, Harbour 2008 et seq., Adger 2010) propose further componential analyses for numbers; for instance, Adger (2010: 192) treats SG as an abbreviation for [+ATOMIC, −AUGMENTED]. Substituting this into the feature set of *dog*, we get the more fine-grained specification [NUM: [+ATOMIC, −AUGMENTED]] or [+[+ATOMIC, −AUGMENTED]]. Note that the granularity level shift not only lifts SG from an attributive value to a taxonomic category but also enriches the feature inventory (i.e., the set of ground terms in the featural metalanguage).

Separating levels of abstraction (or "levels of concern"; see Martin 1986: 3–4) is an important strategy across disciplines (Martin's discussion is on computer science; see Cheng 2015: 22 for a similar remark on mathematics); it is also a recurring theme in this dissertation. Specifically, I distinguish two dimensions of granularity-induced variation for syntactic categories: the *syntagmatic* (or *combinatorial*) dimension and the *paradigmatic* (or *taxonomic*) dimension.

(16) a. **Syntagmatic granularity:** A sequence of combinatorially oriented categories may be collapsed into a single category (e.g., X–Y–Z → A), and a single category may be split into a sequence of categories (e.g., X → X_1–X_2–…–X_n). Compositionality is preserved in both scenarios, with the same rules applicable at all granularity levels.

b. **Paradigmatic granularity:** At a fixed syntagmatic granularity level, each syntactic cat-

> egory X may be distinguished from other categories either by a single atomic feature F or by a (possibly structured) set of features (e.g., [F_1, F_2, …]), each of which may similarly be defined either atomically or componentially. The combinatorial nature of X is unchanged however it is taxonomically defined.[44]

In minimalism, for example, the distinction between the Chomskyan core functional categories and the cartographic functional categories (e.g., C–T vs. Foc–Top–Fin–Mod–T–Asp) is one of syntagmatic granularity,[45] while that between the conventional grammatical categories and their componential analyses (e.g., [CAT: N/V] vs. [CAT: [±N, ±V]]) is one of paradigmatic granularity. For the current purpose, a crucial difference between the two dimensions lies in how the different granularity levels interact; namely, how more fine-grained particulates combine into (equivalents of) their coarser counterparts. As is generally understood, the combination of syntagmatic categories is scope-based however fine-grained they are (even nanosyntactic categories still merge along a c-command hierarchy), and their composition is noncommutative; for instance, if we write · for a general composition function (whose definition varies from case to case), then obviously $[\![\text{Top}]\!]\cdot[\![\text{Fin}]\!] \neq [\![\text{Fin}]\!]\cdot[\![\text{Top}]\!]$ and $[\![\text{C}]\!]\cdot[\![\text{T}]\!] \neq [\![\text{T}]\!]\cdot[\![\text{C}]\!]$.[46] By contrast, the combination of paradigmatic componential features is usually non-scope-based[47] and their composition is commutative; for instance, $[\![+\text{N}]\!]\cdot[\![-\text{V}]\!] = [\![-\text{V}]\!]\cdot[\![+\text{N}]\!]$ (where · is conjunction).[48] This contrast suggests that we cannot freely push syntax all the way down to feature level (pace Collins 2017, Tsoulas 2017) but should at least distinguish syntagmatic and paradigmatic decompositions.

Finally, let us return to the question of how fine-grained the analytical units should be. My stance on this is similar to that of McCarthy (1988: 94–97) and Corbett (2012: 17–21):[49] there is no absolute right or wrong (or more/less elegant) between analyses at different granularity levels; all there is are different perspectives and purposes. Thus, in a data-glossing task the coarse labels SG and PL are perfectly adequate, but if one's purpose is to explain why we have these but not other categories or why certain grammatical patterns are prevalent or missing, then one might wish to bring the analysis to a higher granularity. Both purposes are legitimate and neither invalidates

[44] Gazdar et al. (1985: 20, citing Halle 1969) similarly remark that "there is an exact equivalence between generative systems that use complex symbols (= matrices of distinctive feature specification) and those that do not."

[45] See Ramchand & Svenonius (2014) and Biberauer & Roberts (2015) for similar granularity-oriented comparisons between the Chomskyan and the cartographic functional hierarchies.

[46] The (non)commutativity here is about semantic denotations instead of syntactic terms. So even though categories like Neg may have flexible syntactic positions, as in Neg–T–V vs. T–Neg–V, this does not make their composition commutative, because $[\![\text{Neg}]\!]\cdot[\![\text{T}]\!] \neq [\![\text{T}]\!]\cdot[\![\text{Neg}]\!]$. Besides, the variation Neg–T vs. T–Neg cannot be described as "syntactically commutative" either, because derivationally it is not a simple swap between Neg and T but involves two separate configurations Merge(Neg, TP) and Merge(T, NegP), where the input of Merge changes and hence no commutativity-related issue arises at all.

[47] Recall from note 11, p. 8 that the scopes of contrastive features in Dresher (2009) (and similarly those in Harley & Ritter 2002) are not combinatorial scopes, as they do not encode c-command relations or first-Merge positions.

[48] A notable exception is Harbour (2016) as mentioned in §2.2.2, where the fine-grained person and number features (e.g., [±AUTHOR], [±ATOMIC]) are combined in an order-sensitive mode and composed by (noncommutative) function application. Harbour's proposal in a sense unifies granularity magnification along the two dimensions.

[49] McCarthy points out that reducing a class node on the feature geometry to an *n*-ary [ATT: VAL] feature "changes nothing at [the assumed] level of analysis" (p. 94), and Corbett demonstrates that neither multivalence nor bivalence has convincingly proven empirical superiority over the other.

the other. In fact, a well-designed featural metalanguage should be able to accommodate and disambiguate multiple analytical purposes and granularity levels. A ⟨feature, value, type⟩-based metalanguage achieves the first goal but not the second; one way to improve this situation is to include a granularity level component in the featural descriptions or at least take the relative granularity levels into account when comparing or integrating syntactic categories from different proposals. I write the granularity level in subscript when necessary (e.g., $\mathrm{C}_{\mathrm{CFC}}$).

2.2.5 Interim summary

In this section I examined three key aspects of linguistic feature systems (i.e., classification, notation, organization) and summarized four major components for the featural metalanguage (i.e., specification, valuation, typing, granularity). First, a feature specification Σ consists of one or more atomic terms of the metalanguage (e.g., [+N] consists of two atomic terms, + and N). Depending on term structure, the specification may be atomic (i.e., privative, Σ_{P}) or bipartite (comprising a base specification and an extension, $\langle \mathrm{F}, \alpha \rangle$). Second, a bipartite specification is usually put in a feature-value notation, either coefficiently [αF] or attributively [F: α]. Third, terms $t, t', t'', \ldots$ of the featural metalanguage (especially atomic ones) can be organized by logico-mathematical typing (n) or taxonomic sorting/classing (σ/κ). Fourth, a linguistic object Ω may have different feature specifications at different granularity levels (GRANs), which are in one-to-one correspondence ($\cong$) since they all describe the same object. Below is a formal summary of the featural metalanguage:[50]

(17)
$$\Sigma \coloneqq \Sigma_{\mathrm{P}} \vee \langle F, \alpha \rangle \qquad \text{(privative vs. valued)}$$
$$\langle F, \alpha \rangle = [\alpha\mathrm{F}] \mid [\mathrm{F}\colon \alpha] \qquad \text{(coefficient vs. attributive)}$$
$$t : n \mid t : \sigma \mid t : \kappa \qquad \text{(typing, sorting, classing)}$$
$$\forall \Omega \,.\, \Sigma_{\mathrm{GRAN}_i}(\Omega) \cong \Sigma_{\mathrm{GRAN}_j}(\Omega) \qquad \text{(cross-granularity correspondence)}$$

2.3 Syntactic category

In this section I present my fundamental assumptions about syntactic categories. §2.3.1 elaborates on the aforementioned taxonomic and combinatorial perspectives to define syntactic categories and introduces the key notion mergeme. §2.3.2 further examines the formal nature of syntactic categories and puts forward a mergeme-based template for them, together with a few remarks on the template in comparison with other similar notions in the literature.

[50] I use $\vee$ for disjoint union (i.e., exclusive *or*) and | for nondisjoint union (i.e., inclusive *or*); e.g., while a specification cannot be both privative and bipartite, a bipartite specification can be written both coefficiently and attributively.

2.3.1 Definition

2.3.1.1 Taxonomic vs. combinatorial perspectives

To recapitulate, the taxonomic perspective to define syntactic categories is distinction-driven. The intercategorial distinction may be rendered atomically or componentially; for instance, the N/V distinction may be stated as either [N]/[V] or [+N, −V]/[−N, +V]. There is much debate over what features to use and what content to give them, but the basic idea that the features in question are meant to distinguish some categories from others is uncontroversial. On the other hand, the combinatorial perspective to define syntactic categories is less about taxonomic distinction[51] and more about how various categories interact with one another. This perspective is most systematically pursued in categorial grammars, where all categories except a few basic ones are defined by function construction; for example, given the basic categories NP and S, an intransitive verb is defined as $\frac{S}{NP}$, the category that when combined with NP yields S. To a lesser extent, minimalist grammars are also combinatorially oriented. For instance, Stabler (1997: 71) divides syntactic features into four classes—Base, Select, Licensors, and Licensees—whereby two lexical items like $\textit{make}_{[=d,\ =d,\ v]}$ and $\textit{tortillas}_{[d]}$ ([d], [v] ∈ Base, [=d] ∈ Select)[52] can combine into $\textit{make tortillas}_{[=d,\ v]}$ by matching a [=d] against [d]. Both frameworks express the syntacticness of syntactic categories by giving them a combinatorial signature (functional typing in categorial grammars and merging-by-feature-matching in minimalist grammars).

The two perspectives are currently developing in opposite directions. Taxonomically oriented proposals are getting more and more elaborate as the granularity is pushed to increasingly higher levels, while combinatorially oriented proposals (especially those in minimalism) are getting barer and barer, with various rules and features being recast in configurational or nonsyntactic terms; for example, subcategorization frames are replaced by argument-introducing heads like Voice (Kratzer 1996) and Appl (Pylkkänen 2008), c-selection is partly reduced to the conceptually grounded functional hierarchies (Cinque & Rizzi 2008: 52, Shlonsky 2010: 425), and s-selection is entirely moved out of syntax (Borer 2005a). After this much relocation, what still remain on the combinatorial side of syntactic categories are perhaps the most basic and general properties, such as combinability, projectivity, and linkability. That is, syntactic categories can merge (or unify) their separately encapsulated information; they can project hierarchical structures according to some systematic template (e.g., X-bar) or algorithm (e.g., labeling); and they can be derivationally or representationally linked to one another (e.g., via chains) for information integration or structural transformation.

A noticeable characteristic of the combinatorial properties is that they are predicated of syn-

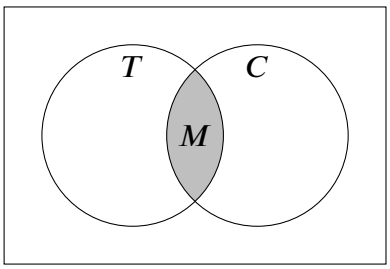
[51] More precisely, the taxonomic distinctions associated with the combinatorial perspective are not between individual categories but between syntactic and nonsyntactic categories. When we restrict our universe of discourse to syntactic categories, however, the latter type of taxonomic distinction can be safely ignored. Henceforth I use *taxonomic* in the restricted sense (i.e., about the taxonomy of individual categories) unless otherwise clarified.

[52] Stabler uses lowercase features to indicate in-situ spell-out.

tactic categories collectively (e.g., combinability is a property of syntactic categories as a class). As such, they may be considered systematic properties of the human language syntax. By comparison, taxonomic properties are predicated of syntactic categories individually (e.g., [+v] is a property of V as a specific category). This is what I mean by a distinction in perspective. Some collectively predicated properties have individually predicated counterparts—call them *instantiations*—for example, combinability may be instantiated as the edge feature (à la Chomsky 2007), and linkability may in certain situations be instantiated as the EPP feature. However, once it is instantiated, a combinatorial property becomes taxonomic, since it may now be used to distinguish one category from another (e.g., $T_{[+EPP]}$ vs. $T_{[-EPP]}$). By contrast, it is much less usual to directly put a feature like [LINKABLE] in a specific category's feature set, since all syntactic categories are potentially linkable and there is no need to particularly mark this general capacity.

2.3.1.2 The mergeme

The notion of mergeme lies precisely at the intersection of the two perspectives to define syntactic categories. Since the two perspectives have different concerns, a *syntactic-category-defining property* (SCDP) from one perspective might not be one from the other. Taxonomically, any contrast that systematically distinguishes one set of SOs from another qualifies as an SCDP, be it combinatorially relevant or not; thus, not only contrasts like [N]/[V] and [+EPP]/[−EPP] but also those like [STRESSED]/[PLAIN] and [ELEGANT]/[VULGAR] may serve as SCDPs.[53] Similarly, what qualifies as a combinatorially oriented SCDP may not be category-defining from the taxonomic perspective either. Take combinability for example. It is a most important defining property of syntactic categories from the combinatorial perspective yet completely "useless" from the taxonomic perspective, for it cannot distinguish any syntactic category from any other. I propose the following classification for SCDPs.



T = Taxonomic
C = Combinatorial
M = Mergeme

Figure 2.5 Three classes of syntactic-category-defining properties

In Figure 2.5 the universe (i.e., the rectangle) contains all category-defining properties from all domains, and the subuniverse we are concerned with (i.e., the two circles) contains all and only SCDPs. SCDPs in the two circles have two distinctions: one concerns the aforementioned individual vs. collective predication, and the other concerns their relation with surrounding SCDPs:

[53] Reusing existing features for extra purposes, or maximizing minimal means in Biberauer's (2017a) terms, is a fundamental strategy in human language.

SCDPs in $\boldsymbol{T}$ are (partially) ordered while those in $\boldsymbol{C}$ are unordered. This is because taxonomic SCDPs (i.e., features) are ontologically interconnected—in the form of a HPSG sort-type hierarchy or a huge feature geometry—whereas combinatorial SCDPs are conceptually independent of one another; for instance, there is no obvious ordering between projectivity and linkability.

The two distinctions suggest that whatever falls inside $\boldsymbol{T} \cap \boldsymbol{C}$ is (i) predicatable both individually and collectively, and (ii) simultaneously ordered (in $\boldsymbol{T}$) and unordered (in $\boldsymbol{C}$) in relation to other SCDPs. There may be multiple SCDPs meeting these conditions, but here I only discuss one: the mergeme. Accordingly, I refer to the intersection of the combinatorial and the taxonomic perspectives on syntactic categories as the *mergeme-based perspective*. As a combinatorial SCDP, the mergeme describes the fundamental abstract pattern by which syntactic categories and their projections are stacked into computable hierarchies; namely, the scope-based ordering commonly known as "extended projection" or "functional hierarchy." As a taxonomic SCDP, it singles out the minimal mergeable and composable units relativized to a granularity level that are used to instantiate the abstract scoping pattern. As the definition of the mergeme relies on several theoretical assumptions, I first spell them out before continuing my discussion.

2.3.1.2.1 Derivation as formal proof First of all, I do not view a syntactic derivation as a psychologically real process but treat it like a formal proof; thus, a derivational description like "a *wh*-phrase first merges as the complement of V and then moves to the specifier of C" does not guarantee at all that this is what happens in the brain when an utterance like *What did you eat?* is produced but merely provides a possible logical deduction from the premises (the lexical items) to the conclusion (the utterance). This point has been repeatedly emphasized by Chomsky since the 1960s (in the context of the competence-performance distinction), as in the following quotes:

> When we say that a sentence has a certain derivation with respect to a particular generative grammar, we say nothing about how the speaker or hearer might proceed … to construct such a derivation. These questions belong to the theory [of] performance. (Chomsky 1965: 9)
>
> [A] generative system involves no temporal dimension. In this respect, generation of expressions is similar to … constructions of formal proofs. … The actual construction of a proof may well begin with its last item, involve independently generated lemmas, etc. … The same is true of generation vs. production of an expression, a familiar competence-performance distinction. (Chomsky 2007: 6)

Thus, an SO ω is well-formed iff it is derivable in the sense that given an array of well-formed SOs as premises and a set of derivational rules, ω can be deduced as a conclusion.[54] This conception of derivation is essentially the same as that in logic, as in Figure 2.6 (see also Tomalin 2006).

[54] Jacobson (2012: 110) similarly remarks that "builds" in statements like "syntax first 'builds' syntactic objects such as trees" is a metaphor for "proves well-formed."

$$\dfrac{\dfrac{p \quad q}{p \wedge q}\ \wedge\text{-introduction} \qquad p \wedge q \rightarrow r}{r}\ \text{modus ponens}$$

(a) A proof tree in logic

$$\dfrac{\dfrac{\textit{the} \quad \textit{cake}}{\{\textit{the}, \textit{cake}\}}\ \text{merge}}{_{\textit{the}}\{\textit{the}, \textit{cake}\}}\ \text{label}$$

(b) A (partial) proof tree in syntax

Figure 2.6 Derivation in propositional logic and minimalist syntax

Any SO that can be associated with a formal proof is derivable. Among the syntactic derivational rules, the two most fundamental ones are perhaps *merge* and *label*—mergeability (i.e., combinability) is what licenses new premises into the derivation, and labelability is a condition that must be met by every valid derivation. For example, in Figure 2.6b *the* and *cake* would not be able to appear above the inference line if they were not mergeable, and the (sub)derivation itself would be invalid if the conclusion {*the*, *cake*} were unlabelable. The labelability condition is important because Merge qua set formation can be mechanically applied to any pair, be it ⟨*the*, *cake*⟩ or ⟨*two*, *cake*⟩ or ⟨*the*, *the*⟩, but apparently not all merged products can be successfully interpreted. In Chomsky's (2013: 37, 43) words, the label is the "identification of the category of a phrase" as well as what licenses an SO so that it "can be interpreted at the interfaces."

Note that labeling is relevant to both the taxonomic and the combinatorial perspectives on syntactic categories. Taxonomically, it records and passes on the categories of merged objects (at least across certain inference steps); for example, in Figure 2.6b the label of {*the*, *cake*} can only come from *the* and *cake* but not anywhere else.[55] Combinatorially, it imposes a well-formedness constraint on the product of Merge.[56] Since labeling serves both taxonomic and combinatorial purposes in syntactic derivation, it naturally goes hand in hand with the mergeme-based perspective on syntactic categories. To anticipate the content in §2.3.2, the mergeme of a syntactic category is its label-provider, so a mergemeless category is also labelless.[57]

2.3.1.2.2 Decomposability as derivability Since I will be dealing with many high-granularity constructions later on (e.g., word-internal structures), I heavily rely on the method of decomposition. For instance, a noun like *cake* may be decomposed into a categorizer *n* and a root $\sqrt{\text{CAKE}}$ as in Distributed Morphology (DM; Halle & Marantz 1993, 1994), and a pronoun like *he* may be decomposed into a series of categories (e.g., Num, Per) as in nanosyntax. Viewing derivations as formal proofs, I present such lexically decomposed structures without assuming that speakers really assemble words from absolute scratch every time they produce a word or phrase (e.g., first select *n*, then select $\sqrt{\text{CAKE}}$, and finally put them together into *cake*). The purpose of decomposition is not to describe psychological processes but merely to show that even when we push

[55] In GB terms this is simply what projection means; e.g., D projects from the head *the* to the phrase *the cake*.

[56] Merge may be seen as a well-formedness constraint too. Brody (2002: 24) views it as a "tri-representational constraint" linking three separate representations, respectively of the two items being merged and the merged product.

[57] Labelless ≠ nameless. An SO with no label, such as a conjunction or a root (Chomsky 2013: 47), still has a syntactic category (e.g., Conj, √) and can be named (in §2.4.1 and Chapter 4 I will call such a category an "acategorial category"). It just cannot label larger SOs containing it or "head/project phrases" in conventional terms.

the analysis to a very high granularity, what we see are still derivable structures; conversely, as long as a decompositional analysis can be demonstrated as derivable (by a formal proof), it is at least theoretically valid (empirical validity is a separate issue). On the contrary, if at some granularity level a proposed analysis is no longer derivable by the preestablished rules, then it means decomposition cannot reach that level, and we should stop. As such, to what extent decomposition can be used as an analytical tool does not hinge on speakers' psychological intuition[58] but hinges on whether the decomposed result is formally derivable; in other words, decomposability is equivalent to derivability.

2.3.1.2.3 Derivability-compositionality correspondence Lastly, I assume a correspondence between syntactic derivability and semantic compositionality and use it to check the validity of a derivation qua proof. In other words, I assume that every derivable SO has a computed meaning. In formal semantics, this follows from the fundamental principle of compositionality (i.e., Frege's principle), which is formulated by Partee (2004: 153) as "the meaning of an expression is a function of the meanings of its parts and of the way they are syntactically combined."[59] In minimalism, on the other hand, the correspondence follows from the "interface condition":

> If language is to be usable at all, its design must satisfy an "interface condition" (IC): the information in the expressions generated by L must be accessible to other systems, including the sensorimotor (SM) and conceptual-intentional (C-I) systems that enter into thought and action. … The last line of each derivation D is a pair ⟨PHON, SEM⟩. D converges if PHON and SEM each satisfy IC. … The NS derivation has to provide the basis for assignment of order at the SM interface, and for multiplicity of semantic properties at the C-I interface. (Chomsky 2004: 106, 110)

Under "semantic properties" Chomsky lists argument structure, scopal properties, and discourse-related properties. The interface condition is relevant to both phonology and semantics, but here I mainly resort to the latter, for many high-granularity SOs lack exponents, such as *n* (the nominalizer) in *cake*.[60] Considering a single exponent may spell out a chunk of structure, phonology is not a very useful gauge to verify or falsify decompositional analyses. Semantically, however, we can consistently verify derivational validity by constructing a function that maps the meanings of the premises to the meaning of the conclusion in some principled mode.

With the background assumptions laid out, now I proceed to officially define the mergeme:

[58] As Chomsky (1965: 8) remarks, generative grammars deal with "mental processes that are far beyond the level of actual or even potential consciousness," and what a generative grammar specifies is "what the speaker actually knows" instead of "what he may report about his knowledge."

[59] Or as Montague (1974: 225) conceives it, there is an algebraic homomorphism from syntax to semantics.

[60] There are even proposals (e.g., Kayne 2016) arguing that certain types of syntactic objects are consistently null.

(18) The *mergeme* represents the minimal units of a particular granularity level that can be merged according to some scope-based ontological hierarchy and interpreted in composition with other units. Such units are minimal in that there are no smaller units at the same granularity level from which they can be derived.

I have three further comments on the above definition. First, the mergeme, being an SCDP, is not a minimal mergeable unit itself but merely represents one, because Merge is defined on SOs rather than SCDPs.[61] Second, the mergeme cannot be identified with the cartographic categorial (or interpretable) feature despite the close resemblance between the one category–one mergeme scenario here and the "one (morphosyntactic) property–one feature–one head" maxim in cartography (Cinque & Rizzi 2010: 74; see also Rizzi & Cinque 2016: 155). On the one hand, the mergeme is both a taxonomic and a combinatorial SCDP, while a categorial feature, being a specific feature, is merely taxonomic. On the other hand, underlying the cartographic maxim is the assumption that a syntactic category is defined by a single (or core) feature (i.e., its categorial feature); this assumption is not shared by the mergeme-based approach, where the mergeme is just one of the defining properties on a syntactic category (whatever the granularity level is). Nonmorphosyntactic properties like [EPP] and [ELEGANT] can also help define categories in the mergeme-based approach but cannot under the cartographic maxim, because that would predict heads like EPP and Elegant, which are unattested.[62]

Third, the minimality of a mergeme-bearing unit is only relativized to its granularity level. This is again contrasted with the cartographic conception, where the categories qua features are the fixed atoms of syntactic computation (Cinque & Rizzi 2010: 74). Compare V and its tripartite decomposition Init–Proc–Res (Ramchand 2008) for example.[63] From the cartographic perspective V is certainly not minimal in the face of Init–Proc–Res, but from the mergeme-based perspective both V and Init/Proc/Res can be considered minimal as long as they do not co-occur at the same granularity level.

[61] I use "mergeable unit" instead of "syntactic object" or "lexical item" in the definition of mergeme to highlight the abstract nature of the entity and facilitate discussion across granularity levels (the term *lexical item* in practice often carries a vocabulary-item-oriented connotation).

[62] This suggests that the "feature" meant by the maxim has a more restricted domain than that of all features; it presumably only ranges over features that can participate in scope-based (or syntagmatic, in the sense of (16)) combination and composition; i.e., the interpretable or categorial features. Thus, unvalued/uninterpretable features are outside the purview of the maxim. Unfortunately this point is usually left implicit, which leaves the maxim in a questionable state (see, e.g., Baker 2018: 19). Two notable exceptions are Kayne (2005: 15), who formulates a "principle of decompositionality" (i.e., "UG imposes a maximum of one *interpretable* syntactic feature per lexical or functional element"), and Shlonsky (2010: 426), who specifies a "division of labor" (i.e., "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") (emphases mine).

[63] Ramchand (2008: 39): "[This is] a splitting up of what we normally think of as V, in the spirit as Rizzi's (1997) splitting up of the C node … or Pollock's (1989) splitting up of Infl."

2.3.2 Template

Based on the foregoing discussion, I propose the following featural definition (or template) for syntactic categories:

(19) syntactic category$_{\text{LANG, GRAN}}$:= [⟨CAT: cat, POS: nat⟩, NONMERGEME FEATURES] (= (1))

This template consists of four parts: (i) the name of the syntactic category being defined (indicated by a placeholder *syntactic category*); (ii) the language variety and granularity level the category inhabits (indicated by the subscripts LANG and GRAN); (iii) the mergeme (indicated by a pair of [ATT: VAL]-format features);[64] and (iv) other (nonmergeme) features on the category (indicated by NONMERGEME FEATURES or NM), which may be purely taxonomic features (e.g., [ELEGANT]) or taxonomically instantiated combinatorial SCDPs (e.g., [EPP]). There is no place in the template for the morphosyntactic features commonly used to handle agreement phenomena (e.g., φ, case) since those are not SCDPs; they are similar to Chomsky's (1995: 231) "optional features" instead, which are "added as LI [= lexical item] enters the numeration." Also missing from the template are the phonological and semantic/encyclopedic features conventionally associated with LIs—those are irrelevant to the definition of syntactic category and only become relevant when we turn to define LI, which may be represented by the following featural template:

(20) lexical item := [[PHON: phon], [SEM: sem], [SYN: syn]]

This is not a novel proposal but merely reformulates a commonly accepted definition for LIs in the featural metalanguage introduced in §2.2. In particular, syn is the sort-type for the entire feature set in (19), which forms a (complex) value of the attribute SYN. I will leave PHON and SEM aside until Chapter 3 and use the rest of this section to further elaborate on the two middle components of the template in (19), namely the granularity level and the mergeme.

2.3.2.1 Granularity levels

The major granularity levels I assume are (from finest to coarsest) cartography (Cart), core functional category (CFC), phase (Ph), extended projection (EP) (based on Biberauer & Roberts 2015) as well as two very coarse levels that I call archicategory (Arch)[65]—to reuse a Praguian term (see §2.2.1.2 and note 8, p. 7)—and undivided categorial space (○). Each granularity level (except ○) is a set G of sets of completely specified mergeme-bearing categories, with each set immediately dominated by G being a *categorial sequence* or *c-sequence*. I illustrate this in (21)

[64] Recall that the typewriter-font cat and nat are sort-types for values of the features [CAT] and [POS].

[65] Depending on one's assumptions about the number of extended projections and the arity of category subtyping, Arch may be a separate level (if one assumes more than two extended projections and binary subtyping) or simply EP (if one assumes only two extended projections or multiary subtyping); see §6.5.1 for elaboration.

with a possible instantiation of the CFC level.[66]

(21) CFC = {{V, *v*, T, C}, {N, *n*, Num, D}, …}

I will return to the c-sequence-internal ordering (i.e., the sequence) later. For now let us focus on the granularity levels. The two very coarse levels Arch and EP (or ArchEP if the two are identified based on note 65, p. 30) have no proper c-sequences but only individual categories; for example, when there are only two extended projections, verbal and nominal, ArchEP = {V, N}. Alternatively we can view V and N in ArchEP as trivial (i.e., singleton) c-sequences, whence ArchEP = {{V}, {N}}. The coarsest level ○ does not have any concretely defined category, though this does not mean that it is completely void of information; among others, it is still a granularity level and part of the SCS. In other words, ○ is concretely empty but abstractly contentful. I stipulate it as a singleton set containing the undivided categorial space. As suggested above, the granularity levels can be ordered by a binary relation is-coarser-than-or-equal-to (≤):[67]

(22) ○ ≤ Arch ≤ EP ≤ Ph ≤ CFC ≤ Cart

The relation ≤ is reflexive, transitive, and antisymmetric, which makes the set of granularity levels into a *partially ordered set* or *poset*.[68] In (22) we further have a *totally ordered set* or *chain*, though this is not always the case, because there may be numerous intermediate granularity levels, not every two of which are comparable. Imagine two arbitrary levels A and B, both of which inherit the basic setting of CFC and further elaborate its verbal c-sequence, though in two different ways: A divides T_{CFC} into Mod–Tns–Asp while B divides V_{CFC}[69] into Init–Proc–Res, as in (23).

(23) a. A = {{V, *v*, Asp, Tns, Mod, C}, {N, *n*, Num, D}, …}

b. B = {{Res, Proc, Init, *v*, T, C}, {N, *n*, Num, D}, …}

While both A and B are ordered between CFC and Cart (both are finer than CFC and coarser than Cart), there is no obvious ordering between the two of them (neither A nor B is finer or coarser than the other). Considering these and other possible intermediate levels, the set of granularity levels may be rather large, with (23) being only a small subset of it.

[66] The concrete categories presented here are merely expository. The general format is compatible with any decision made on this and other details (e.g., additional c-sequences for adjectives/adpositions).

[67] One could also adopt an opposite relation is-finer-than-or-equal-to. The gist remains the same.

[68] See §5.1.1 or Appendix C for definitions of these order-theoretic notions.

[69] This label does not suggest that the category V itself is a core functional category but merely means that V is part of the granularity level labeled CFC; i.e., the coarsest level containing Chomsky's (2001) core functional categories. Since Chomsky (2001) not only uses the core functional categories but at least also uses the basic lexical categories (e.g., V), the granularity level termed CFC cannot exclusively contain the core functional categories either but should at least have an additional "base category" for each c-sequence. This base category (in its coarsest form) is commonly assumed to be a traditional "big X" (e.g., V, N), so I use it for expository purposes.

2.3.2.2 More about the mergeme

In (19) CAT and POS are two appropriateness conditions respectively taking attributive values of sort-types `cat` and `nat`, which are defined by subtype disjunction as follows:

(24) a. $\texttt{cat} \coloneqq \texttt{v} \vee \texttt{n} \ldots$ (c-sequence names)

b. $\texttt{nat} \coloneqq 1 \vee 2 \vee 3 \ldots$ (nonzero natural numbers)

We could alternatively write the mergeme as [MERGEME: `mergeme`], where `mergeme` is the sort-type defined as [CAT: `cat`, POS: `nat`] (or $\texttt{cat}_{\text{CAT}} \times \texttt{nat}_{\text{POS}}$ in product format). This would make the template fully HPSG-style, but for expository convenience (e.g., to reduce brackets) I will continue using pair notation. Recall that the mergeme as a combinatorial SCDP is the abstraction of the scope-based ordering among syntactic categories. As such, it determines a category's first-Merge situation, including its initial (or inherent) label and first-Merge position. We have seen in §2.3.1.2 that labeling goes hand in hand with the mergeme-based perspective on syntactic categories. For a mergeme-bearing minimal category, its mergeme provides its label. Although the pair notation ⟨CAT: `cat`, POS: `nat`⟩ looks nothing like the familiar labels (N, D, T, etc.), we will see that the former uniquely "picks out" the latter; hence, the mergeme may be viewed as a pointer feature or localizer, in that it points to (and integrates) other features or entities in the ontology.[70] Moreover, the mergeme ties category labels to granularity levels; for example, $\text{N}_{\text{EP}} \neq \text{N}_{\text{CFC}} \neq \text{N}_{\text{Cart}}$.[71] Since the mergeme determines the inherent label of its host category, we can identify the two properties having-a-mergeme and labeling; namely, the presence or absence of a `mergeme`-typed value in a syntactic category's feature set (which may be bivalently presented as [±MERGEME]) determines whether it is a labeling or a nonlabeling category (i.e., whether or not it can contribute to phrase-level labeling).

As for the first-Merge position, the values of CAT and POS—which correspond to the disjoined subtypes in (24) according to (12)—respectively specify a c-sequence and a position in that sequence relativized to a (presupposed) granularity level; for example, [CAT: **v**, POS: 3] is T_{CFC} or Asp_{A} (A is the arbitrary granularity level in (23a)). The `nat` value is obtained by a one-to-one function mapping each c-sequence to the natural number chain, as in (25), where $\leq$ is defined as has-a-combinatorial-scope-smaller-than-or-equal-to for syntactic categories and is-less-than-or-equal-to for natural numbers.

(25) $(\text{CFC}_{\mathbf{v}})\ \text{V}^{72} \leq v \leq \text{T} \leq \text{C} \rightarrow 1 \leq 2 \leq 3 \leq 4 \ldots$

The specific categories in (25) form a chain as the major granularity levels do in (24), but again

[70] This characteristic might be generalizable to all SCDPs in $T \cap C$ in Figure 2.5.

[71] To avoid confusion it is better not to use the same categorial label at different granularity levels. Since this is not realistic (in practice categorial labels are more or less "protocolized"), I use subscripts to distinguish homonymous categories of different granularity levels.

[72] $\mathbf{v} \neq \text{V}$. The former is the name of a c-sequence (i.e., a set) while the latter is that of a specific syntactic category.

this is not always true (see §2.4 for further discussion). Besides, the numerical indexing is merely used for expository convenience; what really matters is the actual scopal ordering inside each c-sequence. Thus, the same index may pick out different categories from different c-sequences, and the same information encoded by a numerical index may be alternatively encoded by a relational predicate; for example, at the CFC level [CAT: **v**, POS: 3] ≡ [CAT: **v**, POS: $\lambda c\!:\!\mathrm{CFC}_{\mathbf{v}}\,.\,v \leq c \wedge c \leq \mathrm{C}$].[73] The two representations are equivalent, but the natural numbers are obviously more readable, so I will stick to them in my presentation (until Chapter 6, where a deeper formalization of the c-sequence indexing will be provided).

The idea of indexing the c-sequence numerically is not new. Below are two previous proposals similar to my own:

(26) a. A category label is a pair consisting of a categorial and a functional specification. The categorial specification is designated as [NOMINAL] and [VERBAL] (which may be defined by features) and the functional specification is a valued feature F*n*, with F0 assigned to the lexical categories, F1 to the lowest level functional category, F2 to the next, and so forth. The functional value of a node is independent of its categorial analysis; for example:
V = ⟨[+V, −N], F0⟩, I = ⟨[+V, −N], F1⟩, C = ⟨[+V, −N], F2⟩;
N = ⟨[−V, +N], F0⟩, D = ⟨[−V, +N], F1⟩, P = ⟨[−V, +N], F2⟩.
(based on Grimshaw 2005: 3–4; see also Grimshaw 1991, 2000)

b. A lexical item is a set {K, F_1, …, F_n}, where K is an ordered pair ⟨Cat, N⟩ (notated as [Cat: N]), with Cat being drawn from the set {C, D, T, Num, Asp, …}−∅ and N from the set of natural numbers above 0. Hierarchies of projections are sequences of Ks whose second member is ordered by the relation <; for example:
⟨V, 1⟩ < ⟨*v*, 2⟩ < ⟨Pass, 3⟩ < ⟨Prog, 4⟩ < … < ⟨T, 8⟩ < ⟨Fin, 9⟩ < ⟨C, 10⟩;
⟨N, 1⟩ < ⟨*n*, 2⟩ < ⟨Poss, 3⟩ < ⟨Num, 4⟩ < ⟨D, 5⟩ < ⟨Q, 6⟩.
(based on Adger 2010: 198; see also Adger 2003)

Both Grimshaw and Adger cite cartographic results to support their indexed hierarchies. Grimshaw (2005: 47) follows Cinque (1999) in assuming that UG provides "a set of possible functional specifications" that are "universally organized in a hierarchy," and Adger (2010: 197, 199)—also following Cinque (1999)—assumes that functional categories are "ordered in a number of (universal) scopal hierarchies." Since the cartographically depicted "scopal hierarchies"—UG-given or not[74]—are carefully generalized based on ample empirical evidence, they are at least descriptively adequate, so I will use them to organize my c-sequences wherever applicable.

[73] The relational predicate is read "*c* (a $\mathrm{CFC}_{\mathbf{v}}$ category) has a scope larger than or equal to that of *v* and smaller than or equal to that of C." It returns `true` iff $c = \mathrm{T}$.

[74] There are works arguing that the cartographic hierarchies are not UG-given but reflect the "core function[s]" of categorial "domains" (Wiltschko 2014: 24, 79) or an extralinguistic "cognitive proclivity" (Ramchand & Svenonius 2014: 172). In §2.4 and Chapter 6 I will follow this line of thought and posit a shared "universal spine" (Wiltschko's term) for various c-sequences at the same granularity level.

That said, my mergeme-based approach differs from previous theories in a number of ways. First, my template in (19) is applicable to multiple granularity levels, whereas previous theories usually only consider a single level "hard-coded" into their primitive definitions, as reflected in the fixed hierarchies in (26). Second, I do not give the traditional big X categories (e.g., N, V) any special status; hence, there is no equivalent in my proposal to Grimshaw's F0 or Adger's "special kind of category" that each hierarchy "is rooted in" (p. 197)—I simply assign index 1 to whatever category happens to have a smallest scope in its c-sequence.[75] Third, as we will see in §2.4, the mergeme-based template makes available two theoretically special types of syntactic categories by incomplete specification (i.e., defective and acategorial categories), so nonequipollence has a fundamental place in my theory, much like the place of unvalued features in standard minimalism. In fact, my incompletely specified mergeme and Chomsky's unvalued feature respectively represent nonequipollence in the categorial and the noncategorial facets of syntactic categories.

Fourth, I aim to use the featural metalanguage in a coherent way. There are two potential points of confusion in (26). First, the relation between the natural number values and their base feature(s) is left vague in Grimshaw (2005) (with a generic F) and undefined in Adger (2010) (it is unclear what it means for an index to be the value of a categorial feature; e.g., [T: 8], p. 198). In particular, Adger's (2010) practice is inconsistent with Adger & Svenonius (2011), where [ATT: VAL] encodes the feature class–feature relation (apparently T is not the class for 8). Second, since Adger's indexing is based on the cartographic ordering of categorial features, having both Cat and N in K is redundant, as the categorial features fall in exactly the same order even without the indices. My [POS: `nat`] notation (e.g., [POS: 3]) avoids both notational problems. On the one hand, it explicitly lets the numerical indices specify positions by making them values of a POS attribute, which by the FTC corresponds to the sort-type `nat` of natural numbers and hence well reflects the feature class–feature relation (e.g., [POS: 3] $\leftrightarrow$ $\texttt{nat} \sqsubseteq 3$).[76] On the other hand, since the "categorial features" and the numerical indices do separate jobs in my template (also in Grimshaw's), there is no information redundancy.[77] Last but not least, I adopt the weak $\leq$ (i.e., the order-theoretic standard) instead of the strict $<$ as the partial order relation, for $\leq$ but not $<$ gives us reflexivity, a property that plays an indispensable role in connecting the various c-sequences as well as granularity levels (see Chapter 6 for detail).

2.4 Flexibility

The aim of this dissertation is to explore the formal flexibility in the SCS, under the idea that UG only provides templatic guidance for syntactic category formation and organization but leaves many other issues open to be shaped by non-UG factors, such as the "second" and "third" factors

[75] I will still use big Xs to exemplify but I use them more as placeholders than with firm commitment.

[76] Recall from §2.2.3.1, note 38, p. 18 that the FTC associates generic supertypes with specific attributes.

[77] Neither my CAT values (i.e., c-sequence names) nor Grimshaw's categorial specifications (i.e., the Chomskyan lexical categories) are the same as Adger's categorial features (which are basically the cartographic heads).

in Chomsky (2005).[78] So far I have focused on the template for individual categories, but as we will see, each layer of theoretical abstraction has its own template. The poor-UG view offers new perspectives on language development and variation. In relation to syntactic categories, it has been insightfully explored in Wiltschko (2014), Ramchand & Svenonius (2014), and Biberauer (2017a) inter alia. I will build on previous research results and embed them in a larger categorial universe. The overarching theme of this dissertation is stated in the following slogan:

Slogan 2. Template and flexibility go hand in hand in every corner of the SCS.

In this section I give an outline of the types of categorial flexibility that will be covered in the upcoming chapters. In a nutshell, I distinguish three broad levels of concern and five layers of abstraction:[79] (i) individual level, for the formation of single syntactic categories (first layer) and their organization into single c-sequences (second layer); (ii) global level, for the interconnection between different c-sequences within one granularity level (third layer) and across different granularity levels (fourth layer); and (iii) supraglobal level, for the use of various syntactic categories to generate utterances (fifth layer). In Sections 2.4.1–2.4.3 I go through these in turn.

2.4.1 Individual level

Individual-level flexibility arises from two sources: (i) alteration to the mergeme-based template; (ii) alteration to the c-sequence format. These happen at the first two abstraction layers of the SCS: that of single categories and that of single c-sequences.

2.4.1.1 First abstraction layer

2.4.1.1.1 Underspecified mergeme and defective category First, we can have an underspecified or unvalued mergeme, notated as ⟨CAT:_, POS:_⟩ or [MERGEME: [CAT:_, POS:_]] or simply *u*MERGEME. Since it still carries the two appropriateness conditions, *u*MERGEME may become valued in the course of derivation. And since the values of CAT and POS determine a category's first-Merge position, a *u*MERGEME-bearing category—which I will call a *defective category* and notate as Cat—has no predetermined first-Merge position; in other words, its first-Merge position is flexible. However, this flexibility is only relative. Since valuation copies the values of CAT and POS from the value provider (goal) to Cat (probe), upon valuation the latter obtains a first-Merge position identical to the former's; and since the first-Merge position of a category is the position where it first enters the derivation, a derivational conflict would result if Cat is first-merged at

[78] According to Chomsky (2005: 6, 9), three "factors of language design" enter into the "growth of language in the individual": (i) genetic endowment (UG), (ii) experience (primary linguistic data or PLD, Biberauer 2017a), and (iii) principles not specific to the language faculty (principles of data analysis, efficient computation, etc.).

[79] So I am assigning the two terms *level of concern* and *layer of abstraction* different senses, unlike Martin (1986) (see §2.2.4) and many others, who treat them as synonyms.

any other position than where its value-provider is first-merged.[80] Hence, Cat can only be legitimately merged with and valued by a nondefective category (i.e., a category with a completely specified mergeme), after which it becomes assimilated to that category. This chameleon-like characteristic makes the defective category potentially useful in featurally deriving adjunction. In Chapter 3 I explore this potential by two case studies, one on modifier-head compounds across languages and the other on sentence-final particles in Chinese.

2.4.1.1.2 Unspecified mergeme and acategorial category Second, we can also have an unspecified mergeme, notated as ⟨⟩ or [MERGEME:_], which lacks a `mergeme`-typed value altogether.[81] Since the mergeme is the label-provider, a mergemeless category cannot label (or project/head) phrases. I will call such a syntactic category an *acategorial category*. It cannot be merged according to a scopal position in some c-sequence but must seek alternative ways to join the derivation. I further distinguish two types of acategorialness:

- *Roots* (e.g., $\sqrt{\text{CAKE}}$, $\sqrt{\text{EAT}}$) are acategorial categories with no compositional meanings (though they do have lexical meanings); in other words, they have no taxonomic SCDPs. A root can only feed its (idiosyncratic) interpretation to the interface via categorization. My view on root categorization is much more liberal than that in previous studies (especially in DM); I do not constrain root-level derivation in ways not deducible from the mergeme-based template (including what can/cannot serve as a categorizer). I will present this generalized root syntax in Chapter 4, situated in a comprehensive study of the semi-functional items in Chinese.

- *Logical operators* (e.g., ∧ 'and', ∨ 'or') are acategorial categories with non-mergeme-based compositional meanings (encoded as purely taxonomic SCDPs). The logical functionality is all the semantic content in such categories. I will not go into detail about logical operators in this dissertation (except in §4.2.4) but leave them to future research.

We can distinguish the two types of acategorialness based on the presence/absence of taxonomic SCDPs, which can be described by a feature [±TAX]: roots are [−TAX] while logical operators are [+TAX]. By *acategorial* I only mean that roots and logical operators have no place in c-sequences; that is, they cannot be defined from the mergeme-based perspective. As a result, they are not subject to scopal ordering (e.g., no root is a born c-commander of another root in the same way as C c-commands T) and do not vary across granularity levels (e.g., AND functions in the same way at Cart and CFC). However, being acategorial does not imply being asyntactic (i.e., disqualified as syntactic categories). Recall from Figure 2.5 that there are three perspectives to define syntactic categories. Since logical operators are [+TAX], they are still syntactic categories from

[80] Recall from §2.3.1.2 that I conceive a derivation as a formal proof; hence, there is no problem in constraining the first-Merge position of a category by where it must end up being by the end of the derivation. It simply means that among the many possible premise settings, this is the one that makes the proof work.

[81] The attribute MERGEME itself has no sort-type (it is just a name tag).

the taxonomic perspective; since both roots and logical operators can still merge, they are both syntactic categories from the combinatorial perspective. In sum, the mergeme-based template yields the following classification of syntactic categories.

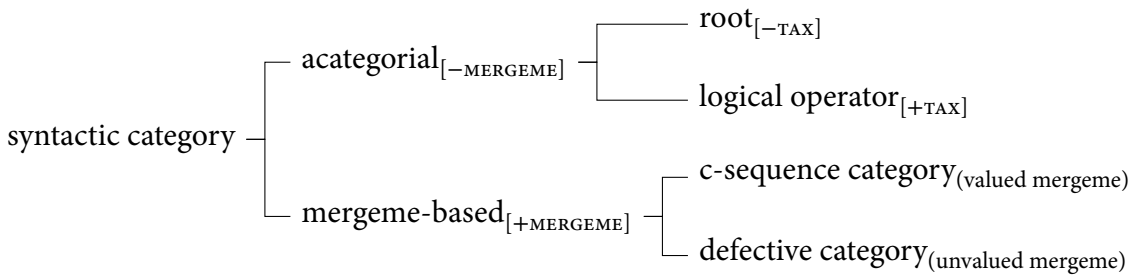


Figure 2.7 Classification of syntactic categories

2.4.1.1.3 Antidecompositional category A flexibility type disallowed by the mergeme-based template is multimergemic specification, which places more than one `mergeme`-typed value in the feature set of a syntactic category (≠ lexical item).[82] Thus, when a mergeme-based category appears to have more than one minimal interpretive unit, it either is not a truly minimal category (but can be decomposed) or has a mixture of mergeme- and non-mergeme-based interpretive units. What my template rules out are categories that are both minimal and have more than one mergeme-based interpretive unit; I call these *antidecompositional categories* for they seemingly have internal structures but cannot be given decompositional analyses. Simply put, such categories are ruled out because the mergeme determines its host category's first-Merge position, and a category cannot have more than one first-Merge position in any derivation (more exactly any derivational layer; see Chapter 4). More rigorously, if there are multiple mergeme values within a category's feature set, that category would correspond to multiple c-sequence elements; however, since each c-sequence element is by definition a category, that would make the category in question simultaneously one category and multiple categories at the same granularity level, which is a contradiction. There is no positive case of antidecompositional categories, but next I discuss two pseudo cases contextualized at the second abstraction layer.

2.4.1.2 Second abstraction layer

Some syntactic categories apparently have more than one mergeme-based interpretive unit, but careful examination reveals that they are not antidecompositional but can perfectly fit into the single-mergeme template. Behind such *pseudo-antidecompositional* or *PAD categories* is flexibility at the second abstraction layer (i.e., that of single c-sequences).

[82] Here it is important to distinguish syntactic categories (abstract constructs, e.g., T, C) from lexical items (concrete entities, e.g., *did*, *if*). A lexical item may certainly contain (or spell out) multiple categories (and hence multiple mergeme values), but syntactic categories qua abstract constructs are necessarily discrete; e.g., while *would* may contain both T and Mod, the categories T and Mod themselves are not defined by the same feature set.

2.4.1.2.1 Valued category First consider cartographic categories like T_{Past}, $T_{Anterior}$, $Asp_{Habitual}$, and $Asp_{Repetitive}$ (via Rizzi & Cinque 2016: 150). These are seemingly antidecompositional as the mergeme-based template can define both categories like T/Asp and categories like Past/Habitual, and they belong to two separate granularity levels (e.g., A and Cart). Yet inasmuch as the cartographic hierarchies within one proposal all belong to the same granularity level, labels like T_{Past} and $Aspect_{Habitual}$ cannot simultaneously reference different granularity levels, which suggests that one of the two components in the label of a *valued category* must be merely decorative and hence dispensable. In the above examples the decorative component is the base (e.g., T, Asp). In a sense the composite labels record the history of category subtyping; for example, $Asp_{habitual}$ and $Asp_{repetitive}$ show that the two (sub)categories Habitual and Repetitive are resulted from subtyping Asp. This is reminiscent of the subtype disjunction definition for sort-types in §2.2.3.1, which corresponds to the [ATT: VAL] notation by the FTC (e.g., [ASP: HABITUAL/REPETITIVE ...] ↔ `asp` ≔ `habitual` ∨ `repetitive` ...). A condition imposed by cartography on such category subtyping is that the subcategories occupy separate c-sequence positions. So, as long as a valued category is associated with a unique c-sequence position, it is encodable in a single mergeme value and does not count as real antidecompositionality. I will not further investigate valued categories in this dissertation (except in §5.2.2 in comparison with flavored categories) since they are already well built into the machinery of cartography.

2.4.1.2.2 Flavored category What I will investigate in more detail is another subtype of PAD category often informally termed *flavored category*, notably in DM-based works on little *v* categories (e.g., v_{CAUSE}, v_{BECOME}, v_{DO}; Cuervo 2003, Folli & Harley 2005). Two oddities make such categories neither the same as valued categories nor properly handleable by cartography. First, the base-subscript relation in a flavored category is different from that in a valued category; the latter is just the attribute-value relation while the former is not. Thus, while [ASP: HABITUAL] is a reasonable attribute-value pair, [V: CAUSE] is not, for it cannot be given a feature class–feature interpretation (just like Adger's [T: 8] in §2.3.2). Intuitively, the subscript "flavors" are really values of a separate event-type-related attribute (e.g., EVENT). Second, although the flavored little *v*s are often presented in a stacked fashion (e.g., in Cuervo 2003), whether there is any fixed scopal ordering among the event types they encode is unresolved. For instance, in Ramchand (2008) the two event types process and state are not ontologically ordered, and the derivational ordering Init–Proc–Res only emerges configurationally: a state merged above a process is interpreted as an initiating or causing subevent (labeled Init), and a state merged below a process is interpreted as a result subevent (labeled Res).[83] Similarly, my study of Chinese resultative verbal compounds in Chapter 5 shows that, in a flavor-based analysis, the two flavors DO (activity) and BE (state) may be flexibly stacked, as in Mandarin *kū-shī* 'cry$_{DO}$-wet$_{BE}$; sb. cries and consequently sth. (e.g., a handkerchief) ends up being wet' and *téng-kū* 'painful$_{BE}$-cry$_{DO}$; sb. is in a painful state and con-

[83] Baker (2018: 6) modifies Ramchand's theory and proposes that the two stative positions are Res and "Volition" instead of Res and Init(iation).

sequently cries'. Such base-generated flexible ordering within the same language variety clearly violates the cartographic tenet of a rigidly ordered functional hierarchy. In Chapter 5 I will explain the oddities of flavored categories in terms of flexibility at the second layer of abstraction, viewing flavored categories as incomparable elements in a poset, as in Figure 2.8.

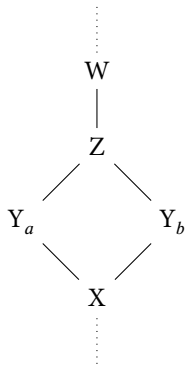


Figure 2.8 A c-sequence with two flavored categories

The flavors *a* and *b* in the Hasse diagram in Figure 2.8 are SCDPs but not mergeme-based; they are purely taxonomic instead. As such, Y_a and Y_b have identical mergeme values but differ in their nonmergeme features, and so flavoring becomes subtyping, though it differs from the subtyping underlying valued categories in that flavor features are not scopally ordered. In sum, I take the template at the c-sequence abstraction layer as the scope-based partial order ($\leq$), which flexibly allows both comparable and incomparable categories. That said, however, a major result of Chapter 5 is that incomparable categories, albeit legal, are "bizarre" due to non-UG factors, so flavored categories should be used cautiously in syntactic analysis.

2.4.2 Global level: Third and fourth abstraction layers

Global-level flexibility goes beyond the mergeme-based template for single categories and the order-theoretic template for single c-sequences. It brings in higher-level categorial constructs like granularity levels and their interconnections. Accordingly, the global-level template-flexibility pairs have a much more abstract nature, which feature theory alone cannot adequately describe. Therefore, I will introduce a more advanced toolkit from mathematics, Category theory,[84] to examine and formalize structures and patterns at the remaining abstraction layers. As Fong & Spivak (2018: iii) point out, Category theory is "unmatched in its ability to organize and layer abstractions" and "to find commonalities between structures of all sorts." Thus, I treat feature theory and Category theory as two halves of a complete metalanguage covering the entire ladder

[84] As mentioned in note 2, p. 2, I capitalize Category-theoretic terms when there are homonymous terms in linguistics; e.g., *Category* is a mathematical term (adj. *Categorical*, v. *Categorify*), while *category* is a linguistic term (adj. *categorial*, v. *categorize*).

of abstraction for the SCS. While a step-by-step introduction of Category theory will only be presented in Chapter 6, below I list some main points that we can anticipate there.

- A single c-sequence as a poset can be viewed as a *Category*. The syntactic categories in the c-sequence are *objects* of the Category and instances of the c-sequence partial order are *arrows* of the Category.
- A single granularity level as a set of c-sequences can be viewed as a Category of posets. The individual c-sequences of the granularity level are objects of the Category and the inter-c-sequence monotone functions are its arrows. This Category can additionally be associated with a universal spine (Wiltschko 2014), which serves as a *zero object* in it.
- The zero object is important because it is connected to all c-sequences at the same granularity level via an *Adjunction*. This adequately formalizes the intuitive parallelism across c-sequences of different major parts of speech, though it also reveals that this parallelism is not directly established but mediated by a shared universal spine.
- All theoretically possible granularity levels together form a Category, whose arrows are various inter-granularity-level *Functors*. In this Category we can formalize the inheritance relation across granularity levels resulted from category subtyping. In particular, ○ serves as a *terminal object* in this Category, which underlies all granularity levels and can be identified with the defective category studied in Chapter 3.

At the abstraction layer of a single granularity level (i.e., the third layer), the template is its associated universal spine and the flexibility space involves all possible c-sequences that can be instantiated from that spine. Next, at the abstraction layer of all granularity levels (i.e., the fourth layer), the template is the abstract mechanism of category subtyping, perhaps together with the crucial first step of category formation (from the absence to the presence of syntactic categories), and the flexibility space is the set of all possible granularity levels realizable in human language. Note that when it comes to implementing these abstract templates, both the instantiations of the universal spine and the results of category subtyping are shaped by specific taxonomic features (i.e., features that are salient in specific language varieties). This reflects a general division of labor between Category theory and feature theory in describing the SCS: Category theory is good at depicting *intercategorial* connections, but it cannot peek inside single categories (as Categorical objects are opaque); feature theory is good at fleshing out *intracategorial* content, but it cannot establish ontological connections independent of concrete linguistic entities (as features are specified for concrete entities).

2.4.3 Supraglobal level: Fifth abstraction layer

All previous abstraction layers reflect the ontological organization of *mergeme-based* syntactic categories. However, recall that the featural template in (19) also allows acategorial categories,

which fall outside the foregoing layers but are still within the SCS. To incorporate acategorial categories into the picture we need one final and highest abstraction layer: that of the entire SCS.

In Category-theoretic terms, the entire SCS is a Category **Syn** containing all syntactic categories of all sizes—not only minimal categories like V/T but also phrasal categories like VP/TP—regardless of their ontological taxonomy; arrows in this Category are not intercategorial relations either but describe derivational processes. In other words, at this level of concern, the focus of attention is no longer how syntactic categories are ontologically connected to one another but how they are combined into meaningful utterances. Accordingly, necessary equipment at this layer is a systematic map between syntactic structures and semantic values, which is standardly Categorified as an interpretation Functor $\Phi : \mathbf{Syn} \rightarrow \mathbf{Sem}$ from the syntax Category to the semantics Category (Lambek 1988). While this purely combinatorial layer attracts the mainstream application of Category theory in linguistics (see Chapter 6 for references), it is not my focus in this dissertation and only mentioned for the sake of completeness. Its template is none other than the basic combinatorial operation (e.g., Merge in minimalism, Unify in HPSG)—perhaps together with a small set of derivational rules (e.g., Agree, Label, etc. in minimalism; Axiom, Cut, etc. in categorial grammars)—and its flexibility space involves the familiar parametric settings that determine how exactly these general syntactic rules are carried out and thereby give rise to various grammatical systems.

To conclude this chapter, I summarize the five template-flexibility pairs I have proposed for the SCS in Table 2.1. Remember that the templates are assumed to be UG-given, and the flexibility types emerge in places UG does not care about.

Table 2.1 Templates and flexibility types in the syntactic category system

Level of concern	Abstraction layer	Template	Flexibility
Individual	single category	mergeme-based feature set	incomplete mergemes
	single c-sequence	scope-based partial ordering	incomparable categories
Global	single granularity level	universal spine	varied c-sequences
	all granularity levels	category subtyping	varied granularity levels
Supraglobal	entire SCS	syntactic rules	parametric variation

Chapter 3

Defective category and feature-based adjunction

3.1 Defective category theory

In Chapter 2 I delineated the SCS as a ladder of abstraction consisting of five layers, each associated with a UG-given template and a flexibility space. In this and the next chapter, I examine the first abstraction layer (i.e., that of single categories) in more detail. The template at this layer is the mergeme-based definition of syntactic categories, as repeated in (1).

(1) syntactic category$_{\text{LANG, GRAN}}$:= [⟨CAT: cat, POS: nat⟩, NM]

This template, or more exactly its mergeme part ⟨CAT: cat, POS: nat⟩, allows two types of flexibility, respectively characterized by an underspecified mergeme ⟨CAT:_, POS:_⟩ and an unspecified mergeme ⟨ ⟩. I begin with the former in this chapter and leave the latter to the next. I call the category defined by the underspecified mergeme the *defective category* and notate it as Cat.[1]

(2) Cat$_{\text{LANG}}$:= [⟨CAT:_, POS:_⟩, NM]

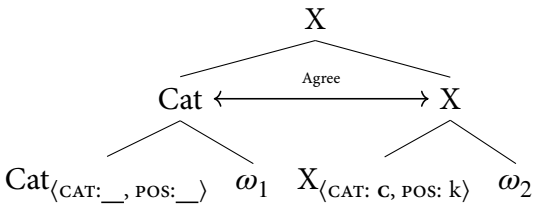
This chapter is organized as follows. In the rest of §3.1 I lay out the specifics of the defective category theory, including its derivational schema and its correlation with (a subtype of) adjunction. In Sections 3.2–3.3 I further illustrate this correlation by two case studies, one on modifier-head compounds (MHCs) across languages and the other on sentence-final particles (SFPs) in Chinese. In §3.4 I summarize the main results.[2]

[1] I leave out GRAN in (2) because Cat, lacking a mergeme value, has no fixed granularity level. In this sense it is similar to acategorial categories, though Cat can be incorporated into a granularity-specific functional hierarchy (via Agree, see below) whereas acategorial categories cannot. LANG, on the other hand, cannot be omitted since different languages may develop different subtypes of defective category (e.g., by different nonmergeme features).

[2] Earlier versions of some material in this chapter can be found in Song (2016b, 2017a).

3.1.1 Derivational schema

In principle Cat can merge with any nondefective category and get its mergeme valued via Agree. I display its derivational schema in Figure 3.1 in two formats, a syntactic tree and a proof tree, in line with my conception of syntactic derivations as formal proofs in Chapter 2. While the two formats are equivalent, the proof tree contains more derivational detail and can more clearly represent multiple independently derived premises (or lemmas), especially when parallel derivational layers are involved. I will come back to the issue of layered derivation in Chapter 4.



(a) Syntactic tree

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\mathrm{Cat}_{\langle \text{CAT:}_,\ \text{POS:}_\rangle}}\,\text{axiom} \qquad \omega_1}{\{\mathrm{Cat}_{\langle \text{CAT:}_,\ \text{POS:}_\rangle}, \omega_1\}}\,\text{merge}}{{}_{\mathrm{Cat}}\{\mathrm{Cat}_{\langle \text{CAT:}_,\ \text{POS:}_\rangle}, \omega_1\}}\,\text{label} \qquad \dfrac{\dfrac{\dfrac{}{\mathrm{X}_{\langle \text{CAT: c, POS: k}\rangle}}\,\text{axiom} \qquad \omega_2}{\{\mathrm{X}_{\langle \text{CAT: c, POS: k}\rangle}, \omega_2\}}\,\text{merge}}{{}_{\mathrm{X}}\{\mathrm{X}_{\langle \text{CAT: c, POS: k}\rangle}, \omega_2\}}\,\text{label}}{\{{}_{\mathrm{Cat}}\{\mathrm{Cat}_{\langle \text{CAT:}_,\ \text{POS:}_\rangle}, \omega_1\}, {}_{\mathrm{X}}\{\mathrm{X}_{\langle \text{CAT: c, POS: k}\rangle}, \omega_2\}\}}\,\text{merge}}{\dfrac{\{{}_{\mathrm{Cat}}\{\mathrm{Cat}_{\langle \text{CAT: }\underline{\text{c}}\text{, POS: }\underline{\text{k}}\rangle}, \omega_1\}, {}_{\mathrm{X}}\{\mathrm{X}_{\langle \text{CAT: c, POS: k}\rangle}, \omega_2\}\}}{{}_{\mathrm{X}}\{{}_{\mathrm{Cat}}\{\mathrm{Cat}_{\langle \text{CAT: }\underline{\text{c}}\text{, POS: }\underline{\text{k}}\rangle}, \omega_1\}, {}_{\mathrm{X}}\{\mathrm{X}_{\langle \text{CAT: c, POS: k}\rangle}, \omega_2\}\}}\,\text{label}}\,\text{agree}
$$

(b) Proof tree

Figure 3.1 Derivational schema for the defective category

The derivation in Figure 3.1 has four premises. In this particular case there are two axioms (Cat, X), X an arbitrary nondefective category (the *k*th element in the c-sequence **c**), and two prederived terms (ω_1, ω_2).[3] But as we will see in §3.2, X may also be prederived and ω_1/ω_2 may also be axiomatic, whence emerge several variants of the schema in Figure 3.1. From the premises, Cat and X respectively merge with ω_1 and ω_2 and label the merged products. Examining the final conclusion of the derivation, we see that the SO headed by Cat (i.e., {Cat, ω_1}) is effectively an adjunct of its sister SO (i.e., {X, ω_2}). To facilitate presentation, I track the mergeme values of Cat and X by right subscripts; the left subscripts (i.e., the labels) are also merely expository. Following the inclusiveness condition in Chomsky (1995) and the labeling algorithm in Chomsky (2013, 2015), I do not assume labeling to introduce extra ingredients (e.g., indices, bar levels) to the derivation but merely view it as minimal search,[4] which locates the label-provider

[3] By *axioms* I mean primitive terms that have just left the lexical array (LA)—these are the minimal categories (i.e., heads) of the derivation. All other terms, being derived, are nonminimal categories (i.e., phrases).

[4] Chomsky (2013: 43): "The simplest assumption is that [labeling algorithm] is just minimal search, presumably appropriating a third factor principle, as in Agree and other operations."

from inside an SO rather than add a tag to it from the outside. In addition, viewing derivations as formal proofs, I abstract away from the issue of "timing"; that is, the derivation in Figure 3.1b is logical rather than temporal. Thus, all the labeling steps may happen at once (if they correspond to psychological processes at all).[5]

I have three further comments on defective category agreement. First, viewing labeling as minimal search gives us the freedom to let Agree proceed via labels. This is also the standard conception in Chomsky (2001: 5), where the Agree-activating probe is a label rather than an entire SO, as the label is "the only element" that is "immediately accessible" to syntactic operations. Zwart (2006) further argues that Agree always happens between two sister phrases (instead of a head and a phrase), which in contemporary terms are just two phrasal labels. Thus, in Figure 3.1 it makes no difference to say "Cat agrees with X" or "$\text{Cat}_{\langle \text{CAT:_, POS:_} \rangle}$ agrees with $\text{X}_{\langle \text{CAT: x, POS: k} \rangle}$," as the information in a label and that in its provider are always in sync. Second, deviating from the standard minimalist conception, the agreement in Figure 3.1 is between two inherent and interpretable features,[6] not between an inherent and an optional feature or an interpretable and an uninterpretable feature as defined in (Chomsky 1995 et seq.)—nor does it rely on φ features. I do not consider this deviation a flaw; on the contrary, inasmuch as the defective category theory is plausible, it gives us a reason not to confine Agree to uninterpretable probes or φ features[7] but to view it as a general, abstract operation. Just as we do not want to exclusively associate Merge with selected classes of SOs/features, nor do we want to do so for Agree.[8] Third, in standard minimalism the probing feature, which is uninterpretable and whose value is redundant, must be deleted upon valuation; by contrast, in Figure 3.1 no feature deletion happens, because there is no uninterpretable feature to begin with.[9] In fact, as we will see, the derivationally obtained mergeme value on Cat serves the important purpose of unifying the adjunct and its host into a single mergeme-based category both in syntax and in semantics,[10] so it is anything but redundant.

The potential explanatory power of the defective category theory is twofold. First, by letting ω_1, ω_2, and X vary, we can derive various adjunction scenarios in purely featural terms (i.e., without techniques like pair-Merge). Second, the nonmergeme features on Cat can be used to encode certain parametric or idiosyncratic variations in adjunction (e.g., varied word orders

[5] In my opinion, timing-related concerns, including interface operations like Spellout and Transfer, are better addressed at a different analytical level, one where procedural descriptions like "the first/second phase" and "the beginning/end of a cycle" make more sense. A good part of current minimalism—e.g., phase theory, labeling algorithm—seems to be motivated by performance-oriented concepts (e.g., computational efficiency, memory load; see Neeleman & Van de Koot 2010: 187), which do not arise at the purely logical level I am concerned with.

[6] The unvalued mergeme on Cat is interpretable in that once valued it can pick out a category from a c-sequence just as any nondefective category can.

[7] Neither of the two points is new. See Pesetsky & Torrego (2007) for a similar proposal to untie feature interpretability from feature valuation, and see Miyagawa (2017) for a detailed investigation of non-φ-based agreement.

[8] I am not denying the importance of uninterpretable probes or φ features in agreement but merely proposing to view them as a particular instantiation of Agree rather than the whole story.

[9] I remain agnostic about the destiny of uninterpretable probes. They may be either deleted as in standard minimalism or "coalesced" as in feature-sharing theories (e.g., Frampton & Gutmann 2000, 2006).

[10] The semantic effect of defective category agreement is exactly the predicate composition operation deemed as an interface condition requirement in Chomsky (2004). See §3.2 for elaboration.

between adjuncts and hosts). Due to space limitations, in this chapter I will only present two case studies, respectively on modifier-head compounds and sentence-final particles. I make this choice because I do not think what are traditionally classified as adjuncts are all in a theoretically homogeneous class.[11] At the current stage of my research, the defective category theory seems especially suitable for compound-like adjunction structures, where the adjunct and its host form a tight interpretive unit. The two cases I study respectively represent the smallest (i.e., word-level) and the largest (i.e., utterance-level) such structures.

3.1.2 Approaches to adjunction

The conception of defective category is not motivated by the need to analyze adjunction but deduced on purely logical grounds: if we define syntactic categories by the mergeme-based template in (1), then there is bound to exist a well-defined syntactic category taking the form of (2); it is just a bonus that this category brings along a new perspective on adjunction. However, since there are dedicated theories of adjunction, in this section I review some of them and specify how the defective category theory is similar/different.

3.1.2.1 Axiomatic approach: From pair-Label to pair-Merge

Chomsky (1995: 248) distinguishes adjunction from nonadjunction (or substitution) by the shape of its label. Given two SOs α and K, their merger qua substitution yields (suppose K projects) {H(K), {α, K}}, where H(K) is the head/label of K; and their merger qua adjunction yields (by adjoining α to K) {⟨H(K), H(K)⟩, {α, K}}, which is a "two-segment category" with an ordered-pair label. The term *segment* is inherited from Chomsky (1986), where it is presented as a development of an idea in May (1985). According to Chomsky (1986: 7), in the structure $[_\beta\ \alpha\ [_\beta\ \ldots\ ]]$, with α adjoined to β, β consists of two segments, and a category is dominated by β only if it is dominated by both segments (so α is not dominated by β). Chomsky does not give *segment* a rigorous definition.[12] The original passage in May (1985) is more lucid, though May speaks of "member/projection" instead of "segment/category":

> [T]he occurrence of a *projection* [is] made up of a set of occurrences of nodes that are featurally nondistinct (that is, identical with respect to syntactic features, bar level, index, etc.). It is these nodes, taken collectively, that constitute the membership of a projection. In effect, this characterization claims that the structural effect of (Chomsky)-adjunction is to create multimembered projections. (May 1985: 56–57)

[11] Theresa Biberauer (p.c.) points out that there are numerous empirical facts supporting this; for instance, not all adverb(ial)s can be put into Cinque's (1999) fixed hierarchical slots, there are central vs. peripheral adverbials as Haegeman's work shows, and there is also a distinction between agreeing vs. nonagreeing adjuncts.

[12] He apparently has something along the lines of "occurrence of a category" in mind (Chomsky 1986: 92, note 10), but that too is only mentioned in passing.

Why does adjunction create multimembered projections / multisegmented categories? Both May and Chomsky accept this as an axiom, but can we reduce it to a theorem? This may not make a difference empirically, but from a theoretical angle having a smaller axiom set makes the framework more parsimonious and hence more minimalist in spirit. The defective category theory provides one way to do this, by reducing the featural nondistinctness of projection members to a consequence of Agree. In Figure 3.1, upon agreement $\{\{\mathrm{Cat}_{\langle \text{CAT: } \underline{\text{C}}, \text{ POS: } \underline{\text{k}}\rangle}, \omega_1\}, \{\mathrm{X}_{\langle \text{CAT: C, POS: k}\rangle}, \omega_2\}\}$ ($\star$) can be labeled by the shared feature $\langle$CAT: C, POS: k$\rangle$, which is exactly the mergeme specification of X. Thus, the label of ($\star$) is identical to that of $\{\mathrm{X}_{\langle \text{CAT: C, POS: k}\rangle}, \omega_2\}$, which makes ($\star$) a multimembered projection.

In post-2000 minimalism, pair-Label is changed into pair-Merge, which takes two SOs α, β as input and returns an ordered pair $\langle \alpha, \beta \rangle$ as output. Here the label of an adjunction structure is no longer a pair, as expounded in the following quotes:

> Adjunction has an inherent asymmetry .... Exploiting that property, let us take the distinction between substitution and adjunction to be the (minimal) distinction between the set $\{\alpha, \beta\}$ and the ordered pair $\langle \alpha, \beta \rangle$ ... [T]he adjoined element α leaves the category type unchanged: the target β projects. Hence, adjunction of α to β forms $\mathrm{K} = \{\gamma, \langle \alpha, \beta \rangle\}$, where γ is the label of β. (Chomsky 2000: 133)

> [Adjunction] takes two objects β and α and forms the ordered pair $\langle \alpha, \beta \rangle$, α adjoined to β. ... [W]e might intuitively think of α as attached to β on a separate plane, with β retaining all its properties on the "primary plane," the simple structure. (Chomsky 2004: 117–118)

The pair-Label theory and the pair-Merge theory mainly differ in where they encode the "inherent asymmetry" of adjunction and, accordingly, in which operation they leave theoretically simple. Pair-Label complicates labeling and leaves Merge simple (i.e., there is only one type of Merge), whereas pair-Merge complicates Merge and leaves labeling simple (i.e., all labels are the result of minimal search).[13] I illustrate this difference in (3).

(3) Given two SOs α and β, adjoining α to β returns

a. $\{\langle \mathrm{H}(\beta), \mathrm{H}(\beta) \rangle, \{\alpha, \beta\}\}$ (by pair-Label)

b. $\{\mathrm{H}(\beta), \langle \alpha, \beta \rangle\}$ (by pair-Merge)

It is not exactly clear what motivated the perspective shift from Chomsky (1995) to Chomsky (2000), though the pair-Merge theory seemingly has more advantages.[14] For one thing, it makes the labeling procedure consistent and hence is more compatible with the recent proposal in

[13] Pair labels cannot be obtained by minimal search, as they are not "term[s] of the structure[s] formed"; i.e., not "identical" to any head (Chomsky 1995: 248).

[14] See Hornstein (2009: 86) for further discussion on the problems of pair labels.

Chomsky (2013, 2015). For another, pair-Merge is more minimalist in spirit as it conforms to the "interface condition":

> [T]he strong interface condition … requires sufficient diversity at SEM. Possibly richness of expressive power requires an operation of predicate composition: that is not provided by set merge [but] is the essential semantic contribution of pair merge. (Chomsky 2004: 118)

That said, pair-Merge is just as axiomatic as pair-Label, for it cannot be reduced to anything else. By comparison, the defective category theory leaves both labeling and Merge simple and derives the inherent asymmetry of adjunction via the independently motivated axiom Agree.

3.1.2.2 Derivational approach: Unlabeled merger and dangling adjunct

There are several alternatives to the standard minimalist theory of adjunction. I review two derivational approaches in this section, both based on labeling, and a lexicalist approach in the next. I begin with the theory in Hornstein & Nunes (2008) and Hornstein (2009). In this theory, when two SOs "concatenate,"[15] the product may either be labeled or not, and when it is not, we get an adjunction structure. To illustrate, in (4) the concatenation of X and Y is labeled while that of X^Y and Z is not, so the latter is adjunction.

(4) $[_X$ X^Y]^Z (Hornstein & Nunes 2008: 66)

In this way Hornstein & Nunes are able to derive a number of characteristic behaviors of adjuncts, such as optional stranding and island effects. I abstract away from the empirical side of their theory and only examine its logic. An examination reveals two problems. First, in an unlabeled concatenation like (4) it is hard to formally determine which SO is the host and which is the adjunct. Hornstein & Nunes choose to let $[_X$ X^Y] further concatenate, as in (5), and thereby specify Z as the adjunct (called "dangling off").

(5) $[_W$ W^$[_X$ X^Y]]^Z

But there is no reason to deduce (5) from (4) other than that we want Z to be the adjunct. Logically speaking the two alternative derivational paths in (6) are also licit.

(6) a. $[_X$ X^Y]^$[_Z$ Z^W]

b. $[_W$ W^$[_X$ X^Y]]^$[_Z$ Z^V]

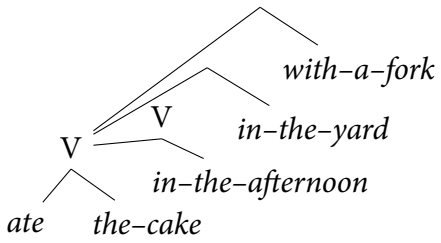
Second, a fundamental theorem in Hornstein & Nunes' theory is that in an unlabeled concatenation like (4) both components may freely concatenate with further SOs, which is why $[_X$

[15] Concatenation is the basic combinatorial operation in this system.

X^Y] can concatenate with W in (5) even though it is already concatenated with Z. Moreover, such reconcatenation can happen multiple times. Consequently, multiple adjuncts may "dangle off" the same host, as in (7a). I also display this structure in tree format in (7b).

(7) a. [$_V$ [$_V$ *ate^the–cake*]^*in–the–afternoon*]]
^*in–the–yard*
^*with–a–fork* (Hornstein & Nunes 2008: 67)

b.



Hornstein & Nunes (2008) motivate reconcatenation based on two assumptions: (i) concatenation is defined on "atoms," which are either LIs or labeled concatenations (p. 65); (ii) labeling prevents all but a head (i.e., label) to be seen from outside a labeled concatenation (p. 79). From these they conclude that components of an unlabeled concatenation "can still be input to further concatenation" (ibid.). This amounts to saying that "If a concatenation c is labeled, then its components cannot be input to further concatenation." implies "If a concatenation c is not labeled, then its components can be input to further concatenation." If we use P and Q to denote two predicates is-labeled and components-can-be-input-to-further-concatenation, then Hornstein & Nunes' reasoning can be written as $P(c) \rightarrow \neg Q(c) \Rightarrow \neg P(c) \rightarrow Q(c)$. No established rule of inference supports this implication. Besides, the possibility of reconcatenation also makes Hornstein & Nunes' theory incompatible with standard minimalism, where remerge or multidominance is prohibited:

> [T]here are no operations "form copy" or "remerge," just simple Merge. … The concepts of multidominance, "late Merge," and some others postulate an extension of Merge. (Chomsky 2013: 40)

Of course, to what extent reconcatenation counts as multidominance depends on how different the concatenation operation is from Merge and crucially on whether concatenation creates sets. Neither question has an explicit answer. In Hornstein (2009), where the framework in Hornstein & Nunes (2008) is further enunciated, the author initially describes Merge as "a species of concatenation" (p. 55) but later uses concatenation as a synonym of Merge, as in "concatenation (aka Merge)" (p. 89). And the following note on the mathematical nature of concatenation sounds just like set formation (if no further complexity is intended):

> I assume that one can have order free concatenation so that A^B and B^A are indistinguishable. One can think of this as concatenation in a 2-space, where the con-

> catenation of A^B does not determine whether A precedes or follows B. Perhaps a better name for the operation might be COMBINE. (Hornstein 2009: 55, note 4)

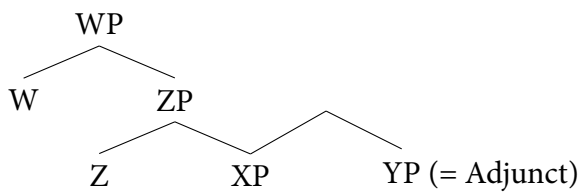
If concatenation is set formation, then reconcatenation inevitably leads to remerge/multidominance, because set-theoretically speaking dominance is just set membership (under transitive closure), and a set (immediately) dominates its members whether it is named/labeled or not. In other words, a label-free system is not also a domination-free system.[16] Thus, as long as a system allows one element to be a member of two or more sets where no set is a subset of any other set (i.e., the given sets all intersect), it de facto uses multidominance. For instance, we can display (5) and (7b) set-theoretically as follows, where the member of the intersection is the SO being multidominated/remerged.[17]

(8) a. $\{\{X, Y\}, Z\} \cap \{W, \{X, Y\}\} = \{X, Y\}$

b. $\{\{\textit{ate}, \textit{the–cake}\}, \textit{in–the–afternoon}\} \cap \{\{\textit{ate}, \textit{the–cake}\}, \textit{in–the–yard}\} \cap \{\{\textit{ate}, \textit{the–cake}\}, \textit{with–a–fork}\} = \{\textit{ate}, \textit{the–cake}\}$

A proposal similar to that of Hornstein & Nunes (2008) is that of Oseki (2015), who also derives adjunction by unlabeled merger. Based on Chomsky (2013), Oseki assumes that when two phrases XP and YP merge with no shared feature, the merger cannot be labeled. Adopting the label accessibility condition (Hornstein 2009: 90, Epstein et al. 2012: 254), which states that unlabeled SOs cannot be accessed by Merge, Oseki then concludes that at this stage the derivation can only proceed by letting one of XP and YP (here XP) participate in further Merge, thus yielding a "two-peaked" structure with one peak containing the adjunct (YP, which "dangles off" in Hornstein & Nunes' terms) as in (9).

(9) (Oseki 2015: 307)

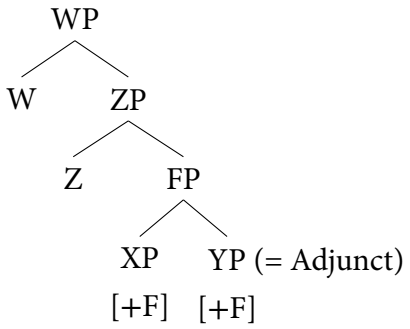


For Oseki the islandhood of "invisible" (i.e., "dangling-off") adjuncts is not derived from unlabeled merger but from compulsory Transfer; namely, the labelless "peak" must be removed from syntax for the derivation to proceed. By comparison, when XP and YP have shared feature(s), their merger can be successfully labeled and no two-peaked structure results, as in (10).

[16] Roy et al. (2016, 2017) similarly propose an unlabeled set-formation operation termed "Juxtapose" for acquisition, where the juxtaposed objects cannot be labeled as the child does not yet know all their formal properties.

[17] Epstein et al. (2012: 261) similarly conceive multidominance structures as "intersecting set-theoretic SOs."

(10)



Here YP, though traditionally called an adjunct, is visible to later derivation. Oseki (2015: 305) illustrates the two types of adjunct with the following examples.

(11) a. *Who did Mary cry [$_{\text{Adjunct}}$ after John hit ~~who~~]? (invisible adjunct)

b. What did John arrive [$_{\text{Adjunct}}$ whistling ~~what~~]? (visible adjunct)

Although Oseki's proposal is inspired by Epstein et al.'s (2012) two-peaked analysis for Spec-TP subjects, the two theories have a number of differences, which I summarize in Table 3.1.

Table 3.1 Comparison of Epstein et al. (2012) and Oseki (2015)

	Epstein et al. (2012)	Oseki (2015)
Cause of two peaks	countercyclic movement	unlabelable merger
Two peaks joined by	internal Merge	external Merge
Labeling of two peaks	both labeled	only one labeled

Table 3.1 reveals three potential problems in Oseki's theory: (i) it crucially relies on the assumption that unlabeled SOs cannot be input to further Merge; (ii) it resorts to multidominance in external Merge; (iii) it transfers an unlabelable SO to the interface. Among the three, (i) may or may not be a real problem depending on how the question of whether every step of Merge must be labeled is answered (Chomsky 2013 answers yes while Bauke & Roeper 2017, Bošković 2016, 2018 inter alia answer no). Next, (ii) is more risky because while (in standard minimalism) internal Merge may sometimes target a nonroot node without violating the no tampering condition (Chomsky 2005)—consider Richards' (2001) "tucking in" for instance[18]—it is arguably more difficult to save external Merge at nonroot positions from no-tampering violation, so multidominance involving external Merge can be truly principle-offending. Finally, (iii) is perhaps the most problematic because according to Chomsky (2013) unlabeled SOs cannot be properly interpreted, so it is unclear how Oseki's theory can meet the interface condition.

There is still another problem in the two-peaked structure, both for Oseki (2015) and for

[18]See Collins (2017: 51) for a strong position against all countercyclic mergers whether external or internal.

Epstein et al. (2012). Since a "peak" comprises two edges, transferring a peak amounts to transferring an entire Merge-created set. In (9) this means that not only the adjunct (YP) but also its host (XP) is invisible to later derivation; in other words, the islandhood of the adjunct affects a larger structure than itself. As Oseki (2015: 308, note 4) points out, this is empirically false; he therefore has to stipulate that XP stays in syntax as a member of {Z, XP}. This stipulation makes Oseki's use of the two-peaked structure at odds with that of Epstein et al. (2012), where the counterpart of XP (i.e., the complement of a phase head) must not stay in syntax as a member of either peak due to the phase impenetrability condition (PIC).[19] What is more, considering adjunction can happen at a wide range of loci, including phase-head-complements, what Oseki must stipulate is actually that XP *optionally* skips Transfer; but why and how?

3.1.2.3 Lexicalist approach: Adjunction via functional head

Contemporary minimalist research has been much influenced by the lexicalist methodology; namely, the practice of encoding syntactic properties and variations in LIs. This is commonly known as the Borer-Chomsky conjecture (BCC):

(12) Borer-Chomsky conjecture (Baker 2008: 156)
All parameters of variation are attributable to differences in the features of particular items (e.g., the functional heads) in the lexicon.

The merit of BCC surpasses Baker's formulation, for not only ideas about parametric variations but also those about core syntactic mechanisms have been gradually "lexicalized"; for example, Merge has been associated with the edge feature, Agree with unvalued features, and Move with the EPP feature. Despite this general trend, however, few studies have attempted to lexicalize adjunction. Thus, the labeling-based perspectives on adjunction are not lexicalist in spirit; they rely on the interaction of derivational procedures instead, much as how syntax was done in the GB era (with various modules interacting at different levels and collectively giving rise to observable effects). In this sense, research on adjunction has not caught up with that in other syntactic domains. But if the guideline of minimalism is (partly) to explain as many phenomena as possible in featural terms, then the lexicalization of adjunction is still a mission.

A notable work in this regard is Rubin (2003), which is the only lexicalist theory of adjunction to my knowledge. Like the derivational proposals, Rubin's proposal is also motivated by a dissatisfaction with the axiomatic approach:

> For Chomsky (2001), adjunction necessarily involves pair-Merge. … [I]t is crucial to ask how … NS determines that pair-Merge is appropriate for adjuncts. We need to avoid circularity here, so we cannot simply say that we want adjuncts to be adjuncts

[19] In fact this is precisely how Epstein et al. derive PIC.

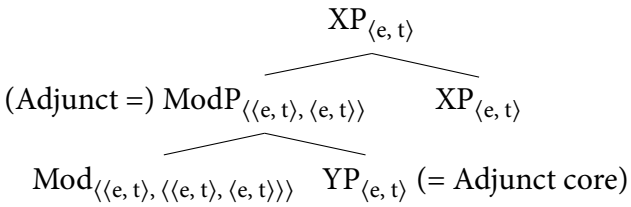
> [and therefore] invoke pair-Merge, which creates adjuncts. Before any two expressions are merged, relational terms such as *adjunct*, *complement*, and *specifier* are premature. Another way to understand the same question is to ask why set-Merge couldn't create adjuncts. (Rubin 2003: 663)

Rubin's solution is to use a special functional head, Mod,[20] which "forms an extended projection around all base adjuncts" such that "[a]ny phrase headed by Mod is subject to pair-Merge" (p. 664). The Mod hypothesis is schematically represented in (13).

(13) [Mod [$_{\text{YP}}$ "Adjunct"]] (Rubin 2003: 661)

According to Rubin, pair-Merge is invoked iff an SO headed by Mod is merged with another SO. Thus, Mod serves as a functional shell for adjuncts just as C and D are functional shells for clauses and nominal phrases, and the syntax "need not search farther than the lexical type of one of the two elements in question, Mod, to determine that the required pair-Merge must apply" (p. 665). More specifically, Rubin defines the semantic type of Mod as $\langle\langle e, t\rangle, \langle\langle e, t\rangle, \langle e, t\rangle\rangle\rangle$ (p. 666); namely, a function from a predicate (i.e., the SO merged with Mod) to a property of predicate (i.e., a predicate modifier). As a result, Mod turns a predicate (the host) into a modified predicate (the adjunction structure) solely by function application, without invoking a separate compositional rule like predicate intersection. Rubin deems this "reduction in … primitive semantic processes" as a major advantage of the Mod hypothesis (p. 667). I display Rubin's composition in (14).

(14)



Another advantage of the Mod hypothesis according to Rubin is that it provides a functional head for morphemes dedicated to modifiers, such as Mandarin *de* (15a) and Tagalog *na* (15b).

(15) a. *nà yì-běn* [$_{\text{ModP}}$ [*zài zhuōzi shàng*] [$_{\text{Mod}}$ *(*de*)]] *shū* [Mandarin]
that one-CL be at table top MODI book
'that book on the table'

b. *bahay* [$_{\text{ModP}}$ [$_{\text{Mod}}$ *(*na*)] [*maganda*]] [Tagalog]
house MODI beautiful
'the beautiful house' (adapted from Rubin 2003: 665)

The use of such "modifier markers" is more complex than meets the eye; for instance, Man-

[20] Unfortunately this overloads the label Mod, which has long been used as an abbreviation of *modality* in generative syntax. Pending a better solution, in this dissertation I limit the "modification" sense of Mod to contexts related to Rubin's proposal and use it in the "modality" sense elsewhere.

darin has several stylistic variants of *de*,[21] and the linearization between adjunct and host is adjunct≺host in Mandarin but host≺adjunct in Tagalog.[22] Such phenomena cannot be easily explained in the axiomatic/derivational approaches, which are too macro-level and not designed to address such micro-level issues as the *de*∼*zhī* alternation. A lexicalist theory like Rubin's is more complete as it can address both macro- and micro-level issues.

That said, Rubin's theory is not optimal. First, as Arsenijevic & Sio (2009: 2) notice, the association of Mod and pair-Merge in Rubin's proposal is inconsistent: two steps of Merge are involved when Mod joins a modifier to a host, but pair-Merge only happens in the second step. Observing the tree in (14), we can find that the way Rubin lets Mod trigger pair-Merge crucially references bar level information; namely, only maximal projections of Mod (but not minimal ones) trigger pair-Merge. Second, the way Mod triggers pair-Merge also crucially references its semantic interpretation, which in minimalism is not available in the syntax. Third, Rubin only defines the type-theoretic function of Mod but does not specify its substantive featural content or conceptual-intentional basis, which makes Mod subject to the same criticism Chomsky (1995: 349) imposes on Agr; namely, it has no C-I interpretation and "is present only for theory-internal reasons." Finally, since Rubin's aim is not to eliminate pair-Merge but merely to let syntax autonomously determine when it applies, his proposal is not fully lexicalist but rather semiaxiomatic; it just adds one more axiom to the Chomskyan system (i.e., Mod^{max} triggers pair-Merge).

3.1.2.4 Interim summary

Above I reviewed three generative approaches to adjunction (in the Chomskyan school). I summarize their major insights and problems in Table 3.2.

[21] A most common one is *zhī*, which is used in elegant or formal registers; e.g.,

(i) *shuǐ* *shì* *shēngmìng* *zhī* *yuán* [Mandarin]
water is life MODI source
'Water is the source of life. (elegant register, slogan)'

Another variant is *dī*, often seen in internet language or pop culture terms (see Tan & Chen 2017); e.g.,

(ii) *ǒu* *dī* *gē-shén* *a*
I.FUNNY MODI song-god SFP
'oh my singing god (funny register, TV show name)'

[22] Compared with other methods of adjunct linearization, a functional-head-based approach has more flexibility. Take Kayne's (1994) LCA-based approach for example. Since the LCA systematically demands left-adjunction, whenever an adjunct appears on the right of its host, this must be due to movement of the host triggered by some third-party functional head. By comparison, for Rubin (2003) adjunct linearization may be accounted for either in terms of third-party-head-triggered movement or directly in terms of some intrinsic property of Mod.

Table 3.2 Three approaches to adjunction in Chomskyan syntax

Approach	Representative	Insight	Problem
Axiomatic	Chomsky (1995, 2000)	inherent asymmetry in adjunction	ad hoc
Derivational	Hornstein (2009), Oseki (2015)	different adjunct types	countercyclic
Lexicalist	Rubin (2003)	functional shell for adjuncts	info-demanding

The axiomatic approach correctly highlights the asymmetry in adjunction, but the way it creates asymmetry is ad hoc. The derivational approach recognizes the fact that adjuncts fall into different types (dangling-on/off or one/two-peaked), but the way it derives the canonical (island-like) adjunction structure is countercyclic (i.e., relies on multidominance). Rubin's (semi)lexicalist proposal groundbreakingly reduces adjunction to a functional category, but the particular category it employs relies on bar-level and semantic information to operate.

The defective category theory retains the key insights of all three approaches and avoids their problems. First, Cat creates a two-membered projection (or two-segment category) solely by set-Merge and the ordinary labeling algorithm, so it brings out the inherent asymmetry of adjunction without invoking extra axioms. Second, the defective category theory is not proposed as a unified theory for all adjuncts but merely provides a new perspective on a subtype of adjunction, so it leaves room for other structural possibilities of adjuncts as the derivational approach does. Since it makes no use of unlabeled merger or halted derivation, it also avoids the multidominance pitfall. Third, Cat is a functional shell for adjuncts just like Mod, and it is a conceptually-intentionally motivated one, because its mergeme feature, once valued, is interpretable. Meanwhile, since Cat does not function by triggering other operations but directly and deterministically establishes an adjunction structure via Agree, it requires no more information than that standardly available in the syntax (i.e., unvalued/valued features). What is more, since the Cat hypothesis dispenses with adjunction-specific combinatorial rules like pair-Merge altogether, it represents the lexicalist approach even better than the Mod hypothesis does.

3.2 Modifier-head compounds

Compounding is "an important means of extending the lexicon of a language" (Lieber & Štekauer 2009: 3), and modifier-head compounds (MHCs) like *blackboard* and *water bottle* are a most common type of compound. In this section I first delineate how MHCs are syntactically derived (§3.2.1) and semantically interpreted (§3.2.2) under the defective category theory and then move on to discuss implications of the theory (§3.2.3). I list a few remaining questions in §3.2.4.

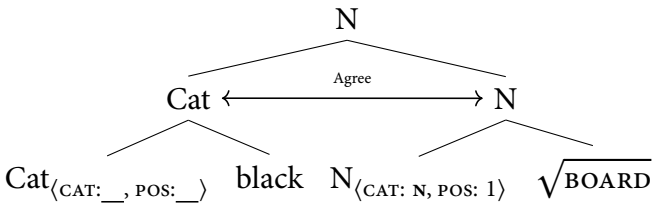
3.2.1 Derivation

The derivation of MHCs straightforwardly follows from the schema in Figure 3.1. All we need to do is substitute X, ω_1, and ω_2 respectively by the compound's overall category, the modifier material, and the head material. Imagine modifier-head compounding as a function f, as in Figure 3.2, where f takes a triple input and yields an MHC output. The role of the defective category theory is to give f a specific definition (via the constant Cat and the schema in Figure 3.1).

$$\xrightarrow{(\mathrm{X}, \omega_1, \omega_2)} \boxed{f} \xrightarrow{\mathrm{MHC}}$$

Figure 3.2 Modifier-head compounding as a function

Let us begin with the most common scenario: that of compounds like *blackboard*. In this scenario ω_1 is a preexisting word (i.e., a separately lexicalized object) and ω_2 a root, which becomes the lexical head of the compound upon merging with X. I illustrate the derivational process in Figure 3.3 in two formats.



(a) Syntactic tree

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\mathrm{Cat}_{\langle \mathrm{CAT}:__,\ \mathrm{POS}:__\rangle}}\ \text{axiom} \qquad \mathrm{black}}{\{\mathrm{Cat}_{\langle \mathrm{CAT}:__,\ \mathrm{POS}:__\rangle}, \mathrm{black}\}}\ \text{merge}}{{}_{\mathrm{Cat}}\{\mathrm{Cat}_{\langle \mathrm{CAT}:__,\ \mathrm{POS}:__\rangle}, \mathrm{black}\}}\ \text{label} \qquad \dfrac{\dfrac{\dfrac{}{\mathrm{N}_{\langle \mathrm{CAT}:\ \mathrm{N},\ \mathrm{POS}:\ 1\rangle}}\ \text{axiom} \qquad \dfrac{}{\sqrt{\mathrm{BOARD}}}\ \text{axiom}}{\{\mathrm{N}_{\langle \mathrm{CAT}:\ \mathrm{N},\ \mathrm{POS}:\ 1\rangle}, \sqrt{\mathrm{BOARD}}\}}\ \text{merge}}{{}_{\mathrm{N}}\{\mathrm{N}_{\langle \mathrm{CAT}:\ \mathrm{N},\ \mathrm{POS}:\ 1\rangle}, \sqrt{\mathrm{BOARD}}\}}\ \text{label}}{\{{}_{\mathrm{Cat}}\{\mathrm{Cat}_{\langle \mathrm{CAT}:__,\ \mathrm{POS}:__\rangle}, \mathrm{black}\}, {}_{\mathrm{N}}\{\mathrm{N}_{\langle \mathrm{CAT}:\ \mathrm{N},\ \mathrm{POS}:\ 1\rangle}, \sqrt{\mathrm{BOARD}}\}\}}\ \text{merge}}{\dfrac{\{{}_{\mathrm{Cat}}\{\mathrm{Cat}_{\langle \mathrm{CAT}:\ \underline{\mathrm{N}},\ \mathrm{POS}:\ \underline{1}\rangle}, \mathrm{black}\}, {}_{\mathrm{N}}\{\mathrm{N}_{\langle \mathrm{CAT}:\ \mathrm{N},\ \mathrm{POS}:\ 1\rangle}, \sqrt{\mathrm{BOARD}}\}\}}{{}_{\mathrm{N}}\{{}_{\mathrm{Cat}}\{\mathrm{Cat}_{\langle \mathrm{CAT}:\ \underline{\mathrm{N}},\ \mathrm{POS}:\ \underline{1}\rangle}, \mathrm{black}\}, {}_{\mathrm{N}}\{\mathrm{N}_{\langle \mathrm{CAT}:\ \mathrm{N},\ \mathrm{POS}:\ 1\rangle}, \sqrt{\mathrm{BOARD}}\}\}}\ \text{label}}\ \text{agree}$$

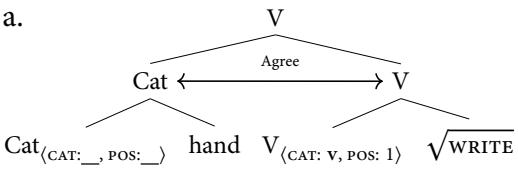
(b) Proof tree

Figure 3.3 Derivational path of *blackboard*

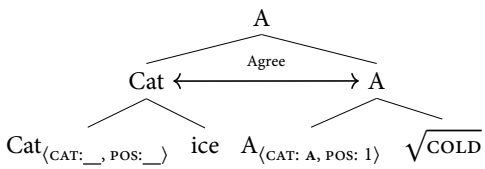
In Figure 3.3 since N labels both {N, $\sqrt{\text{BOARD}}$} and {{Cat, black}, {N, $\sqrt{\text{BOARD}}$}}, it participates in two cycles of categorization, one for $\sqrt{\text{BOARD}}$ and the other for the entire compound. If each categorization cycle provides a locus of lexicalization (or idiomatic interpretation; see Marantz 1995, Harley 2014), then N also links two products of lexicalization, one for the head (a

type of entity) and the other for the compound (another type of entity). The categorization/lexicalization of the modifier material *black* is a separate issue. As Figure 3.3b shows, an MHC is derived from three axioms (Cat, X, ω_2) and one lemma (ω_1). The derivational history of the lemma does not matter at the compounding level; we may as well view ω_1 as a flattened sound-meaning pair like a root, say ⟨/blæk/, 'of the color black (a type of quality)'⟩.[23] By Figure 3.3 we can derive all *blackboard*-type MHCs. Below are two more examples, one verbal (*hand-write*) and the other adjectival (*ice-cold*). I display them only in the syntactic tree format to save space.

(16) a.



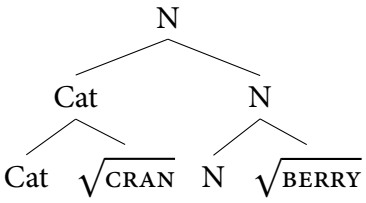
b.



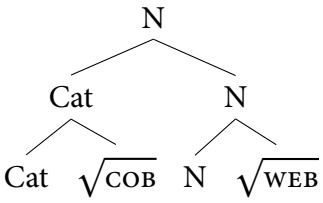
As with N above, in (16) V/A simultaneously categorizes the compound and its lexical head and thereby links two lexicalization products ('write'∼'write by hand' and 'cold'∼'cold like ice').

Since the derivational history of ω_1 is irrelevant, ω_1 may be of any category. The above-mentioned *black*, *hand*, and *ice* are adjectival/nominal, but ω_1 may also be verbal, as in *bite plate* and *stir-fry*.[24] It may even be a bound morpheme without clear meaning (at least synchronically), such as *cran-* in *cranberry* and *cob-* in *cobweb*. In this scenario we can view ω_1 as a root, which only gets categorized and lexicalized as part of the ambient compound, as in (17).

(17) a.

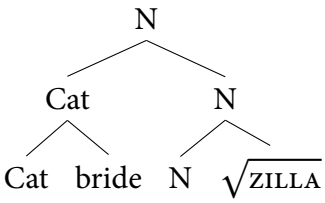


b.

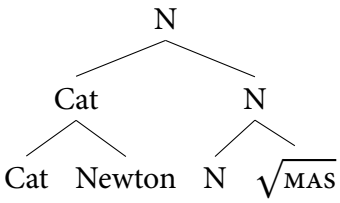


Similarly, ω_2 may also be a bound root that only gets categorized as part of the compound, such as *-zilla* in *bridezilla*/*momzilla*/*weddingzilla* and *-mas* in *Christmas*/*Michaelmas*/*Newtonmas*. I illustrate their derivations in (18).

(18) a.



b.



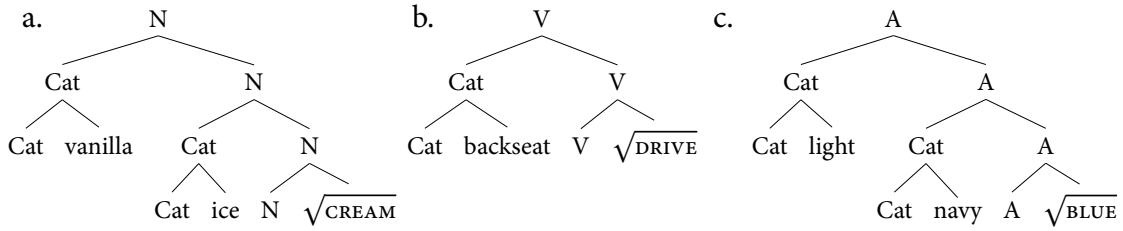
In addition, ω_1 and ω_2 may both be filled by more complex SOs, such as *ice cream* in *vanilla*

[23] Of course, such a flattened object differs from a real root in having specific phonological and semantic interpretations, for it has gone through categorization in some previous cycle. As such, we could call it a *root-like chunk*.

[24] Apparently ω_1 may be prepositional too; e.g., *uptake*, *download*. But I will generally abstract away from prepositional modifiers, since they involve an extra complication beyond my scope; see §3.2.4 for a comment.

ice cream, *backseat* in *backseat-drive*, and *navy blue* in *light navy blue*. I illustrate these in (19).

(19) a. b. c.



Note that in (19a) and (19c), since the category of the big compound's lexical head is inherited from a smaller compound, X is not really a sister node of ω_2 but essentially its label.

3.2.2 Composition

It is generally agreed that the meanings of compounds are not neatly compositional. In Bauer's (2017: 106) words, "the semantics of all kinds of compounds seems to be variable." Fabb (1998: 66) similarly states that "the meaning of a compound is usually to some extent compositional, though it is often not predictable." Yet as the first half of Fabb's statement indicates, there is still some compositionality in compound semantics. In this section I will show that all MHCs have a fully compositional facet in their semantics, the categorial facet, which works in tandem with the defective category schema.[25] I will still use *blackboard* for example. To figure out how the syntactic representation in Figure 3.3a is compositionally interpreted, we need to determine what each terminal node denotes and what compositional rule each merger invokes. What we readily know are: (i) Cat has no complete denotation of its own but will obtain one from N, (ii) *black* has a usual adjectival denotation (i.e., a predicate over a quality variable), and (iii) the merged SO {N, $\sqrt{\text{BOARD}}$} has a usual nominal denotation (i.e., a predicate over an entity variable). Formally:[26]

(20) a. $[\![\text{Cat}]\!] = \ ?$

b. $[\![\text{black}]\!] = \lambda x{:}\texttt{quality}\,.\,\text{Black}(x)$

c. $[\![\{\text{N}, \sqrt{\text{BOARD}}\}]\!] = \lambda x{:}\texttt{entity}\,.\,\text{Board}(x)$

What we do not know yet are: (i) the denotation of N, (ii) the denotation of $\sqrt{\text{BOARD}}$, and (iii) the mode of the N-$\sqrt{\text{BOARD}}$ composition. In other words, the crux of the matter is how to compose categorizers and roots. Since there is little previous research on root-level compositional semantics,[27] in what follows I will explore a few alternative possibilities before drawing a conclusion.

[25] The claim may be extended to all compounds, but I focus on MHCs here and leave the extension to future research.

[26] I put types in the typewriter font as in Chapter 2 and adopt the *var*:`type` format to indicate variable typing, which is standard practice in typed lambda calculus (Barendregt 1992). Besides, I capitalize the initial letters of predicate names to distinguish them from object language words.

[27] The only dedicated study to my knowledge is Kelly (2013), which I will review below. There have been semantically informed decompositional studies of lexical categories, but those are either non-root-based (e.g., Ramchand 2008, 2018) or more focused on the (morpho)syntactic side of the formalization (e.g., Borer 2005a,b, 2013).

3.2.2.1 Possibility I: Roots denote type-generic predicates

Mainstream formal semantics does not decompose bare LIs but lets them denote single semantic objects; for example, the bare noun *board* as a whole denotes $\lambda x : \texttt{entity} \,.\, \text{Board}(x)$ (a first-order predicate). This convention is retained in more recent studies that otherwise do use lexical decomposition. In the neo-Davidsonian representation in (21), for example, the bare verb *butter* denotes BUTTER(e), which integrates an idiosyncratic tag "BUTTER" and an event variable "e."

(21) a. Jones buttered the toast slowly in the bathroom with a knife.

b. $\exists e[\text{BUTTER}(e) \wedge \text{AGENT}(e) = j \wedge \text{THEME}(e) = t \wedge \text{SLOWLY}(e) \wedge \text{LOCATION}(e) = b \wedge \text{INSTRUMENT}(e) = k]$ (Landman 2000: 1–2)

The same is true in Ramchand's first-phase syntax (22a) and quotational semantics (22b). I have abbreviated Ramchand's denotations and only kept the bare verb parts.

(22) a. $[\![\text{push}]\!] = \ldots \lambda e[\ldots \textit{push}(e) \ldots]$ (Ramchand 2008: 61)

b. $[\![\text{run}]\!] = \ldots \lambda e[\text{run}(e)] \ldots$ (Ramchand 2018: 15)

Here too the idiosyncratic and compositional parts of a bare verb are integrated into a single expression. The examples in (21)–(22) show that the root/categorizer-oriented thinking—or the idea of separating compositional and idiosyncratic content in meaning representation—has not had much influence in formal semantics. Despite this state of affairs, a root-level semantics is not too difficult to implement. Kelly (2013) adopts a conjunctivist approach (as in event semantics) and lets the root and the categorizer both denote first-order predicates, though he assigns the former a more general argument type than the latter. Thus, when the two compose—by argument identification—the argument of the overall predicate is restricted to the categorizer's type. Kelly takes eventuality to be a generic type covering all of nouns, verbs, and adjectives. He then divides eventuality into two subtypes, event (for verbs) and state (for nouns and adjectives),[28] and assigns these to the DM categorizers. Meanwhile, the generic type eventuality is assigned to roots under the assumption that roots are "all underlyingly properties of eventualities, but are vague between properties of events and properties of states" (Kelly 2013: 81). Kelly illustrates the root-categorizer composition with the following examples:

(23)

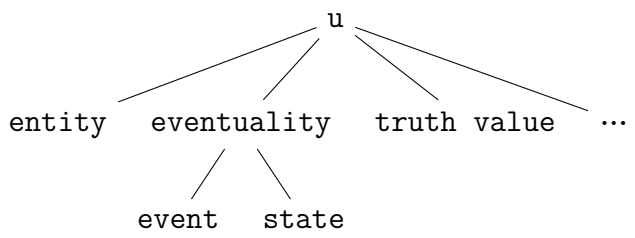
Root	**Categorizer**	**Composition**
$[\![\sqrt{\text{RED}}]\!] = \lambda e \,.\, red'(e)$	$[\![a^0]\!] = \lambda e \,.\, \text{state}(e)$	$\lambda e \,.\, \text{state}(e) \wedge red'(e)$
$[\![\sqrt{\text{DOOR}}]\!] = \lambda e \,.\, door'(e)$	$[\![n^0]\!] = \lambda e \,.\, \text{state}(e)$	$\lambda e \,.\, \text{state}(e) \wedge door'(e)$
$[\![\sqrt{\text{BREAK}}]\!] = \lambda e \,.\, break'(e)$	$[\![v^0]\!] = \lambda e \,.\, \text{event}(e)$	$\lambda e \,.\, \text{event}(e) \wedge break'(e)$

(Kelly 2013: 82, 95)

[28] Kelly (2013: 78) assumes that *n* and *a* are "semantically equivalent" and that their difference is "purely syntactic," with *n* requiring "a referential index provided by a determiner" while *a* not so.

Throughout (23) *e* is implicitly typed as eventuality, and the more specific types (i.e., those in the categorizers) are expressed by first-order predicates. As such, Kelly treats (nongeneric) types and roots as the same sort of semantic object, though he presumably gives them different theoretical status, as reflected in his use of two different typefaces (roman and italic). To facilitate comparison with the alternative possibilities to be introduced, I adapt Kelly's idea into a more familiar type system (i.e., that in Kratzer & Heim 1998, with a Davidsonian extension), where nouns denote predicates over entities rather than eventualities. I also propose a truly generic type u, defined as the most general supertype in the heterogeneous universe of discourse introduced in §2.2.3.1. All the conventional types are thus subtypes of u.[29] Below I illustrate this ontology with a HPSG-style type hierarchy (I have kept Kelly's subtyping of `eventuality`).

(24)



The root-categorizer composition proceeds as in Kelly's model. I simply replace `eventuality` by u as the type of roots and `state` by `entity` as the type of the nominalizer. Accordingly, I update Kelly's examples in (23) to (25). I change Kelly's eventuality variable *e* to the type-neutral *x* and mark its background type u explicitly. I also adapt the notation to my own style and mark types in the *var*:`type` format in the composition column to highlight the different theoretical status between type declarations and root predicates.

(25)

Root	**Categorizer**	**Composition**
$[\![\sqrt{\textsc{red}}]\!] = \lambda x{:}\mathtt{u}\,.\,\mathrm{Red}(x)$	$[\![\mathrm{A}]\!] = \lambda x{:}\mathtt{u}\,.\,\mathrm{State}(x)$	$\lambda x{:}\mathtt{state}\,.\,\mathrm{Red}(x)$
$[\![\sqrt{\textsc{door}}]\!] = \lambda x{:}\mathtt{u}\,.\,\mathrm{Door}(x)$	$[\![\mathrm{N}]\!] = \lambda x{:}\mathtt{u}\,.\,\mathrm{Entity}(x)$	$\lambda x{:}\mathtt{entity}\,.\,\mathrm{Door}(x)$
$[\![\sqrt{\textsc{break}}]\!] = \lambda x{:}\mathtt{u}\,.\,\mathrm{Break}(x)$	$[\![\mathrm{V}]\!] = \lambda x{:}\mathtt{u}\,.\,\mathrm{Event}(x)$	$\lambda x{:}\mathtt{event}\,.\,\mathrm{Break}(x)$

How does the Kellyan approach fare in the composition of MHCs? It correctly composes X and ω_2; for instance, $[\![\{\mathrm{N}, \sqrt{\textsc{board}}\}]\!] = [\![\mathrm{N}]\!] \wedge [\![\sqrt{\textsc{board}}]\!] = \lambda x{:}\mathtt{u}\,.\,\mathrm{Entity}(x) \wedge \lambda x{:}\mathtt{u}\,.\,\mathrm{Board}(x) = \lambda x{:}\mathtt{u}\,.\,\mathrm{Entity}(x) \wedge \mathrm{Board}(x) = \lambda x{:}\mathtt{entity}\,.\,\mathrm{Board}(x)$. Besides, upon the Cat-X agreement, the Cat-ω_1 merger also gets a computed denotation, as in (26).

(26) $[\![\{\mathrm{Cat}_{\langle \textsc{cat}:\underline{\mathrm{N}},\, \textsc{pos}:\underline{1}\rangle}, \mathrm{black}\}]\!] = \lambda x{:}\mathtt{entity}\,.\,\mathrm{Black}(x)$

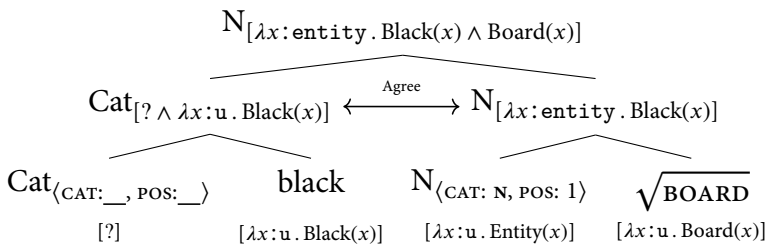
Note that a nominally valued Cat is not the same as a nominalizer; that is, the (re)categorization in (26) does not yield the independently lexicalized nominal meaning for *black* (i.e., a color name).

[29] The conventional basic types are all sort-types; hence, the logical metalanguage they are part of is essentially *many-sorted* (see Wang 1952, Meinke & Tucker 1993 inter alia for discussions of many-sorted logic).

This is because (i) *black* is not a root (so the categorization is not that needed for the lexical noun *black*, which is presumably {N, $\sqrt{\textsc{black}}$}), and (ii) in the course of derivation Cat and N are categorially unified (with N becoming a two-segment category) and come to share a single label (N); hence, {Cat, black} is not interpreted alone at the C-I interface but in conjunction with {N, $\sqrt{\textsc{board}}$}. In other words, the valued Cat in (26) does not categorize *black* into a noun but categorizes it into part of a bigger noun. The ultimate compositional representation of *blackboard* is given in (27a), and the same information is displayed in tree format in (27b).

(27) a. $[\![\{\{\text{Cat}, \text{black}\}, \{\text{N}, \sqrt{\textsc{board}}\}\}]\!] = \lambda x\colon \texttt{entity}\,.\,\text{Black}(x) \wedge \text{Board}(x)$

b.



However, there are two problems in the above composition. First, it does not really give us the meaning of *blackboard* but merely gives us that of Cat-N, which denotes an entity-typed semantic category. As to the definition of that category, all (27) tells us is that it is somehow related with "Black" and "Board." Thus, we are brought back to our point of departure; namely, the meaning of an MHC is only compositional in its categorial facet but not in its lexical-encyclopedic facet. The latter's noncompositional nature leads us to ask what benefit having root denotations in the semantic model really gives us on the one hand and makes it inappropriate to represent the integration of root interpretations by logical conjunction on the other.

Second, the above composition is tailored for lexical categories; that is, it only works for DM-style categorizers and lexical-purpose roots, which we know a priori would compose into first-order predicates (i.e., the denotations of bare lexical categories). Thus, Kellyan root semantics is bundled with the presupposition that roots can only ever be used for lexical categories. While this is an established assumption in most root-based theories, it makes the peculiar prediction that the lexical/functional division is black-and-white; yet half-lexical–half-functional items are widely attested in world languages (e.g., the numerous classifiers in East Asian languages; such items will be investigated in Chapter 4), and a proper theory for such items requires a principled mechanism to integrate functional categories and idiosyncratic-information-bearing units (i.e., roots in our terminology). In short, the lexical-category-specific nature of possibility I makes its applicability and explanation power unduly limited.

3.2.2.2 Possibility II: Roots denote type-open predicates

A major intuition possibility I implements is the type uncertainty of roots; it does so by assigning the root a generic type. But we could implement the same intuition by simply leaving the root's type slot open; namely, we can choose not to assign the root predicate any concrete type at all but completely let it be determined by the categorizer. Intuitively, the composition goes as follows:

(28) $\text{categorizer}_{[\texttt{tp}]} + \text{root}_{[\lambda x:_\,.\,\mathrm{P}(x)]} = \text{categorized root}_{[\lambda x:\underline{\texttt{tp}}\,.\,\mathrm{P}(x)]}$
(tp is a type, P a predicate name)

The above composition cannot be expressed in the usual (simply typed) lambda calculus, where lambda abstraction is exclusively over terms; to express (28) we need to abstract over types. In Churchian type theory (Church 1940 et seq.), type-level abstraction and application are handled by second-order typed lambda calculus ($\lambda 2$; see Nederpelt & Geuvers 2014 for an introduction). We have already seen $\lambda 2$ in note 24, p. 13 in the context of coefficient feature values. Both features and roots are high-granularity constructs outside the purview of traditional formal semantics, and the fact that they both require a more advanced logical metalanguage suggests that as we push syntactic analysis to a higher granularity, we need to update our semantic toolkit as well. The intuitive composition in (28) is expressed in $\lambda 2$ in (29).

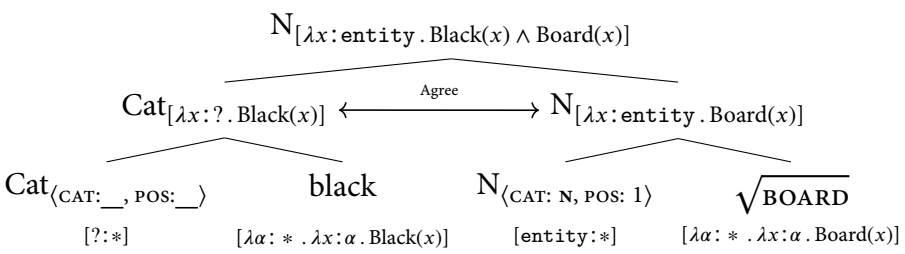
(29) a. $[\![\text{root}]\!] = \lambda\alpha : * \,.\, \lambda x : \alpha \,.\, \mathrm{P}(x)$

b. $[\![\text{categorizer}]\!] = \texttt{tp} : *$

c. $[\![\{\text{categorizer}, \text{root}\}]\!] = [\![\text{root}]\!]([\![\text{categorizer}]\!]) = [\lambda\alpha : * \,.\, \lambda x : \alpha \,.\, \mathrm{P}(x)](\texttt{tp} : *)$
$= \lambda x : \texttt{tp} \,.\, \mathrm{P}(x)$

In (29), α is a type variable, x a term variable of type α, and $*$ the type of all types (called a *kind*). The categorizer and the root compose by type application, with the root providing the function and the categorizer providing the argument. As a result, the type information in the categorizer is integrated into the root predicate. Below I illustrate how this works with the example *blackboard*.

(30)



The composition in (30) is very similar to that in (27b), except for two points. First, the root now has an indispensable place in the semantic model, because it is where the term variable is introduced; the categorizer only specifies the type for that variable but does not "identify" with it.

Second, the denotation of Cat is no longer a bare question mark but has a kind assignment $*$.[30] As such, under possibility II all syntactic categories are associated with some type information in the lexicon: roots with a type variable, nondefective categories with type constants, and Cat with a "kinded" type slot. By contrast, under possibility I Cat is lexically void of any type information.

Nevertheless, possibility II is not rid of possibility I's problems. First, we still lack a deductive connection between the idiosyncratic content in *black*/$\sqrt{\text{BOARD}}$ and that in *blackboard*, so the indispensability of root denotations here is driven by theory-internal reasons rather than virtual conceptual necessity. Second, since possibility II composes categorizers and roots into first-order predicates just as possibility I does, it also suffers from the nongeneralizability problem and leaves little room for semifunctional items.

3.2.2.3 Possibility III: Roots lack model-theoretic denotations

Faced with the common problems in the foregoing possibilities of root-level semantics, in this section I consider a third possibility: perhaps roots do not participate in model-theoretic composition at all. This brings us back to a basic idea in DM, as Marantz states below:

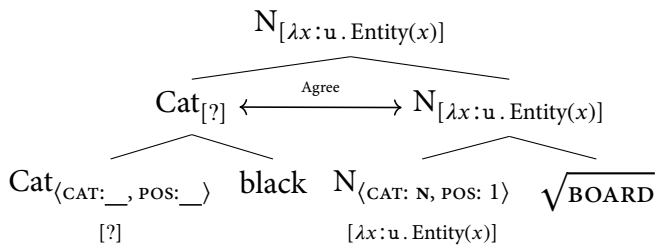
> [I]t is not clear that the computational system of language … must know whether a node contains "dog" or "cat." Distributed Morphology takes the position that this difference between "dog" and "cat" is a matter of Encyclopedic knowledge and that such knowledge is used in semantic interpretation of LF, but not in grammatical computations over LF or involving LF. (Marantz 1995: 4)

The key message here is the separation of compositional/logical and noncompositional/idiosyncratic meanings at the syntax-semantics interface: roots by themselves are invisible not only to feature-driven syntactic operations like Agree and Move but also to syntactically based semantic composition. As a result, the semantic effect of a root is only manifested when compositional and noncompositional meanings are integrated, which is arguably not a model-theoretic process. Two consequences follow from this conception: (i) the categorizer must denote a full-fledged function with a typed variable;[31] (ii) the denotation of a categorizer-root merger is identical to that of its categorizer part. Thus, the composition of *blackboard* proceeds as in (31).

[30] This information is nontrivial, for $*$ is not the only kind. Other kinds include $* \rightarrow *$, $* \rightarrow * \rightarrow *$, etc., which are inhabited by various function types. The question mark here can be viewed as an arbitrary type constant similar to the arbitrary term constants in predicate logic (see Partee et al. 1993: 153).

[31] Here I am assuming the categorizer introduces both the variable and its type, but it could also just introduce a type, which is then fed into a variable-introducing function (as in Acquaviva's 2019 model to be introduced below). Such technical variation is allowed under possibility III as long as root content is excluded from the computation.

(31)



Here the compositional meaning of *blackboard* is merely that it denotes a particular entity category. And it is the entire SO {{Cat, black}, {N, $\sqrt{\text{BOARD}}$}} that gets mapped to the lexicalized meaning 'blackboard'. That is, the defective category schema merely provides a categorial skeleton for the lexicalization of the compound but does not encode the lexicalized result. From this perspective, the conventional practice of incorporating root (or any idiosyncratic) content in model-theoretic denotations (e.g., λx . Board(x), $\exists e$[Run(e)]) is at best expository and at worst completely wrong. This is reminiscent of Chomsky's objection to the referentialist approach to natural language semantics:

> [A] lexical item provides us with a certain range of perspectives for viewing what we take to be the things in the world, or what we conceive in other ways; these items are like filters or lenses, providing ways of looking at things and thinking about the products of our minds. The terms themselves do not refer …. (Chomsky 2000: 36)

Chomsky's LIs have phonological, semantic, and syntactic features all packed in, but in the above context he is mainly concerned with the semantic—or more exactly encyclopedic—dimension; namely, the dimension that brings out, for example, the "*house-home* difference" (ibid.). This type of information is precisely what I have attributed to the root. Hence, Chomsky's view that LIs do not refer and my conclusion that roots do not denote convey basically the same idea. Acquaviva expresses a similar attitude regarding DM roots:

> [A]n extensionalist semantic approach, where basic terms of the semantic representation are ultimately defined by what they are true of, in one or more than one world … cannot possibly shed much light on those aspects of lexical semantic competence based on oppositions in conceptualization rather than in distinct extensions: consider again *home* vs. *house*, or *broad* vs. *wide*, or *use* vs. *utilize*, to say nothing about notorious problematic cases like *time*, *air*, or *god*. (Acquaviva 2014a: 281)

In a later study of nominal structure, Acquaviva (2019) explicitly denies the root any type and only introduces the first-order predicate conventionally associated with a (bare common) noun via a functional head PΣ (property of sum) merged above the nominalizer, as in (32).

(32)

Syntactic object	Semantic type	Description
root	-	purely differential content
n-root	e	introduces an entity type
PΣ-*n*-root	⟨e, t⟩	introduces variable, creates lattice

(based on Acquaviva 2019: 51)

In this model the nominalizer introduces an entity type which is then named by a root, and the nominalized root becomes the smallest nominal structure conceived as "an unanalyzable name, a label maximally underdetermined except for the fact of being formally distinct from other names" (Acquaviva 2019: 45).[32] In other words, the root is just a modifier or name tag for the categorizer/type, though importantly it is not a modifier in the model-theoretic sense (i.e., not a predicate). As Acquaviva (2009: 4) points out, the meaning of a root is "too elusive to be pinned down," for "something so radically underspecified cannot even convey the distinction between argument and predicate." Rather, the semantic effect of this root-to-type modification probably only takes place in the final integration of compositional and noncompositional meanings, where various noncompositional ingredients modify the compositional structures (i.e., SOs) they are part of. This sounds like Marantz's (1995) view just above (31), and Marantz in a later article indeed makes a remark similar to that of Acquaviva:

> [T]he idiosyncrasies in use of verbal roots must be separated from the general, non-idiosyncratic connections between structure and meaning .... The little v semantically introduces an eventuality, ... [and] roots may modify the event introduced by the little v. (Marantz 2013: 154, 157)

So Marantz also lets the categorizer declare a type and lets the root modify that type. Interestingly, both Acquaviva's proposal for the nominal domain and Marantz's proposal for the verbal domain resemble an eclectic combination of our possibility II (where categorizers denote types) and possibility III (where roots do not denote).

3.2.2.4 Interim summary

In §3.2.2 I have examined the semantic composition of MHCs based on the syntactic derivation in §3.2.1. As it turns out, a major obstacle in implementing such a composition is the semantic nature of the root-categorizer merger. As Acquaviva (2014b: 53) points out: "[T]he dissociation between root and category as distinct syntactic heads ... raises the question of how to properly describe the content of the former in isolation from the latter; that is, how to state what *cat* means as a root abstracting away from its being a noun." I have explored three possibilities building on a number of previous studies, as summarized in Table 3.3.

[32] This formal distinction, if relevant to syntax at all, need not be based on semantic content but may also be brought out by other means such as the numerical indices (e.g., $\sqrt{}_{279}$) promoted in Harley (2014).

Table 3.3 Three possibilities of root-categorizer composition

Possibility	Root denotation	Categorizer denotation	Composition
I	Type-generic predicate	typing predicate	conjoined predicate
II	Type-open predicate	type	typed predicate
III	None	typing predicate or type	typing predicate or named type

Among the three possibilities, III is the overall most optimal insofar as the root part is concerned, whereby the compositional meaning of an MHC is all and only that of its categorial skeleton (i.e., the Cat-X two-segment category), and the lexicalized meaning of the MHC is only incorporated at the C-I interface when compositional and noncompositional semantics are integrated. Meanwhile, under possibility III, both the categorizer denotation in possibility I and that in possibility II seem compatible with our proposed derivational path for MHC. I opted for the former in (31) for expository convenience, but the latter may turn out to be more decompositional in spirit (as in Acquaviva's 2019 model).

3.2.3 Implications

After delineating the derivation and composition of MHCs based on the defective category theory, next I discuss some implications of the theory. For space reasons I will not be able to elaborate on any of the empirical domains, but I hope to present a clear picture as to what kinds of phenomena we might use the defective category theory to explain.

3.2.3.1 Modifier-head compounding is universal

This implication consists of two parts: (i) adjunction is universal; (ii) modifier-head compounding is universal. I begin with adjunction. Since the defective category is defined by mergeme underspecification, it is a free rider of mergeme-based categories. Under the assumption that all human languages have mergeme-based categories (i.e., no language is based solely on acategorial categories), this means that all languages have a readily usable defective category; and if Cat underlies certain adjunction structures, then adjunction as a grammatical phenomenon should be universal. This prediction is rather unexciting, but the novel point here is that the universality of adjunction is a logical consequence of the universality of Cat, which in turn is a consequence of the universality of mergeme-based categories. A corollary of this train of thought is that adjunction should be among the first to emerge in both language acquisition and language evolution, since Cat is available once any other mergeme-based category is available.

That said, what specific Cat-based phenomena are available in a language crucially depends on what Xs are usable. In particular, since the schema creates a modificational structure, for

a Cat-based phenomenon to emerge in category X it is not enough to only have a working X, but X must also be associated with some roots or root-containing chunks. In other words, X should be freely modifiable, or it would be pointless to have a modificational structure for it. In sum, a potentially universal Cat-based phenomenon should have an X that is (i) itself universal, and (ii) universally associated with roots. The most obvious candidates for X are the traditional lexical categories. Their universality is generally accepted (Baker 2003, Panagiotidis 2015) on the one hand,[33] and they are the quintessential categories associated with roots on the other. Thus, we expect the traditional lexical categories to produce some universal Cat-based phenomenon, which is precisely modifier-head compounding.

The universality of modifier-head compounding is supported by typological studies. According to Bauer (2009: 344), modificational compounding has been suggested to be a language universal (Fromkin et al. 1996: 54–55, Libben 2006: 2), which is further supported by evidence from language acquisition (Clark 1993) and language contact (Plag 2006). Sometimes this universality is masked by varied classification or terminology. For example, descriptive grammars of Ainu (e.g., Refsing 1986, Shibatani 1990) do not mention compounding at all, though the language evidently does have MHCs, as in (33a). Evenki has also been claimed to lack compounds (Nedjalkov 1997: 308), but a quick look into alternative sources reveals quite a few, as in (33b).[34]

(33) a. Ainu (language isolate; via Bauer 2009)

atuy asam	'bottom sea; sea bottom'
kamuy napuri	'mountain god; holy mountain'
supuya kur	'trace smoke; smoke trace'

b. Evenki (Tungusic; Hu & Chao 1986)

əjʤi ʃɛɛ	'brick tea'
aaxɪn ʤɔlə	'liver stone; marble'
ʊnaaʤɪ utə	'girl son; daughter'

3.2.3.2 Modifier-head compounding has no intrinsic categorial bias

Perhaps at odds with a common impression, the defective category theory predicts no categorial bias in modifier-head compounding: not only is the derivational schema in Figure 3.1 compatible with any lexical category, but there is no way to encode any categorial restriction in it at all. This means that any observed categorial bias in modifier-head compounding must be due to independent factors instead of compounding-specific ones. I illustrate this point with an informal survey of English verbal MHCs, as exemplified in (34).[35]

[33] Here all that matters is that such big X categories universally *exist*; whether they also have universal definitions is orthogonal (they probably do not; see Biberauer 2017a).

[34] Throughout this dissertation I use semicolons to separate verbatim glosses and free translations when they differ.

[35] For expository convenience I hyphenate all verbal MHCs regardless of their usual orthography.

(34) hand-wash, stir-fry, baby-sit, double-check, sleep-walk, day-dream …

Bauer (2017: 99) notices a "common comment" in the literature that compound verbs "are somehow rare or restricted" in English. For instance, Marchand (1969: 100) asserts that verb compounding "does not exist in Present-day English"; Adams (2001: 100) similarly states that "[v]erb compounding is a productive process in some languages, but English is not one of them." A language that does have productive verb compounding is Chinese, as exemplified in (35).

(35)	*shǒu-xǐ*	'hand-wash'	*màn-pǎo*	'slow-run; jog'	[Mandarin]
	shǒu-xiě	'hand-write'	*duǎn-pǎo*	'short-run; sprint'	
	fān-chǎo	'stir-fry'	*dà-xiào*	'big-laugh; guffaw'	
	gān-chǎo	'dry-fry'	*dà-mà*	'big-scold; tongue-lash'	

The Chinese compounds in (35) do not feel that heterogeneous from the English ones in (34); some even have verbatim counterparts. This leads us to query whether the alleged unproductiveness of verb compounding in English really reflects any grammatical restriction. In fact even "unproductiveness" itself is questionable. Bauer has repeatedly pointed this out since the 1980s:

> There are plenty of this type of verb [N-V compound; e.g., *block-bust*, *carbon-date*, *sky-dive*] being coined in current English. (Bauer 1983: 208)

> [C]ompound-verb formation is alive in English. Note that our data include examples with nouns, verb, and adjectives in the first element [e.g., *custom-produce*, *freeze-dry*, *slow-bake*] and with variable meaning relationships between the elements, just as we would expect to find with compounds. (Bauer & Renouf 2001: 110)

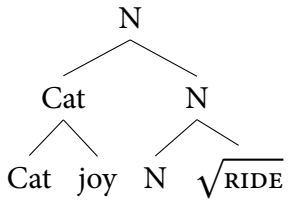
> [T]he presence of such verbs [e.g., *tunnel-chase*, *badge-flash*, *air-quote*] is a feature of individual writers, but it is clear that such constructions can easily be found. … [V]erbal compounding is no longer marginal in current English. (Bauer 2017: 138)

Bauer (2017: 171) suggests that compound nouns may appear more usual than compound verbs simply because "nouns are generally more numerous in languages than verbs."

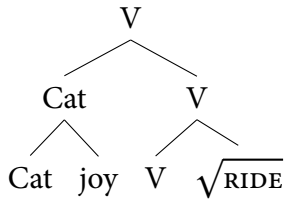
An often-cited fact about verbal MHCs is that they are mainly coined by back-formation, which is why Marchand calls them "pseudo-compound verbs"; for instance, *baby-sit* comes from *baby-sitter* and *jam-pack* from *jam-packed*. However, it is hard to see why this etymological pattern should make compound verbs less authentic. In Bauer & Renouf's (2001: 109) words, Marchand's (mis)classification is "an uncharacteristic failure to distinguish between a synchronic structure and the diachronic process leading to that structure." Ackema & Neeleman (2004: 59) also argue that "a child learning English does not know how *to baby-sit* was coined [but] simply finds a structure in the input which is best analyzed as an N-V compound." Insofar as the defective category theory is concerned, back-formed MHCs fit in the derivational schema equally

well as non-back-formed ones. It is only when (part of) a nonverbal compound is converted into a verb as a whole (e.g., *lipstick*$_V$ 'apply lipstick' < *lipstick*$_N$, *moonlight*$_V$ 'work at a second job' < *moonlight-er*$_N$) that the etymology may affect synchronic analysis, though even then the wholly converted verb may readily receive an MHC reanalysis when one is possible; for example, the past tense of *joyride*$_V$ 'go on a joyride' < *joyride*$_N$ can be either *joyrided* (as particularly noted in Mencken 2009: 281) or *joyrode* (as given in *The Oxford American Dictionary of Current English*). In the latter form the original nominal head *ride* is reanalyzed as a verbal head, as illustrated in (36a–b); compare it with the conversion structure in (36c).

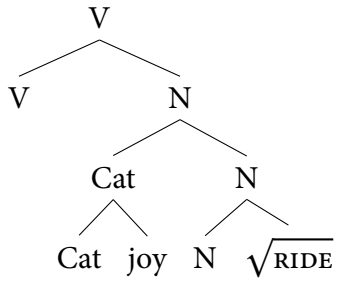
(36) a.



b.



c.



In my informal survey, I collected 200 English verbal MHCs from daily sources (news, books, TV shows, etc.) in a one-year duration. See Table 3.4 for a sample and check Appendix A for the entire list. The survey clearly shows that apart from the often-cited examples (e.g., *baby-sit*, *stir-fry*) many more verbal MHCs are being readily coined by English speakers.

Table 3.4 Verbal modifier-head compounds in English

3D-print	binge-watch	chain-smoke	finger-paint
gene-edit	shame-eat	cold-call	humble-brag
backseat-drive	house-sit	Facebook-stalk	jam-pack
batch-download	spoon-feed	fat-shame	lip-sync
beauty-sleep	butt-dial	time-travel	window-shop

Below are some example sentences attested on Google.

(37) a. I asked Arthur to observe Sherry's behavior with other people, and sure enough, he began to realize that Sherry *backseat-drove* with everyone.

b. This woman *house-sat* for her friends and received a £300 bill.

c. As long as you are out shopping, look for new pillows and sheets …and cute pajamas! Unless you like to *beauty-sleep* in the nude, in which case, go for it!

d. Why certain people may be more likely to *Facebook-stalk* their exes?

Apart from the colloquial registers, various technical registers also witness abundant verbal MHCs. Take generative syntax for example. I find the following terms (among others) in the literature.[36]

Table 3.5 Verbal modifier-head compounds in Chomskyan syntax

set-merge	head-move	left-adjoin
pair-merge	A-move	right-adjoin
self-merge	Ā-move	head-adjoin
late-merge	wh-move	Chomsky-adjoin
parallel-merge	sideward-move	left-Chomsky-adjoin

These are all attested in the third-person singular form, as in (38), which is evidence that they are indeed used as verbs.

(38) a. The adjective *late-merges* into the structure. (Smith 2015: 170)

b. The antecedent component *sideward-moves* to another θ-position. (Hicks 2009: 546)

c. The entire Num *head-adjoins* to the left of D. (Bowers 2018: 140)

Since the colloquial and technical registers are two major domains where new concepts are constantly created, it is normal that they also produce most of the verbal compounds. This clearly shows that when the need for new vocabulary increases, verb compounding can be perfectly harnessed as a lexicon-expanding tool; hence, ostensible categorial biases in compounding are unlikely to be caused by restriction on the compounding mechanism itself.

3.2.3.3 Modifier-head compounding is parameterizable

While there is no way to encode categorial bias in the compounding mechanism, nothing precludes other aspects of modifier-head compounding from being parameterized, for Cat is just another functional category, and according to BCC that makes it an eligible locus of parametric variation. More specifically, in the feature specification of Cat, as repeated in (39), the non-mergeme part can accommodate taxonomic features with parameterizing effects.

(39) $\text{Cat}_{\text{LANG}} := [\langle \textsc{cat}{:}_, \textsc{pos}{:}_\rangle, \textsc{nm}]$

Take linearization for example. Cat may be configured to specify the ordering of the modifier in an MHC with respect to the head—either as modifier≺head (i.e., Cat≺X) or as head≺modifier

[36] In particular, the existence of terms like *left-Chomsky-adjoin* suggests that verb compounding can be recursive just like nominal compounding.

(i.e., X≺Cat). To illustrate, MHCs in English and Chinese have the modifier≺head ordering (40a–b), whereas those in French and Vietnamese have the head≺modifier ordering (40c–d).[37]

(40) a. blackboard, school bus, bookstore [English]

b. *shū-diàn* 'book-store', *yóu-huà* 'oil-painting', *shuǐ-bēi* 'water-bottle' [Chinese]

c. *sauce tomate* 'sauce tomato; tomato sauce', [French]
timbre-poste 'stamp-post; postage stamp',
papier carbone 'paper carbon; carbon paper'

d. *vườn cau* 'garden areca; areca garden', [Vietnamese]
bảng đen 'board black; blackboard',
khăn bàn 'towel table; table cloth'

The above contrast can be expressed as a linearization-related nonmergeme feature on Cat, which I represent informally as "I precede/follow my host," as in (41).

(41) a. $\text{Cat}_{\text{ENG/CHI}}$ ≔ [⟨CAT:_, POS:_⟩, "I precede my host"]

b. $\text{Cat}_{\text{FRE/VIE}}$ ≔ [⟨CAT:_, POS:_⟩, "I follow my host"]

This feature takes effect when the syntactic structure is interpreted at the SM interface. Since everything in the mergeme-based template is an SCDP (§2.3.1.2), this linearization feature is essentially part of what it means to be a Cat-introduced modifier in a particular language.[38] Note two additional points regarding MHC linearization. First, while the examples in (40) are all nominal, the same ordering is seen in compounds of other categories, as in the English and Vietnamese compound verbs below.[39]

(42) a. hand-wash, sleep-walk, stir-fry, cold-call [English]

b. *ăn sáng* 'eat morning; have breakfast' [Vietnamese]
ngủ trưa 'sleep noon; take a siesta'
giặt khô 'wash dry; dry-clean'
ăn xin 'eat ask; beg'

Second, the linearization of MHCs in a language follows the same pattern as that of attributive adjectives. As in (43), adjectives normally precede the nominals they modify in English and Chinese but follow them in French and Vietnamese.[40]

[37] More exactly the Vietnamese pattern is restricted to native compounds. Compounds borrowed from Chinese keep their modifier≺head ordering; e.g., *đại-học* 'big-school; university', *ngữ-pháp* 'language-rule; grammar', *quốc-ca* 'nation-song; national anthem' (Nguyễn 1997: 77).

[38] The idea presented here is similar to Biberauer et al.'s (2014) movement diacritic ^, except that ^ is on head items while "I precede/follow my host" is on nonhead items.

[39] It is not difficult to imagine "anglicized" versions of the Vietnamese examples in (42b); e.g., one may "morning-eat" at home and "noon-sleep" at school.

[40] Some French adjectives are prenominal; e.g., *belle femme* 'beautiful woman', *vieil homme* 'old man', *gros chat* 'big cat'.

(43) a. black board, hot water, quiet night [English]

b. *rè shuǐ* 'hot water' [Chinese]
piàoliàng de huà 'pretty MODI painting; pretty painting'
dà yǎnjīng 'big eye'

c. *sac vert* 'bag green; green bag', [French]
voiture allemande 'car German; German car',
film triste 'film sad; sad movie'

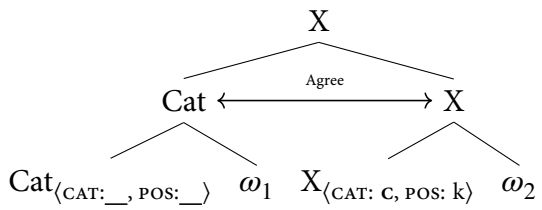
d. *quyển-sách thú-vị* 'book interesting; interesting book', [Vietnamese]
cây xanh 'tree green; green tree',
thành-phố cổ 'town old; old town'

This situation is expected if, as I speculated in §3.1, the applicability of the defective category theory is not restricted to modifier-head compounding but covers a particular type of adjunction in general. As such, when a linearization feature is postulated on Cat, this may affect a group of constructions including modifier-head compounding, adjectival attribution, and so on.

3.2.3.4 Modifier-head compounding blocks head movement

The last implication of the defective category theory I want to highlight concerns the correlation between modifier-head compounding and head movement. In short, MHCs cannot go through head movement because they are not syntactic heads, nor can their lexical heads move because those are blocked in situ by a minimality condition. Before I can elucidate this prediction, however, I need to first review the notion head. The defective category schema is repeated in (44).

(44)

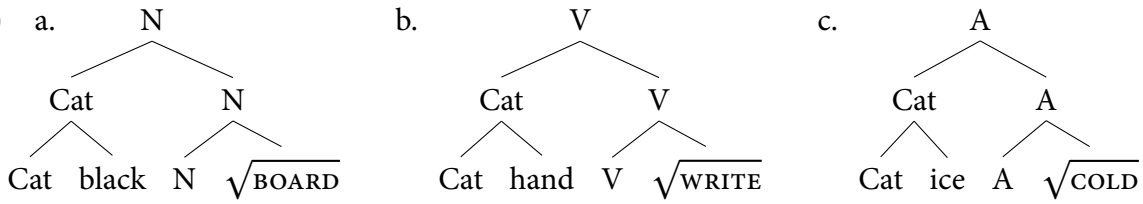


In the above configuration X and Cat are clearly heads.[41] But once we look at concrete examples like those in (45), we immediately find a significant mismatch between heads like X/Cat and heads in head movement theories.

And some English adjectives are postnominal; e.g., *secretary general*, *court martial*, *bargains galore*, *proof positive*. While I must leave such complications to future research, this combined with the observation in note 37, p. 71 seems to suggest that the parameterization of Cat-based phenomena is stratified into different levels (e.g., macro, micro) in some format similar to Biberauer & Roberts' (2015) parameter hierarchy.

[41] ω_1 and ω_2 are not heads, as they do not project/label any phrase structure of their own (in the current derivational layer). Hence, *head* and *terminal node* are not synonyms in my system.

(45) a. b. c.



A head in standard GB/minimalism is a Chomskyan LI like *board*, *write*, or *cold*. This is also what *head* refers to in classical theories of head movement (e.g., Chomsky 1957, Travis 1984, Baker 1988; see Roberts 2010, 2011 for overviews); thus, by "V-to-T movement" what is meant is that an LI of category V (e.g., *write*), not the abstract category V itself, moves to a syntactic position of category T^0. However, this definition of head movement loses ground as we push analytical granularity to the root level, where an item like *write* is no longer a syntactic head but a vocabulary item spelling out a minimal phrase {V, $\sqrt{\text{WRITE}}$}. In other words, we no longer have a neat correspondence between LIs and syntactic heads, and the classical perspective on head movement must be properly reformulated if we want to maintain its insights. I do not have the space to explore how the reformulation might proceed, but I should emphasize that among various possibilities the "root movement" or "root incorporation" method, albeit widely practiced in the literature (e.g., Harley 2009a, Borer 2013, Bauke 2014), is incompatible with my system,[42] because my definition of the Root category in Chapter 2 precludes it from being targeted by any feature-driven (i.e., nonfree) syntactic operation. Insofar as my following discussion is concerned, all we need to know is that *head* in classical head movement corresponds to {X, root} in my system; namely, an overt or root-supported head (see Chapter 4). Henceforth, I assume without arguing that a root-supported head can undergo classical head movement (however reformulated) even though it is technically a nonhead at the granularity level I work with, and by *head movement* I mean the movement of a root-supported head (i.e., a classical head).

So, what happens if an MHC appears in a context with a head movement requirement? Suppose (44) is embedded in a structure where a higher category K attracts a lower category X, for example, via probe-goal agreement (Roberts 2010). If syntactic operations like Agree and Move proceed based on labels, then all three X nodes in (44) qualify as goals since their labels are identical. According to Chomsky's (1995: 311) minimal link condition (MLC), a minimality condition as stated in (46), in this scenario it is the highest X node that gets attracted by K and move.

(46) Minimal link condition
K attracts α only if there is no β, β closer to K than α, such that K attracts β.

However, the highest X in (44) cannot land in a head position, for it is not a syntactic head in its derivational layer but an ordinary (i.e., set-merged) phrase; it merely looks like a head at the interface as a result of defective category agreement, which produces a two-segment X. Since the syntactic head (the lowest X) cannot be attracted and the attracted node (the highest X) cannot

[42] Though a PF-oriented version of this method might be compatible; see De Belder (2017) for a possibility.

head-move, the derivation enters a dilemma and halts. In conclusion, MHCs are incompatible with contexts with a head movement requirement; and if my discussion in this section is on the right track, it could be used as evidence for the syntactic nature of (certain) head movement phenomena (in line with Roberts 2010).

Why can't the head movement rule specifically target a syntactic head? There are perhaps two reasons. First, a movement rule sensitive to the head/nonhead distinction must have access to more information than labels can provide. More exactly it needs bar-level information; hence, such a rule is at odds with the minimalist spirit. What is more, since the highest X in (45) can further project (e.g., a transitive compound verb like *hand-write* can take a direct object), to distinguish all the X nodes we need at least four bar levels, which is unusual practice even back in the heyday of GB. Second, even if we allow the movement rule to recognize different SO sizes, say at least a head/nonhead distinction, we would still end up with a wrong prediction, because the syntactic head in (45), namely the lowest X, is not the classical head that is supposed to move, namely $\{X, \omega_2\}$. And there is no evidence that an abstract categorizer can move without its root. Therefore, I conclude that we had better not (re)introduce a bar-level-sensitive movement rule but let the desired effects be derived via other means.

To illustrate the incompatibility between modifier-head compounding and head movement, below I present a typological pattern it gives rise to concerning verb movement, as stated in (47). Note that this quasi parameterization of modifier-head compounding does not conflict with the conclusion in §3.2.3.2, since it is not encoded in Cat but postulated as an independent constraint.

(47) A language can have (productive) modifier-head verb compounding iff it has no verb movement requirement.

I restrict my discussion to V-to-T/C movement, abstracting away from movement within the VP (or *v*P) domain, because argumentation for/against the latter crucially depends on one's view about VP-internal structure, particularly on issues like the nature/number of the little *v* head(s) and the way roots get involved in structure building. For instance, does "V-to-*v* movement" mean moving the verb to a certain voice/transitivity head or moving the verbal root to the verbalizer (or a combination of the two)? A standard minimalism practitioner may prefer the former, a DMer subscribing to radical decomposition may prefer the latter, a DMer allowing functional category bundling may prefer a combination, while the question may never arise for a spanning theorist. Given the considerable controversy and complication beyond the purpose of this section, I will not extend the prediction in (47) to the VP-internal level, though see Song (2017a) for a relevant suggestion (the general idea is that any *v* head may serve as a locus of MHC formation).

The prediction in (47) in borne out if we compare English and Romance languages. It is well established that the former does not have the type of verb raising attested in the latter (see Pollock 1989, Biberauer & Roberts 2010, Schifano 2018 inter alia), as in (48) (French) and (49) (Spanish).

(48) a. John often kisses Mary. [English]

b. *Jean embrasse souvent Marie.* [French]
John kisses often Mary
'John often kisses Mary.' (Pollock 1989: 367)

(49) a. What did Mara buy yesterday? [English]

b. *¿Qué compró Mara ayer?* [Spanish]
what bought Mara yesterday
'What did Mara buy yesterday?' (Suñer 1994: 335–336)

Accordingly, neither French nor Spanish has the type of compound verb productive in English, as in Table 3.6, which shows that English verbal MHCs are all translated periphrastically.[43]

Table 3.6 Translations of English compound verbs in French and Spanish

English	French	Spanish
hand-wash	*laver à la main* 'wash by hand'	*lavar a mano* 'wash by hand'
dry-clean	*nettoyer à sec* 'clean in dry'	*limpiar en seco* 'clean in dry'
sleep-walk	*marcher en dormant* 'walk sleeping'	*caminar dormido* 'walk sleeping'
double-check	*revérifier* 'reverify'	*volver a revisar* 'inspect again'
window-shop	*faire du lèche-vitrines* 'do lick-windows'	*ir de escaparates* 'go of shop windows'
baby-sit	*faire du baby-sitting* 'do babysitting'	*hacer de canguro* 'do kangaroo'
hitch-hike	*faire du stop* 'do stop'	*hacer autoestop* 'do car-stop'

In both French and Spanish there is a nonnegligible number of nominal MHCs (Ten Hacken 2016, Kornfeld 2009), as exemplified in (50). This means that modifier-head compounding as a word-creating mechanism is active and that it is only verb compounding that is blocked.

(50) a. *sauce tomate* 'sauce tomato; tomato sauce' [French]
maître chien 'master dog; dog trainer'
timbre-poste 'stamp-post; postage stamp'
bande-son 'band-sound; soundtrack'

[43] Retrieved from `https://www.collinsdictionary.com` and `https://dictionary.cambridge.org`

b. *baloncesto* 'ball.basket; basketball' [Spanish]
célula madre 'cell mother; stem cell'
bocacalle 'mouth.street; street intersection'
perro policía 'dog police; police dog'

Meanwhile, East Asian languages like Chinese and Japanese do not have verb movement like English, and they do have productive modifier-head verb compounding, also like English.

(51) a. *shǒu-xǐ* 'hand-wash' *tàn-kǎo* 'charcoal-broil' [Mandarin]
fān-chǎo 'stir-fry' *gān-xǐ* 'dry-clean'

b. *uti-korosu* 'shoot-kill; kill by shooting, shoot to death' [Japanese]
arai-otosu 'wash-make drop; make (dirt, mud, etc.) drop by washing, wash off'
furi-mazeru 'shake-mix; mix by shaking, mix up'
naki-sakebu 'cry-shout; shout while crying, scream'

The Chinese compound verbs in (51a) clearly pattern like English as mentioned in §3.2.3.2. That the Japanese compound verbs in (51b) also fit in a Cat-based analysis is supported by similar verbs like *uti-korosu* 'shoot-kill; kill by shooting, shoot to death', *naguri-korosu* 'strike-kill; kill by striking, strike to death', *kiri-korosu* 'slash-kill; kill by slashing, slash to death', *osi-korosu* 'crush-kill; kill by crushing, crush to death', *yaki-korosu* 'burn-kill; kill by burning, burn to death', and so forth, which show that the nonhead in a Japanese (lexical) compound verb is a modifier of the head (see Tsujimura 2014: 196–197 for a similar remark and Usuki 2007 for a "manner conflation"-based implementation).

An especially interesting situation arises in German (among other West Germanic languages), where the verb movement requirement is inconsistent. In main clauses the finite verb obligatorily raises to the second position (52a), while in embedded clauses there is no such requirement (52b); nonfinite verbs are not raised either (52c).

(52) a. *Petra kommt nach Erfurt.* [German]
Petra comes to Erfurt
'Petra comes to Erfurt.'

b. *Ich weiß, daß Petra nach Erfurt kommt.*
I know that Petra to Erfurt comes
'I know that Petra comes to Erfurt.'

c. *Gestern habe ich fünf neue Apps gekauft.*
yesterday have I five new apps bought
'I bought five apps yesterday.' (Durrell 2011: 451)

This mixed pattern combined with (47) predicts that if German had verbal MHCs at all, they should only be well-formed in embedded clauses or in nonfinite forms but not in finite forms in main clauses. This seems true: German does have verbal MHCs, as in (53), and they indeed can only occur in the above-mentioned contexts (i.e., contexts without V2 requirement), as in (54)–(55). In fact, such complex verbs are so noticeable that they have been singled out in the literature, dubbed as "immobile verbs" in Vikner (2005) and "C^0-phobic verbs" in Ahlers (2010).[44]

(53) *bau-sparen* 'building-save; save with a building society'
bauch-reden 'belly-talk; ventriloquize'
bauch-tanzen 'belly-dance'
schutz-impfen 'protection-vaccinate; inoculate'
berg-steigen 'mountain-climb; go mountain climbing'
ehe-brechen 'marriage-break; commit adultery'
kopf-rechnen 'head-calculate; do mental arithmetic'
bruch-rechnen 'fraction-calculate; do fractions'

(54) a. *Sie will bausparen.*
she wants building-save
'She wants to save with a building society.'

b. *…weil er bauspart*
…because he building-saves
'…because he saves with a building society'

c. **Spart er bau?* / **Bauspart er?*
saves he building building-saves he
Intended: 'Does he save with a building society?' (Vikner 2005: 88)

(55) a. *Erna hat sehr viel bauchgetanzt.*
Erna has very much belly-danced
'Erna has belly-danced a lot.'

b. ?*Unglaublich, dass Emil so gut bauchtanzt.*
unbelievable that Emil so good belly-dances
'(It is) Unbelievable that Emil belly-dances so well.'

[44] See also McIntyre (2002), Fortmann (2007), Freywald & Simon (2007), and Song (2016a) inter alia. In addition, Zeller (2002: §2.2.2) and Haider (2010: 58ff.) focus on a "doubly prefixed" subtype of immobile verb; e.g., *vor-an-melden* 'pre-on-notify; preannounce, preregister', *ur-auf-führen* 'first-up-lead; perform for the first time'. And Koopman (1995) discusses a similar phenomenon in Dutch.

c. **Emil bauchtanzte letzten Sommer.*
Emil belly-danced last summer
Intended: 'Emil belly-danced last summer.' (Ahlers 2010: 16)

Previous accounts for immobile verbs usually resort to a structural uncertainty or a syntactic conflict (either constraint- or rule-based), as summarized below:

- **Structural uncertainty:** Vikner (2005). Immobile verbs are not yet resolved between a V^0 structure (inseparable; e.g., *ver-stehen* 'PRF-stand; understand') and a V^* structure (separable; e.g., *ab-schicken* 'off-send; send off'); they must fulfill requirements of both types (i.e., be simultaneously inseparable and separable), which is impossible.

- **Syntactic conflict:**
  - **Constraint-based:** McIntyre (2002). Immobile verbs are structurally complex (e.g., $[[\textit{ketten}]_N\ [\textit{rauchen}]_{V^0}]_{V?}$ 'chain-smoke', where V? = V^0 or V′) and subject to two constraints, MINIMAL V2 (only attract minimal V head) and BACKFORMATION INTEGRITY (keep phonological similarity between a back-formed verb and its source), which cannot be simultaneously met. Consequently, a third constraint INDECISION (avoid contexts where principles are in competition) takes effect.
  - **Rule-based:**
    - Fortmann (2007). Immobile verbs are syntactically complex (e.g., $[_{V^*}\ \textit{wett}\ [_V\ [_N\ \cancel{\textit{wett}}\ [_N\ [_V\ \textit{renn}]\ \text{-}\varnothing_N]]\ \text{-}\varnothing_V\ \text{-}\textit{en}]]$ 'contest-run; race', where ∅ represents a null categorizer and V^* is as in Vikner 2005) and hence cannot head-move; they cannot strand their nonheads either because that would break an interpretive chain, which is responsible for the complex verb's compositional meaning (Fortmann observes that immobile verbs are interpreted compositionally).
    - Ahlers (2010). Immobile verbs are compounds qua complex heads ($[_{V^0}\ [_{X^0}\ \textit{berg}]\ [_{V^0}\ \textit{steigen}]]$ 'mountain-climb'); hence, their inflection-bearing part (the inner V^0, which is attracted by C^0) is blocked in situ under the lexical integrity hypothesis (Di Sciullo & Williams 1987).

My analysis for immobile verbs is similar to that of Ahlers (2010) in spirit, except for two technical differences: (i) the compound is not a syntactic head but an ordinary set-merged phrase; (ii) the verbal head is blocked in situ not by lexical integrity but by MLC (i.e., minimality). Not analyzing the compound as a V^0 gives us a way to account for the occasional Ā-movement of the nonhead of an immobile verb, which is predicted to be illegal by Ahler's theory; for instance, my informants find the corrective sentences in (56) acceptable (though to varied degrees).[45]

[45] An informant comments that she would not produce (56a) and (56c) in life because *bauchreden* and *bausparen* are "very rare words" and hence "people would prefer leaving [them] as one so that the meaning is more clear." This is somewhat similar to McIntyre's (2002) BACKFORMATION INTEGRITY constraint.

(56) a. (?)*BAUCH möchte er reden, nicht fuß.*
belly would like he talk not foot
'He would like to *belly*-talk, not foot.'

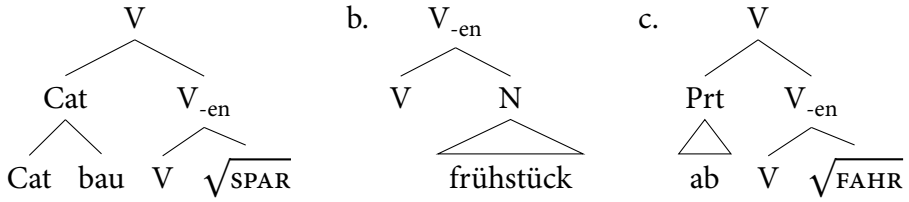
b. *VOR hat er uns angemelded, nicht nach.*
pre has he to.us announced not nach
'He has *pre*-announced to us, not after.'

c. (?)*BAU hat er gespart, nicht Paul.*
building has he saved not Paul
'He has *building*-saved, not puilding.'

Examples like the above suggest that a defective-category-based analysis for immobile verbs can potentially cover more data than its alternatives.

Of course, there is much more to research on immobile verbs; for instance, Vikner (2005) reports that back-formed verbal MHCs in Danish (and other Germanic languages with VO-order) are mobile, and even among those Germanic languages that have immobile verbs there is much interregional/interspeaker variation as to which verbs are mobile and which are not. Ahlers (2011) responds to Vikner's puzzle by proposing that back-formed verbs in Danish are systematically reanalyzed as simple V^0s, and several of the aforementioned authors similarly account for the crosslinguistic/interpersonal variation by multiple possible structures for a back-formed complex verb. These ideas are well compatible with my analysis. Specifically, the following three structures can coexist in one grammatical system (I mark the infinitive suffix *-en* as a subscript to the rooted verbal head):

(57) a. b. c.



The three structures respectively represent ordinary verbal MHCs (*bau-sparen* 'building-save; save with a building society'), which are immobile in German; simple verbs with a conversion etymology (*frühstücken* 'have breakfast'), which are mobile as a whole; and particle verbs (*ab-fahren* 'off-drive; leave'), which are mobile and separable. The only structural difference between (57a) and (57c) is that Cat is present in the former but absent in the latter. Consequently, *bau* in (57a) is an adjunct of the verbal head, while *ab* in (57c) is a complement. McIntyre (2002: 7) notes that back-formed immobile verbs in common usage may eventually "become fixed in either the inseparable or the separable use." In my terminology, this is a change from (57a) to either (57b) or (57c). Based on the foregoing discussion, I tentatively propose the typology in Table 3.7.

Table 3.7 Correlation between verb movement and modifier-head verb compounding

Verb movement	Verbal MHC	Example
Yes	No	Romance languages OV-Germanic languages, V2 contexts VO-Germanic languages(?)
No	Yes	OV-Germanic languages, non-V2 contexts English, East (and Southeast?) Asian languages

Some other facts follow from the typology in Table 3.7. For instance, while Romance languages and Vietnamese both have the head≺modifier order in nominal MHCs, the former do not have verbal MHCs while the latter does; and even though various nominal MHC calques have entered Romance languages from English (e.g., French *taxe carbone* 'tax carbon; carbon tax', Spanish *balompié* 'ball-foot; football'), no calques of verbal MHCs have made their way in (as evidenced by the periphrastic expressions in Table 3.6). Finally, while I remain agnostic about the exact situation in non-English VO-Germanic languages (there is significant interregional/interpersonal variation; Timo Ahlers, p.c.), the typology in Table 3.7 does not force us to group English with other VO-Germanic languages with respect to verb compounding. Back-formed complex verbs may well systematically assume the structure in (57a) in English but that in (57b) in Danish. I speculate that (57a) may be the default structure for MHC-based backformation, for it is the readily available one (with just a change in X), and any reanalysis results from extra syntactic pressure; thus, a fully V2 system and a partially one, or a [+V2, +VO] system and a [+V2, −VO] one, may lead to different results. These issues of microvariation merit further research.

3.2.4 Remaining questions

I list three remaining questions before finishing this section:

1. **Can the proposed theory be applied to prepositional modifiers?**
   In §3.2.1 we saw that the modifier material of an MHC (ω_1) could have various categorial origins. I discussed N, V, A but set aside P, despite apparent examples like *uptake*, *download*, and *aboveboard*. While these fit well into a Cat-based analysis, the picture becomes a lot murkier if we look beyond English. In German, for instance, complex verbs with (single) P-origin preverbs are either separable (e.g., *ab-fahren* 'off-drive; leave') or inseparable (e.g., *über-setzen* 'over-put; translate') but never immobile; hence, they are not MHCs. Similarly, in Romance languages there are many P-source prefixes, as in French *sur-face* 'sur-face', *em-barquer* 'em-bark', *anté-natal* 'ante-natal' and Spanish *pre-aviso* 'fore-warning' , *pos-poner* 'post-pone', *sub-urbano* 'sub-urban', yet neither do the nominal examples show the

noun≺modifier order nor do the verbal examples shun verb movement. Thus, complex words with P-origin prefixes are unlikely to have an MHC structure in syntax even though their prefixes semantically modify their lexical heads.

2. **Can the proposed theory be extended to more bound roots?**
Another type of special compounding material we saw in §3.2.1 was the bound root; namely, a root with no independent categorization or lexicalization outside the compound environment. We saw that either the modifier material or the head material could be a bound root (e.g., $\sqrt{\text{CRAN}}$-*berry*, *bride*-$\sqrt{\text{ZILLA}}$). However, trouble arises when we attempt to extend the analysis crosslinguistically. For instance, while *insect-ology* may be given an MHC structure in English, its cognate in French *insectologie* may not, again because French MHCs assume the head≺modifier order (e.g., *sauce tomate*). This situation is even clearer in quasi-complex words where both the quasi-modifier and the quasi-head are bound roots, such as *psych-ology*, *biblio-phile*, and *con-cept*. All these words have cognates in other European languages, but none of them shows MHC-like behavior or variation.

3. **Does Cat have dedicated exponent(s)?**
When introducing Rubin's (2003) theory of adjunction (§3.1.2.3), I endorsed his proposal that modification-specific morphemes like Mandarin *de* and Tagalog *na* are exponents of the adjunct-introducing functional head (Mod for Rubin, Cat for me). However, further reflection reveals that this may need more justification. We ideally want to apply the functional head to a variety of adjunction scenarios, but its exponent seems to vary from case to case. In Mandarin, *de* only appears in phrasal modification but not in compounds or other phenomena that otherwise seem to fit into a Cat-based analysis (such as the sentence-final particles to be discussed in the next section). Similarly, in German the compound-specific linking elements are plausible candidates for Cat exponents, but they are absent from other adjunction scenarios. As such, whether and how Cat could be associated with overt markers is open to further research. A possibility is to assign exponents to Agree-valued versions of Cat; for instance, a Cat valued by a big X category may get a different exponent from a Cat valued by a non-big-X category. This could be the case in German, where linking elements accompanying nominal-origin modifiers may appear in MHCs of all big X categories, such as *Arbeit-s-zimmer* 'work-LK-room; study', *zwang-s-räumen* 'force-LK-remove; evict', and *geist-es-krank* 'spirit-LK-ill; mentally ill'. A similar situation is observed in Japanese, where the so-called continuative form (*renyookei*: *-i* after consonant-ending morphological roots and ∅ after vowel-ending ones; Tsujimura 2014: 168) accompanying verb-origin modifiers may appear in compounds of all big X categories, such as *nom-i-mizu* 'drink-CONT-water; drinking water', *yude-∅-tamago* 'boil-CONT-egg; boiled egg', *fur-i-mazeru* 'shake-CONT-mix; mix by shaking, mix up', and *mus-i-atui* 'steam-CONT-hot; steaming hot'. We could speculate that, in the defective category schema, when ω_1 is nominal in German or verbal in Japanese and X is a big X category, Cat is spelled out by a

certain linking element or the continuative form (which particular element surfaces is determined by the entire compound in German and by ω_1 in Japanese).[46] Such an analytical possibility is unavailable in Rubin's theory.

Questions 1 and 2 together suggest that the defective category theory is more suitable for analyzing phenomena involving structures larger than a head (X^0); namely, phenomena that are "more syntactic" in nature. Question 3 reflects potential interaction between Cat and elements in its local domain (X, ω_1, ω_2).

3.3 Sentence-final particles

3.3.1 Introduction

Sentence-final particles (SFPs) are a special class of vocabulary items in (South)East Asian languages; they encode a wide range of discourse meanings. My discussion in this section is mainly based on Mandarin Chinese; see (58) for examples.

(58) a. *tā chū-qù mǎi dōngxi le* [Mandarin]
he exit-go buy thing SFP
'He's gone shopping.'

b. *tā hěn hǎo-kàn ba*
she very good-look SFP
'She's very good-looking, right?'

c. *nǐ hǎo ma*
you good SFP
'How are you?'

d. *xiǎoxīn ou*
be careful SFP
'Be careful, okay?' (adapted from Simpson 2014: 157)

In (58) *le* asserts a currently relevant state (CRS), *ba* solicits the listener's agreement, *ma* signals a yes/no question, and *ou* conveys a gentle reminder. I have avoided using the term *marker*—*le* is not a dedicated CRS marker, nor is *ma* a dedicated yes/no marker. In Mandarin a CRS may well be expressed without *le* (59a), and a yes/no question can also be formed without *ma* (59b).

[46] Two further similarities between German linking elements and Japanese continuative form: (i) in both languages the infix is semantically empty and may have a purely (morpho)phonological status (Nübling & Szczepaniak 2008, Nishiyama 2016); (ii) in both languages the same forms have independent functions outside compounds—case/number-marking in German (e.g., *Lehrer-s* 'teacher-GEN') and "infinitive"-marking/deverbal-noun-forming in Japanese (e.g., *kat-i-ni iku* 'win-INF-to go; go to win', *kat-i* 'win-N; victory'; Tsujimura 2014, Nishiyama 2016).

(59) a. *cái gāng wǔ diǎn, shū-diàn biàn yǐ guān-mén (le)*
just barely five o'clock book-store simply already close-door SFP
'It's only five and the bookstore has already closed.'

b. *nǐ zuìjìn kě hǎo (ma)*
you recently at all good SFP
'How are you recently?'

In nonconversational contexts (e.g., literary proses, diaries, letters) the SFP-less versions in (59) sound more natural. Such optionality suggests that the indispensable functionality of SFP is not grammatical marking but a provider of speaker/listener-oriented information; namely, SFPs put the materials they are attached to into interactive or communicative uses.

This hypothesis does not contradict the widespread idea that SFPs may be subclassified; however, it does mean that whatever the subclassification is, speaker/listener involvement is a basic hallmark of all SFPs. Hence, there is no such dedicated subclass as "attitude SFP" (pace Zhu 1982, Paul 2015), for all SFPs are mingled with speaker attitude; thus, *le* not only indicates a CRS but more importantly asserts it to the listener, which differentiates it from other CRS-expressing means in the language. This situation is reflected in decompositional analyses, according to which many SFPs are not monomorphemic. For instance, Sun (1999: 184) decomposes Mandarin SFPs based on four vowel series, each presumably associated with an emotional class,[47] as in Table 3.8a; Sybesma & Li (2007: 1773) similarly decompose Cantonese SFPs based on a division of the Cantonese syllabic structure, as in Table 3.8b.

Except *i*- and *w*-, which are presumably byproducts of phonological fusion,[48] all of the rest consonants in Sun's decomposition, once completed with a basic vowel *a* or *e*, fall at the low-to-mid end of Paul's (2015) three-layered SFP classification, while all the vowels as standalone SFPs fall at the high end. I illustrate this correspondence in Table 3.9. The approximate function column is mainly based on Hu (1981) but also takes Li & Thompson (1981), Zhu (1982), Sun (1999), Simpson (2014), and Paul (2015) into account.[49]

Apart from their basic syllabic forms in Table 3.9, the four consonants *l*, *m*, *n*, and *b* may also combine with other vowels, in which case they would largely take on those vowels' emotional values, though note that the exact interpretation of an SFP often involves further lexicalization (a point I will return to in the next section). In nondecompositional analyses (e.g., Zhu 1982, Paul 2015) such combinations are usually treated as fusions of two complete SFPs, as in (60).

[47] According to Sun (1999: §4.2–4.3), *a* encodes a default or general emotionality, *ou* an attention-seeking or reminding emotion, and *e* (or its variants *ei*/*ê*) a fact-based or confident emotion. Sun does not specify the emotional class of *o*, but see Table 3.9 and note 49, p. 83 for a possibility.

[48] e.g., *nǐ a* 'you SFP' → *nǐ ya*, *hǎo a* 'good SFP' → *hǎo wa*.

[49] Among these sources only Sun (1999) mentions *o*, but this SFP has become increasingly popular in recent years, especially in internet language (Chen 2009, Li 2017).

Table 3.8 Decomposition of Chinese sentence-final particles

(**a**) Mandarin

	∅	***i-***	***n-***	***l-***	***b-***	***m-***	***w-***
-a	*a*	*ia*	*na*	*la*	*ba*	*ma*	*wa*
-o	*o*	*io*	*nio*	*luo*	*bo*	*mo*	*wo*
-ou	*ou*	*iou*	*nou*	*lou*	*bou*	*mou*	(*wou*)
-e	*e*	*ie*	*ne*	*le*	*bei*	*me*	*wei*

(**b**) Cantonese

Initial	**Rhyme**	**Coda**	**Tone**
g(e)3: asserting relevance	*aa*: smooth alerting	*k*: intensive emotion	*1*: looking-forward/ hearer orientation
l: realization of state	*e*: suggestive		
m: yes/no question	*o*: noteworthiness		*4*: speaker orientation
n/l: evaluative	*aa4*: question		*5*: evidentiality
z: restriction	*e*: default		*3*: default

(60) a. *xià-yǔ* *la*(= *le a*)
fall-rain SFP
'It's raining! (emotional assertion of a CRS)'

b. *hǎo-hǎo* *shuō* *bei*(= *ba ei*)
good-good talk SFP
'Just explain yourself nicely (and you'll be fine)! (softened command, urging tone)'
(based on Zhu 1982: 207–208)

In addition, the four basic consonant-based SFPs (*le*, *ma*, *ne*, *ba*) all have lexical origins in history, as described in Table 3.10 (Sun 1999),[50] while the vowel-based ones are born SFPs—they seem to be directly established on major vowels available in the language.

Among the four basic consonant-based SFPs in Table 3.10, *le*—or more exactly its consonant part *l-*—is the only one that has gone through step-by-step grammaticalization from the bottom up (in the sense of Roberts & Roussou 2003). This is in line with the general assumption that *le* is the most grammatically oriented (or lowest) SFP in Mandarin (see Erlewine 2017 for a proposal). By comparison, the others have been directly recruited from adverbials (*ma*, *ne*) or quasi adverbials (*ba*). So, these other SFPs (as well as the vowel-based ones in Table 3.9), if they are associated with functional categories, are actually "grammaticalized" from the outside in.[51] While

[50] The MidC pronunciations are based on Wang Li's reconstruction (`http://xiaoxue.iis.sinica.edu.tw`).
[51] This idea converges with that in Biberauer (2018).

Table 3.9 Basic sentence-final particles in Mandarin and their functions

SFP	Approximate function	Paul's hierarchy
le	CRS	low C∼Force (mid C)
ma	yes/no question	
ne	emphasizing, soliciting special attention	
ba	uncertainty, tentativity	
a	general emotion or softening	Attitude (high C)
ou	gentle reminder	
o	a softer variant of *ou*	
ei/e	attracting attention	

Table 3.10 Etymologies of basic consonant-based sentence-final particles in Mandarin

SFP	Etymology
le	MidC **lieu* 了 'end (v)' (= Mandarin *liǎo*) > completive verbal complement > perfective marker > SFP for CRSs
ma	MidC **mǐu* 無 'not have (v)' (= Mandarin *wú*) > sentence-final adverbial '… or not' > SFP for yes/no question
ne	MidC **dzɒi-lǐə* 在裏 'be.at-inside; be inside (VP)' (= Mandarin *zài-lǐ*) > sentence-final adverbial 'surely true' > SFP for affirmation (shortened to *li*) > merged with MidC interrogative SFP **nǐə* (= Mandarin *ni*) > lightened to *ne*
ba	MidC **bǐe* 罷 'stop (v)' (=Mandarin *bà*) > completive verbal complement > SFP for request or suggestion '… and that's it'

this may look bizarre, it has a natural place in the defective category theory (see §3.3.3).

3.3.2 Similarities with modifier-head compounds

There are three remarkable similarities between SFPs and MHCs. First, they both recruit materials from elsewhere in the language. The modifiers and heads of MHCs are recruited from root-containing items, while most SFPs are recruited from (quasi) adverbials or interjections. Recruiting existing material as a means of lexicon expansion is much less restricted than grammaticalization. Thus, we often see new SFPs emerging in an almost random fashion, as in (61).

(61) a. *xièxie ha* [Mandarin]
thanks SFP
'Thanks! :-)'

b. *hǎo da*
good SFP
'Okeydokey!'

c. *yào kū re*
will cry SFP
'I'mma cry.' (from Baidu)[52]

All three SFPs in (61) have gained popularity in internet language: *ha* is recruited from the sound of laughter and used as a smiley version of *a* to make the conversation friendlier, *da* is originally a fusion of *de* + *a* but has been independently lexicalized as a means of sounding cute and innocent, and *re* is a deliberately distorted variant of *le*, conveying a jokingly pretentious attitude.[53] The past decade has witnessed the development of many such novel SFPs, but the most striking innovation in current internet language is the surge of emoticons and emojis. These are often used in a quasi-SFP fashion—call them *visual SFPs*—as exemplified in (62).

(62) a. *zhīdào le* ~
know SFP EMOTICON
'Got it!'

b. *nǐ zài gàn shénme* 🙃
you be at do what EMOJI
'What are you doing?'

The wide tilde ~ in (62a) is used not as a punctuation mark but as an emoticon showing that the speaker is friendly and sincere, while the same sentence without it may sound nonchalant (Chen 2017).[54] Similarly, the upside-down smiley in (62b) is used to convey an amicable curiosity and create a relaxed atmosphere; the same sentence without it has a higher chance to be understood as an abrupt interrogation. While these two symbols largely correspond to the auditory SFPs *o* and *ya*, many more emoticons/emojis have no auditory counterparts, as the emotions they encode are rather complex or subtle. See (63) for some real-life examples.

[52] Baidu is the Chinese counterpart of Google.

[53] *Re* was initially used by fans to mock the pop singer Jolin Tsai but quickly expanded to ordinary contexts. Similar novel SFPs include *lu* (≈ *le*), *nu* (≈ *le*), *hou* (≈ *a*), etc., and the entire subcultural phenomenon they are part of is called "Linguage" (the language of Lin; see Ye 2007 for a sociolinguistic investigation).

[54] Despite the high popularity of this speaker-attitude-conveying usage of ~ in Chinese (many young netizens apply it to every sentence in their text messages), it is much less common in English.

(63) a. *jiù zhèyàng* O_o?
just like this EMOTICON
'Just like this? (I didn't expect that.)'

b. *dēng bú liàng le* →_→
light not bright SFP EMOTICON
'The light is kaput. (Life sucks.)'

c. *qiú jiàgé* 🙏
beg price EMOJI
'What's the price (pleeease)?'

d. *méi yìsi* 🙄
not have meaning EMOJI
'Boring (whatever / I don't care).' (from Baidu)

In (63a) the emoticon O_o (a pair of eyes of different sizes) expresses a confused or surprised feeling; it is used within the scope of the question mark and hence intended as part of the sentence, reflecting the speaker's perplexed tone behind the screen. Similarly, in (63b) the emoticon →_→ (rightward-looking eyes) expresses a frustrated and jokingly world-weary attitude, with which the speaker laments his bad luck. As for the emojis, the praying gesture in (63c) adds to the eagerness of the request and largely corresponds to an emphasized *please*, and the eye-rolling face in (63d) indicates that the speaker is rather bored and uninterested. Syntactically, the visual SFPs above have the same distribution as the vowel-based conventional SFPs in Table 3.9, as they can be (i) directly attached to SFP-less sentences as in (62b) and (63a, c, d), (ii) combined with consonant-based SFPs as in (62a) and (63b), (iii) repeated (or lengthened) for emphasis as in (64a), and (iv) used alone as "visual interjections" as in (64b).

(64) a. *wèi-shénme rén shēng-bìng le bù néng chī hǎixiān* ~~~(≈aaa)*!*
for-what human get-illness PERF not can eat seafood EMOTICON
'Why can't one eat seafood after falling ill! (I'm eager but don't mean to interrogate.)'

b. —*qiángxíng ànmó* —😂(≈ha)
by force massage EMOJI
'—Just force-massage yourself (to get slim). —Hahaha (I'm laughing my tears out)!'
(from Baidu)

The foregoing illustration shows that the inventory of SFPs can be arbitrarily enlarged as long as there are enough recruitable materials. Before the internet era the number of SFPs had been rather limited due to the limited number of emotion-conveying materials (mainly vowels) in the language, but with the advent of the internet the quantity of such materials has greatly increased with the aid of an additional (visual) modality. What is more, the visual materials can efficiently

convey much more sophisticated emotions, some of which (e.g., 💃, 💅) are barely effable, though they are perfectly integrated into text-based communication, as in (65).

(65) a. *wǒ zhǎngde dàodǐ zěnmeyàng, lián wǒ yě bù zhīdào* 💃
I look on earth how even I also not know EMOJI
'How on earth do I look? That is unknown even to me. (confident tone)'

b. *nǐ bìng bù dǒng wǒ* 💅
you at all not understand me EMOJI
'You don't understand me at all. (posh tone)' (from Baidu)

The way new materials are recruited as SFPs, each for a subtle emotion, highly resembles the way the basic lexical vocabulary is established: via the creation and lexicalization of a multitude of roots, each for a subtle concept. So, contrary to the common assumption, SFPs are more akin to an open-class, lexical category than a closed, grammatical category.

A second similarity between SFPs and MHCs is the lack of categorial bias. Just like modifier-head compounding can be applied to any lexical category, so most SFPs in Chinese (perhaps all but *le*) can be applied to phrases of any category, not just sentences (CPs), as illustrated in (66).[55]

(66) a. —*nǐ zài gàn shénme?* —[$_{VP}$ *chī-fàn*] *a*
you be at do what eat-meal SFP
'—What are you doing? —Just eating.'

b. —*nǐ xiǎng diǎn shénme hē-de?* —[$_{DP}$ *yì-bēi chéng-zhī*] *ba*
you want order what drink-NML one-glass orange-juice SFP
'—What drink would you like to order? —A glass of orange juice, please.'

c. —*bīng yóu shénme zǔchéng?* —[$_{NP}$ *shuǐ*] *lo*
ice by what compose water SFP
'—What is ice made up of? —Obviously water.'

d. —*nǐ zuìjìn hǎo ma?* —[$_{AP}$ *hǎo*] *a*
you recently good SFP good SFP
'—How are you recently? —Good. (positive tone)'

e. —*nǐ zhīdào yàoshi zài nǎr ma?* —[$_{PP}$ *chōuti-lǐ*] *ba*
you know key be at where SFP drawer-in SFP
'—Do you know where the key is? —Probably in the drawer.'

[55] Ritter & Wiltschko (2019) propose a "nominal speech act structure," which is built on an extended-projection-based reasoning; i.e., the verbal domain and the nominal domain both have a speech act projection. While I do not have space to review this proposal, my own analysis in the next section is fundamentally different in that I do not place SFPs in any "extended" projection at all. See Biberauer (2017b) for a similar conclusion.

f. —*wǒ jīntiān yíng-le yì-qiān-kuài qián!* —[$_{\text{Interj}}$ *wa-o*] *ne*
I today win-PERF one-thousand-CL money WOW SFP
'—I won 1000 RMB today! —Oh wow!'

As (66) shows, SFPs can be attached to VPs, DPs, NPs, APs, PPs, and even interjections, and such usages are no less frequent than CP-final SFPs.[56] This means that there is no c-selection restriction on SFPs. Hence, the relation between SFPs and the materials they attach to is not one of extended projection or cartographic extension, contrary to a mainstream view (noticeable exceptions are Sz.-W. Tang 2015 and Biberauer 2017b, whose perspectives I am sympathetic to). Interestingly, although the SFP *le* is limited to CPs,[57] the fused form *lo* is not, as evidenced by (66c); the same is true for other complex SFPs with *l(e)* as the first component (e.g., *la*, *lou*, *lei*). In fact, these *le*-based SFPs can be added to *le* without being redundant, as in (67).

(67) a. *zhīdào le la*
know SFP LE
'Ok, ok, I know (stop nagging me).'

b. *mǎshàng dēng-jī le lou*
immediately board-plane SFP SFP
'Yay, I'm gonna board!' (from Baidu)

This distinction between *le* and its derivative SFPs confirms my earlier comment that the content in a particular SFP, whatever its source, is largely a matter of lexicalization; once *lo*, *la*, and the like become independent SFPs, the grammatical restrictions on *le* no longer apply to them.

A third similarity between SFPs and MHCs is in linearization. We saw in §3.2 that the linearization of MHCs is separate from that of heads and complements, because the modifiers in MHCs are not complements but adjuncts; for example, English is head-initial in all the lexical categories (e.g., *book of linguistics*, *write letters*) but English MHCs are always "head-final" (e.g., *linguistics book*, *hand-write*). Similarly, the linearization of SFPs with regard to their hosts seems to be separate from that of heads and complements too; for example, Chinese is head-initial in all the lexical categories like English, yet Chinese SFPs are always "head-final." Beyond Chinese,

[56] A potential question here is how to be sure that the examples in (66), except (66f), involve direct mergers instead of movements to, e.g., Spec-SAP plus ellipses. While I must leave a detailed justification to future research, I present a preliminary line of thought here: In order for an ellipsis-based analysis to work, the SFP head (SA under an extended-projection-based analysis) must be able to probe and attract various constituents to its specifier position; however, all the constituents in (66a–e), being base-generated in the thematic domain (i.e., within the first clausal phase), must go through two intermediate phase edges (Spec-*v*P and Spec-CP) to bypass the PIC. This requires that *v* and C both be [+EPP] in Mandarin and predicts massive scrambling-like movement scenarios unattested in the language. In addition, it is unclear how the nonreferential bare nominal in (66c), which is embedded inside another nonclausal phase (DP), can be probed by SA, not to mention that an ellipsis-based analysis cannot explain (66f) at all. By comparison, the defective-category-based analysis to be presented in §3.3.3 covers all the data in (66) and does not predict unattested phenomena.

[57] Though this restriction is flouted in "Linguage" (note 53, p. 86); e.g., *tiān le lu* 'sky SFP SFP; oh my god', where *le* is attached to a noun and followed by another *le*-based SFP (*lu=le+u*).

the linearization of SFPs seems to be separate from that of heads/complements in other East and Southeast Asian languages as well: the head-complement ordering varies much in the area, yet the speaker/listener-related discourse particles always come after their hosts. An especially interesting case is Singapore English, which has developed Chinese-style SFPs (Gupta 1992, Lim 2007, German & Prévot 2016), and those indeed come after their hosts regardless of the general head-complement directionality in English, as in (68).

(68) *My parents very old fashion ah21? Then your parents leh55?* [Singapore English]
'Are you saying that my parents are old-fashioned? Then what about your parents?'
(Lim 2007: 446)

Besides, if emoticons/emojis are visual SFPs as I have conjectured, then the scope of potential evidence for the non-head/complement-based linearization of SFPs extends way beyond Asia, since emojis are gaining popularity globally, and wherever they are used in a quasi-SFP fashion they are attached to their hosts at the end, as in (69).[58]

(69) a. Time to rewatch the episode cause apparently I wasn't hurt enough 😂 [English]

b. *Bitte jetzt nicht fragen, was das sein soll* 😂. [German]
please now not ask what that be should EMOJI
'Please don't ask now what it's supposed to be (lol).'

c. *Mi madre me ha prohibido recomendar una serie que* [Spanish]
my mother me has prohibited to recommend a TV series that
no esté terminada 😂
not is finished
'My mom banned me from recommending her any TV series that hasn't finished (lol).'
(from Twitter)

The overall positioning of emojis is much more flexible; here I restrict my discussion to their positioning when they are used to add emotional modification to the materials they co-occur with. At least in Chinese, this criterion clearly distinguishes sentence-final and sentence-initial emojis. While sentence-final emojis function like auditory SFPs, sentence-initial ones are either independent interjections (70a)[59] or responses to something in the preceding discourse (70b).

(70) a. 🙂 *zuìjìn xiǎo-xué-shēng yòu kāishǐ kuáng le* [Mandarin]
EMOJI recently small-school-student again begin crazy SFP
'Haha (sarcastic). Recently elementary school students are becoming crazy again.'

[58] While the English and the Spanish speakers in (69a) and (69c) do not use punctuation marks, the German speaker in (69b) does and puts the period on the right of the emoji, which shows that the emoji is part of the sentence.
[59] In fact even the smiley face in (70a) may be a response to some antecedent, because the example is a comment on a reposted video. It is difficult to find examples of sentence-initial emojis truly lacking preceding material.

b. —*shénme jīngxǐ?* —😳 *gàosu nǐ le zěnme xíng?*
what surprise EMOJI tell you PERF how okay
'—What surprise? —Emm (embarassed). How can I tell you beforehand?'
(from Weibo)[60]

Since such sentence-initial emojis are interpreted separately from the materials they co-occur with, they should not be analyzed on a par with sentence-final emojis.

To summarize, there are three major similarities between SFPs and MHCs: (i) both constantly recruit and lexicalize new materials and hence are more lexical than grammatical, (ii) both can be applied to all the lexical categories without selectional restriction, and (iii) both are linearized with respect to their hosts separately from the head-complement linearization mechanism. Thus, we are led to the following hypothesis:

(71) Sentence-final particles and modifier-head compounds have a similar syntactic status.

Next I will show how the defective category theory can help us flesh out this hypothesis.

3.3.3 A defective-category-based analysis

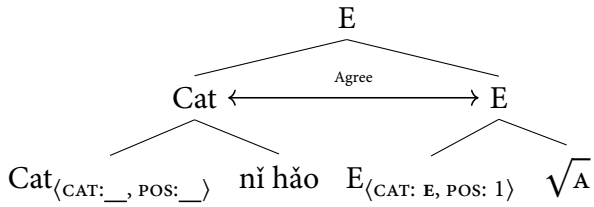
Recall from §3.1 that the defective category schema has four premises: the defective category (Cat), a nondefective category heading the construction (X), some adjunct material (ω_1), and some head material (ω_2). For MHCs, X is a big-X category, ω_1 is the modifier, and ω_2 is the lexical head. What are the premises for an SFP-containing construction if we are to analyze it in the same way? Since an SFP puts its host material into speaker/listener-level use, the overall category of the construction should reflect such a function.

In the literature SFPs and other pragmatic particles are placed in an extended projection—or a cartographic zone of projections—above CP, which contains variedly named illocution-oriented categories like Attitude (Paul 2014, 2015), Speech Act (SA; Speas & Tenny 2003, Hill 2007, Haegeman & Hill 2013, Haegeman 2014), Ground-Response (Heim et al. 2016, Wiltschko & Heim 2016, Wiltschko 2017, 2018), and the like. I have no space to review these previous proposals, but one point is worth noting: even if all the above-mentioned categories are on the right track, Chinese-style SFPs, especially the visual ones, still feel like a different type, for even in languages that fit well into cartographic analyses (e.g., Italian, West Flemish) utterances of various sorts can still be followed by emojis. Thus, I claim that in those languages it is the emojis (in internet language) or suprasegmental elements like intonations and gestures (in spoken language) rather than the SA particles that are on a par with Chinese-style SFPs.[61] I postpone a comparative study of cartographic and anticartographic pragmatic particles[62] to future research; for now I focus on the

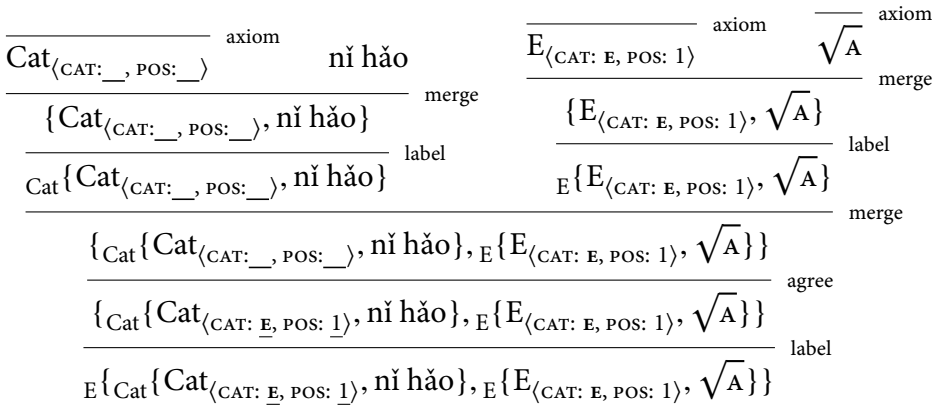
[60] Weibo is the Chinese counterpart of Twitter.
[61] See Heim et al. (2016) for a similar idea.
[62] I use the latter term to refer to pragmatic particles that do not fit into a cartographic or extended-projection-based

Chinese-style SFPs.

To avoid confusion with the cartographic categories, I label the category for SFP-containing constructions as Emotion (E for short); this takes the place of X in the defective category schema. More specifically, Emotion does not encode any concrete emotion in itself but functions like a categorizer; it is the miscellaneous materials categorized by it (i.e., ω_2) that contribute the more substantive emotional content. In this sense, SFPs are like roots. So, in an SFP-containing construction, Emotion is the labeling category and some root-like emotion-bearing material is recruited as ω_2; the {E, ω_2} merger yields a categorized/lexicalized SFP item. Meanwhile, Cat adjoins some other material (ω_1) to this "lexical head" SFP, and since the derivational history of ω_1 does not matter for the defective category schema, ω_1 may be any SO of any category or size (CP, VP, DP, etc.). In Figure 3.4 I illustrate this with the derivation of *nǐ hǎo a* 'Hello!'.



(a) Syntactic tree



(b) Proof tree

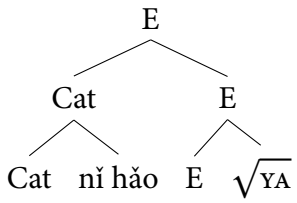
Figure 3.4 Derivational path of Mandarin *nǐ hǎo a* 'Hello!'

There are two major innovations in Figure 3.4. First, the relation between an SFP and its host is one of head and adjunct, not head and complement. The head is the SFP, and the adjunct is the material the SFP attaches to. Note that this attachment is merely phonological, as ω_1 is typically much heavier than ω_2; semantically the center of a Cat-E construction is the {E, ω_2}-conveyed emotion, and the content in ω_1 merely piggybacks on it. Thus, *a* in *nǐ hǎo a* indicates the speaker's emotional involvement (i.e., that he is not nonchalant), and this emotionality wraps the greeting in a specific tone. The same ω_1 material may be wrapped in different tones via different ω_2

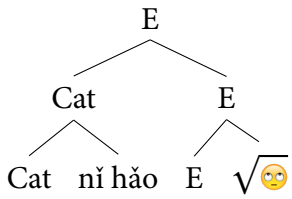
model. The concept is similar to what Biberauer (2017b) calls "acategorial final particle."

materials, as in (72).

(72) a.



b.



In (72a) *ya* indicates a cute and slightly surprised emotion (usually used by girls or when talking to children); in (72b) 🙄 gives the ostensible greeting a sarcastic tone and shows that the speaker is talking to the listener unwillingly and does not care if they are really well (in spoken language this is typically rendered by a toneless utterance that ends abruptly, often accompanied by a lack of eye contact). So, emotional wrapping, which I conceive to proceed via the Cat-E construction, plays a crucial role in interpersonal verbal communication.

Second, the mergeme value of E is ⟨CAT: E, POS: 1⟩. Recall from Chapter 2 that the value of CAT stands for a c-sequence (i.e., a major part of speech). So, setting E as a CAT value is in accordance with our conclusion from §3.3.2 that SFPs are more like a lexical category. The POS value of E is a point of further research. There are two possibilities:

- **Possibility I:** There is only one Emotion category; namely, the c-sequence E is a singleton set {E}. In this case the POS value is trivial since E can only be ordered with respect to itself (i.e., 1 is its identity); however, it is still formally necessary, because CAT merely picks out a c-sequence, and it is POS that further picks out an individual category therefrom. As such, the two parts of a mergeme are always both necessary, because they target information at two interlocking abstraction layers. This separation makes sure that the SCS ontology can normally instruct syntactic derivation even at granularity levels where there is only one c-sequence or only one category per c-sequence.[63]

- **Possibility II:** There are multiple Emotion categories (E_1, E_2, etc.); namely, the c-sequence E is nonsingleton. In this case, further questions arise as to the relative ordering of the multiple E-categories, their derivational path, and so on. Theories assuming a speech act zone (e.g., Wiltschko & Heim 2016) may be compatible with this possibility.

For my current purpose I stick with the simpler possibility I, because there is mostly (if not always) a single SFP unit in each Cat-E construction. By an SFP unit I mean a {E, ω_2} base to which some ω_1 can adjoin. Such a unit may be simple (e.g., *a*, *ne*) or complex (e.g., *ne ba*, *you lei*); the important point is that ω_1 receives emotionality from the SFP unit as a whole rather than successive-cyclicly (i.e., it is not the case that ω_1 first merges with SFP_1, and then ω_1-SFP_1

[63] Impoverished granularity levels may occur at early stages of language acquisition; e.g., at some stage the child may only have one nouny category and one verby category, respectively defined as ⟨CAT: N, POS: 1⟩ and ⟨CAT: V, POS: 1⟩. Since the two categories are readily usable, in the current framework we need the trivial POS value to pick them out of their singleton c-sequences just like we need a nontrivial POS value to pick a category out of a nonsingleton c-sequence at a nonimpoverished granularity level; e.g., ⟨CAT: V, POS: 2⟩ for v from $\{V \leq v \leq T \leq C\}$.

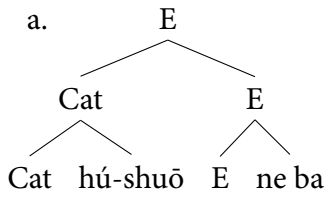
merges with SFP_2). See (73) for examples.

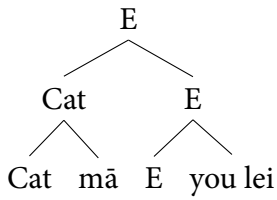
(73) a. *hú-shuō* [*ne* *ba*] [Mandarin]
recklessly-talk SFP_1 SFP_2
'You must be talking nonsense!'

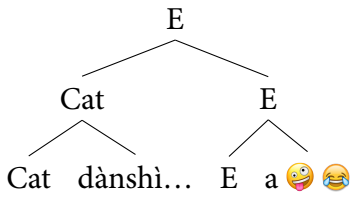
b. *mā* [*you* *lei*]
mom SFP_1 SFP_2
'Mamma mia!'

c. *dànshì* *běi-shàng-guǎng* *yě* *hùn* *bú* *xiàqù* [*a* 🤪 😂]
but Beijing-Shanghai-Guangzhou also get by not down SFP_1 SFP_2 SFP_3
'But it's not easy to get by in the BSG area either (lolol)!' (from Baidu)

In (73a) *hú-shuō* 'talk nonsense' is wrapped in the complex emotion conveyed by *ne ba* 'a very confident guess' in one go, hence the bracketing; similarly, in (73b) *mā* 'mom' is combined with *you lei* 'a strong exclamation' as a whole, and in (73c) the longer material 'but it's not easy to get by in the BSG area either' is rendered with the more complex SFP unit *a*🤪😂 'a self-mocking and helpless emotionality'. The three sentences have the following tree structures:

(74) a.


b.


c.


For now I remain agnostic about how the complex SFP units themselves are derived; it may be desirable to explore possibility II for this purpose. This is reminiscent of MHCs like *vanilla ice cream*, where *vanilla* is adjoined to *ice cream* as a whole rather than first *ice* and then *cream*. As a caveat, "stacked SFPs" involving *le* do not normally form a unit, as in (75).

(75) a. [*tā* *bù* *chōuyān* *le*] *ma*
he not smoke LE SFP
'Does he no longer smoke?'

b. [*bié* *chàng* *le*] *ba*
don't sing LE SFP
'Can you stop singing.' (adapted from Paul 2015: 252–253)

Examples like the above have been taken as evidence for the hierarchical subclassification of SFPs (e.g., Paul 2015). They differ from the examples in (74) in that *le* forms part of the ω_1 material and hence is not a bona fide SFP. The special status of *le* is intuitively clear; unlike other SFPs, *le* has an ordinary grammaticalization route (Table 3.10) and does not convey much emotion

(the act of relating a CRS is not an emotion). Some authors (e.g., Liu 1985, Mei 1994, Yin 2013, Song 2015) further suggest that maybe the CRS-relating functionality is not inherent in *le* either but in another particle *ye*, which is either null or fused into *le* in Standard Mandarin but had been separate in history and is still visible in current Mandarin dialects. Following this view, the sentential *le* is more or less still a tense/aspect marker instead of an SFP; this is also the original view in Zhu (1982), as pointed out by T.-Ch. Tang (2015) (pace Paul 2015).

3.3.4 Further implications

Before ending this section, I briefly discuss a few further implications of the defective category theory for SFPs. First, since ω_2 gets categorized along with the Cat-E construction, we expect {E, ω_2} to be an independently usable item outside the Cat-E construction much like the lexical head in an MHC (e.g., *board* in *blackboard*). This is true for many SFPs, especially vowel-based ones (e.g., *a* 'ah', *o* 'oh', *e* 'uh'). When used on their own, SFPs are just interjections; hence, the category for interjections is also E. As to the consonant-based SFPs, insofar as Sun's list (Table 3.8a) is concerned, all items with onsets *i-*/*w-* and some *n-* items (*na*, *nou*) have interjective uses, while items with the onsets *l-*, *b-*, and *m-* do not. What requires further investigation is whether this interjection-based contrast reveals something about SFP structure (e.g., whether the consonantal onsets *l-*/*m-*/*b-* are more CP-grammatical than emotional as decompositional studies suggest) or is just a matter of lexicalization.

Second, since ω_1 and ω_2 form a quasi compound in the Cat-E construction, various kinds of phonological or semantic fusion may take place between the final element(s) of ω_1 and ω_2. The phonological fusion of SFPs is a well-established observation (e.g., /n/+/a/=/na/, /ŋ/+/a/=/ŋa/), and if my earlier hypothesis that fused SFPs like *la* and *lou* involve separate lexicalization (i.e., their senses are not simply composed from their components) is right, then we also have evidence for semantic fusion. I speculate that the linearized string of a Cat-E construction may undergo a reanalysis, with the final element of ω_1 being reanalyzed as part of ω_2. This is presumably how the sentence-final adverbials in Table 3.10 become SFPs; it is also what happens to *le*, though the reanalysis of *le* may not be mature yet considering its strong tense/aspect reading. In sum, the defective category schema provides two routes for something to be "grammaticalized" as an SFP: (i) via direct recruitment as ω_2 material, (ii) via reanalysis from ω_1 to ω_2.

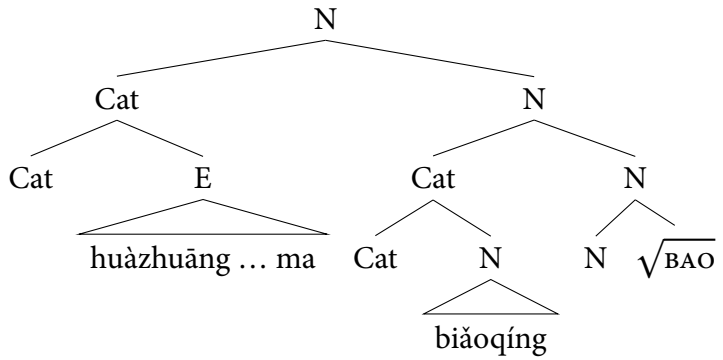
Third, since the category E is neither verbal nor nominal, it in theory cannot be c-selected as an argument or a complement, for it cannot bear case or a theta-role (unlike DP and CP). Thus, we expect SFP-containing constructions to only appear in main "clauses" but not in embedded ones, which is another well-established observation. Meanwhile, since adjuncts need not bear case or a theta-role, we expect that expressions with SFPs may normally be used as adjuncts. This is borne out in examples like the following:

(76) a. *yì-liǎn* [*wǒ rènshi nǐ ma*] *de shénqíng* [Mandarin]
one-CL I know you SFP MODI look
'a do-I-know-you face'

b. *zhè-zǔ* [*wǒ huà-zhuāng hǎo-kàn ma*] *biǎoqíng-bāo*
this-CL I apply-makeup good-looking SFP facial expression-package
'this am-I-good-looking-with-makeup meme package' (from Baidu)

In (76a) *wǒ rènshi nǐ ma* 'Do I know you?' is used as an attributive modifier for *shénqíng* 'look', and in (76b) *wǒ huà-zhuāng hǎo-kàn ma* 'Am I good-looking with makeup?' is recruited as the modifier material for an MHC. See (77) for a tree diagram for the latter.

(77)



The same logic could be extended to explain the following examples, where the SFP appears in a conditional adverbial:

(78) a. [*zǒu ba,*] *bù hǎo;* [*bù zǒu ba,*] *yě bù hǎo*
go SFP not good not go SFP also not good
'It'd be inappropriate if I leave but also inappropriate if I stay.'

b. [*xǐhuān ne,*] *jiù mǎi-xià;* [*bù xǐhuān ne,*] *jiù bié mǎi*
like SFP then buy-ASP not like SFP then don't buy
'Buy it if you like it and don't if you don't like it.' (from Baidu)

Fourth, since E is a mergeme-based category and the Cat-E construction is a constituent, we expect that SFP-containing expressions can be coordinated. This is borne out examples like (79).

(79) a. *háizi, nǐ* [*lěng ma? è ma?*]
kid you cold SFP hungry SFP
'Kid, are you cold or hungry?'

b. [*kǒuhào-shēng* *la,* *luó-gǔ-shēng* *la,* *biānpào-shēng* *la,*]
chant-noise SFP gong-drum-noise SFP firecrackers-noise SFP
xiǎng-chè *yúnxiāo*
resound-through clouds
'Noises of chants, gongs and drums, and firecrackers resound to the skies.'
(from Baidu)

In (79a) the two questions ending in *ma* are coordinated and share a common subject *nǐ* 'you'; in (79b) the three items ending in *la* are coordinated[64] and together function as the subject of *xiǎng-chè yúnxiāo* 'resound to the skies'. In particular, particles like *la* in (79b)[65] have previously been treated as pause fillers (Lü 1999) or dedicated "enumerating particles" (Ma 1992), but they can perfectly fit into an SFP-based syntactic structure in the current approach.

3.4 Chapter summary

The mergeme-based definition for syntactic categories, understood as a UG template, allows for several types of formal flexibility. In this chapter I have explored the theoretical foundation and empirical application of a first type, the defective category (Cat). Cat can merge with any nondefective category X and get its mergeme valued via Agree. This chameleon-like behavior makes it a suitable tool to derive certain adjunction structures in purely featural terms. The defective category schema also provides a categorial skeleton for semantic composition, though syntactic roots do not participate in model-theoretic composition. This result will continue to be useful in Chapter 4, where I explore root syntax in more detail. Compared to three previous generative approaches to adjunction (i.e., axiomatic, derivational, and lexicalist), the Cat-based approach retains their key insights and avoids their problems.

I used two case studies to demonstrate how Cat can be put to empirical use, one on modifier-head compounds and the other on sentence-final particles. The two phenomena, though previously seldom studied together, have a number of similarities. They both recruit materials from elsewhere in the language, both lack intrinsic categorial bias (i.e., can appear in any lexical category), and are both linearized separately from the usual mechanism for linearizing heads and complements. The modifier of an MHC is analogous to the host of an SFP, and the lexical head of an MHC is analogous to the SFP itself. The defective category theory makes a number of predictions in both case studies. For MHCs it predicts that modifier-head compounding is universal, that it has no intrinsic categorial bias, that it can be parameterized, and that it blocks head movement; for SFPs it predicts that SFPs and interjections share the same category, that there are two routes for the "grammaticalization" of SFPs, that SFP-containing expressions cannot appear in

[64] It is normal for coordinators to be null in Chinese.
[65] Similar particles include *a*, *ya*, etc.

embedded clauses but can in adjuncts, and that SFP-containing expressions can be coordinated.

In summary, while the valued/unvalued feature distinction is a common one in minimalist syntax, when applied to the mergeme it gives the system some quite desirable generative power. What is more, this power does not require any other operations than those standardly available in the system; namely, set-Merge, Agree, and Label. All the results in this chapter point to one conclusion: studying the intrinsic flexibility in the categorial system can lead us to novel perspectives on fundamental mechanisms available in human language syntax.

Chapter 4

Root category and semifunctionality

4.1 Generalized root syntax

4.1.1 Root-supported head

Recall the mergeme-based template for syntactic categories, as repeated in (1).

(1) syntactic category$_{\text{LANG, GRAN}}$ ≔ [⟨CAT: cat, POS: nat⟩, NM]

The above template allows for certain types of flexibility in syntactic category formation. In Chapter 3 I studied one type, the defective category, which is characterized by an unvalued mergeme. In this chapter I continue to examine another type characterized by the lack of a mergeme,[1] as in (2); I call this an *acategorial category* since it is not part of any c-sequence (or does not have any "categorial feature" in conventional parlance).[2]

(2) acategorial category$_{\text{LANG}}$ ≔ [(NM)]

The defective category and the acategorial category are both deduced via underspecification; recall from §2.2.1.2 that underspecified features have a special theoretical status in most generative frameworks (which are therefore all nonequipollent). Besides, even the only content in (2) is optional, which gives rise to two subtypes of acategorial category, as in (3). These respectively define *logical operators* (e.g., ∧, ∨) and *roots* (e.g., $\sqrt{\text{DOG}}$, $\sqrt{\text{RUN}}$).

(3) a. logical operator$_{\text{LANG}}$ ≔ [NM]

b. Root$_{\text{LANG}}$ ≔ [ ]

This radically underspecified Root category is the same as that in frameworks like DM and ex-

[1] The material in this chapter is partly based on Song (2018a).

[2] I omit the subscript GRAN as acategorial categories are not subject to granularity variation (see note 1, p. 43).

oskeletal syntax (XS, Borer 2005a,b, 2013), hence my adoption of the term. I further distinguish *Root* and *root*. The former is a syntactic category type while the latter is a specific item in that category; thus, $\sqrt{\text{DOG}}$, $\sqrt{\text{RUN}}$, and the like are roots, not "Roots."[3] Note that the lowercase *root* is still an abstract syntactic notion and should not be confused with the traditional morphological root—henceforth *m-root*, and by analogy a syntactic root is an *s-root* (all my *roots* are s-roots unless otherwise clarified)—for example, $\sqrt{\text{DOG}}$ is an s-root and *dog* an m-root. An m-root is a concrete sound-meaning pair, as in (4a); accordingly, the sort-type `m-root` can be conceived as a product type as in (4b).

(4) a. *dog* ≔ [[PHON: /dɔg/], [SEM: 'a type of animal']]

b. `m-root` ≔ [[PHON: `phon`], [SEM: `sem`]] ≡ `phon` × `sem`

An s-root, on the other hand, contains vague linguistic information that only gets concretized via *categorization* (i.e., merger with a suitable functional category). S-roots, lacking concrete sound/meaning, are of a different sort-type; see (5) for a representation. Note the different value types of the PHON and SEM features in (4b) and (5).

(5) `s-root` ≔ [[PHON: `rphon`], [SEM: `rsem`]] ≡ `rphon` × `rsem`

Abstracting away from detailed phonological/semantic representations, I notate `rphon`-typed values by small capital root names enclosed in forward slashes and `rsem`-typed values by capital root names enclosed in single quotes, as in (6).[4]

(6) $\sqrt{\text{DOG}}$ ≔ [[PHON: /DOG/], [SEM: 'DOG']]

All my above discussion is based on abstract syntactic entities. They may be used as derivational atoms (or axioms in proof-theoretic terms) but do not directly correspond to LIs in standard minimalism. I assume an LI to be a quadripartite feature set, as in (7).

(7) lexical item$_{\text{LANG, GRAN}}$ ≔ [[PHON: `phon`], [SEM: `sem`], [SYN: `syn`], OP]

The PHON and SEM slots in (7) encode phonological and semantic/encyclopedic information; their values are of sort-types `phon` and `sem` (in this respect LIs and m-roots are alike). The SYN slot has a `syn`-typed value, which is just the mergeme-based template in (1). The OP slot is an implementation of Chomsky's (1995) "optional features"; namely, features that only get bundled in when the LI enters the derivation (via the numeration). Typical examples of optional features are φ features. Neither NM nor OP has a suitable sort-type (as their values have no fixed templates), so I leave them as such.[5] Also, optional features are outside syntactic categories and only inside LIs.

[3] I shall never use *Root* in plural (except when *roots* appears in the beginning of a sentence or title), as there is only one Root category in my system; i.e., that in (3b).

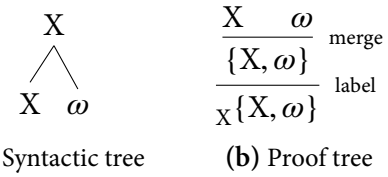
[4] Harley (2014: 253) uses numerical indexing to distinguish one root from another; e.g., $\sqrt{\text{TAPE}}$ is $\sqrt{_{279}}$. However, inasmuch as the indexing serves to individuate roots in syntax, it has the same effect as a small capital root name.

[5] This reflects a distinction between minimalism and HPSG: every complex feature structure must have an explicitly

To repeat Chomsky's (1995: 231) example, *airplane* and *airplanes*, though categorially identical, have different optional features (e.g., [SG] vs. [PL]) and hence count as different (choices of) LIs. Of course, in a decompositional framework LI is not a derivational axiom, so its definition must involve a granularity level indicator. See (8) for example (MP = *The Minimalist Program*).[6]

(8) $$dog_{\text{ENG, MP}} := \begin{bmatrix} \text{PHON} & \text{/dɔg/} \\ \text{SEM} & \text{'a type of animal'} \\ \text{SYN} & \begin{bmatrix} \text{CAT} & \text{N} \\ \text{POS} & 1 \end{bmatrix} \\ \text{OP} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix} \end{bmatrix}$$

Observe that no part in (8) directly corresponds to the root $\sqrt{\text{DOG}}$ in (7). This is because the sound and meaning in (8) belong to the categorized LI, not to the acategorial root. A categorized LI is a syntactic construction comprising at least a non-Root category and a root.[7] I represent root categorization in two formats in Figure 4.1.[8]



(a) Syntactic tree **(b)** Proof tree

Figure 4.1 Derivational schema for root categorization

The schema in Figure 4.1 is very tolerant. On the one hand, ω can be either an atomic root (i.e., an axiom) or a root-like chunk (i.e., a lemma); on the other hand, X can be any non-Root category. This latter point is a major difference between my perspective on root categorization and previous perspectives. In previous theories (especially DM) only the traditional lexical categories or some equivalents thereof[9] may categorize roots, as reflected in the "categorization assumption":

> Roots cannot appear (cannot be pronounced or interpreted) without being categorized; they are categorized by merging syntactically with category-defining functional heads. … Concerning the functional heads themselves, we assume that there exist different types of n, v, and so on. (Embick & Marantz 2008: 6)

defined sort-type in the latter but not in the former.

[6] I display the feature set as an attribute-value matrix for better visual effect.

[7] There may be more complex categorization scenarios; e.g., in the defective category schema in §3.1.1, a single non-Root category X may simultaneously categorize two roots via the mediation of Cat.

[8] The schema is very similar to the head part of the defective category schema.

[9] e.g., Borer (2013) does not use DM-style categorizers but lets roots be directly categorized by the functional contexts they are embedded in. However, since the purpose of such contextual categorization is still to put the root into a traditional lexical category, Borer's system is more different from DM in implementation (i.e., how categorization is done) than in essence (i.e., the *category* in categorization refers to a traditional lexical category).

> It is not just any functional morpheme that can categorize a Root. There is a special set of morphemes in the grammar that perform this function; they are sometimes called *categorizers*. (Embick 2015: 45–46)

I endorse the first half of the categorization assumption (i.e., that roots must be categorized) but object to its second half, which basically says that only traditional lexical categories (incarnated in DM as little *x*s) may be categorizers. This is not a logical sequitur but a stipulation based on an idiomatic usage of the term *category*—as a synonym of *traditional lexical category*. This usage is common in generative syntax; for instance, *categorial features* conventionally mean lexical categorial features (so [N], [V] are categorial while [D], [T] are not). But if only the traditional lexical categories qualify as *categories* and the adjective *categorial* only applies to them, then how should we call functional categories (D, T, etc.) and their defining features (whatever those are)?[10] They would end up being "noncategories" and "noncategorial," which leads us to the awkward if not nonsensical conclusion that functional categories are not categories. On the other hand, if the purpose of root categorization is to give an acategorial root a category and thereby license its syntactic presence, then why must we confine this category to a traditional lexical category? What is worse, since in DM even the little *x*s are functional in nature, the confinement cannot be deduced from a lexical/functional division but must be listed as a separate axiom.

The above confusion reflects a stereotyped thinking pattern regarding categorization: the initial motivation for the root proposal was indeed from the lexical domain (Halle & Marantz 1993, 1994), but after that the marriage of roots and traditional lexical categories has simply stayed as a received norm, as evidenced by the many studies focusing on roots, little *x*s, and their local interactions (especially on phenomena like allomorphy and allosemy; see Matushansky & Marantz 2013 to get an impression). I do not think the results of previous studies or the groundbreaking status of DM would be undermined if we remove the second half of the categorization assumption, because logically the DM machinery merely requires that little *x*s *be* categorizers but not that *only* they be categorizers. That is, the desirable formal relation is (9a) instead of (9b).

(9) a. traditional lexical category(c) $\Rightarrow$ categorizer(c)
"c qualifies as a categorizer if it is a traditional lexical category"

b. traditional lexical category(c) $\Leftrightarrow$ categorizer(c)
"c qualifies as a categorizer iff it is a traditional lexical category"

I claim that the additional assumption of necessity in (9b) is a data-induced bias: previous data on roots are exclusively from the lexical domain, which probably has more to do with the fact that few have looked beyond the lexical domain than with any objectively proven interlocking relationship between roots and little *x*s in human language. Against this backdrop, an important aim of this chapter is to dissolve the marriage between roots and traditional lexical categories

[10] The confusion is even more serious when we compare different minimalist theories; e.g., in cartographically oriented works like Adger (2010) the domain of categorial features does include [D], [T], etc.

and thereby reduce the categorization assumption to its simpler form, which I call the *generalized categorization assumption*, as in (10).

(10) Generalized categorization assumption
Roots (or root-like chunks) cannot appear without being categorized; they are categorized by merging syntactically with category-defining functional heads. A category-defining functional head (or categorizer) is any head with a nonempty syntactic feature specification (i.e., any non-Root category).

The first clause in (10) is almost identical to Embick and Marantz's formulation (except for the adjustment regarding root-like chunks), while the second clause (underlined) is my innovation; it basically describes the schema in Figure 4.1, which can thus be called the *generalized root categorization schema*. I call my theory of root-based derivation *generalized root syntax* (GRS).

As aforementioned, the generalized root categorization schema is very tolerant. In the simplest scenario, X is just a traditional lexical category and ω a root, and the categorization result is a bare lexical category item as in DM. Beyond this scenario, X may also be any other non-Root category and ω a prederived chunk. Exploring the nature of non-traditional-lexical-category-based categorization is another aim of this chapter. In a nutshell, what we obtain from such categorization are half-functional–half-lexical objects, which I call *semifunctional items*.[11] The syntactic center of a semifunctional item is still X, while ω merely supports X at the interfaces, giving X a stand-alone phonological form (i.e., a free-morpheme status) and an encyclopedic reading. The same is true when X is a traditional lexical category; for example, in a bare noun like *dog*, the root $\sqrt{\text{DOG}}$ interpretively supports the abstract category N, which in itself may be null at PF and merely denote an `entity` type at LF (see §3.2.2). In sum, we can unify traditional lexical category items and semifunctional items under the umbrella term *root-supported head*.

There are two further categorization scenarios depending on the nature of X. First, X may be a prederived mergeme-bearing object; this happens when a root supports a span of abstract categories, so GRS is compatible with spanning. Second, X may also be a logical operator, which is allowed because the syntactic feature specification of a logical operator, albeit [−MERGEME], is nonempty.[12] A prediction of GRS, therefore, is that root-supported categorial spans and logical operators should manifest lexical idiosyncrasies as other root-supported heads do. My case study in §4.2 touches on these more complex scenarios, but I will generally leave them to future research

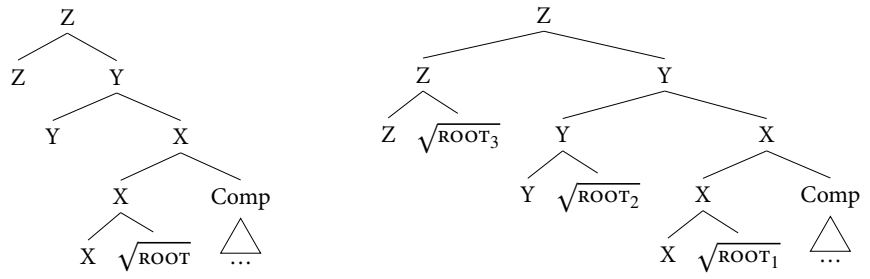
[11] Similarly, Svenonius (2014) doubts "whether there is any difference between lexical and functional items after all."

[12] This has an implication for labeling too. While logical operators cannot label phrases (à la Chomsky 2013), they *can* label root-supported heads; e.g., when a conjunction is supported by a root, as in Chinese {AND, $\sqrt{\text{HÉ}}$} (see §4.2.4), what labels the SO is AND rather than the root. As such, we could potentially distinguish different (cascading) degrees of "labeling capacity" for different types of syntactic category, with mergeme-bearing categories having full labeling capacity (i.e., they can label any SOs), logical operators having restricted labeling capacity (i.e., they can only label root-supported heads), and roots having no labeling capacity (i.e., they cannot label SOs at all). In other words, a syntactic category's labeling capacity is determined by the type of information in its feature specification. This, together with the remark in note 57, p. 27 about the label/name distinction, suggests that the term *label* in the labeling algorithm, as well as the granularity level the algorithm applies, needs more careful characterization.

and merely concentrate on the simplest scenario (i.e., atomic X + atomic ω) in this chapter.

4.1.2 Layered derivation

The output of the generalized root categorization schema is a categorized root, but how can we put it in derivational use? This is a question for all root-based theories. In both DM and XS categorized roots qua traditional lexical category items can take complements, as in Figure 4.2a, where X is a lexical category and Y, Z are functional categories. My system simply generalizes this situation to every head in a projection line—at every point of head-complement merger on the spine the head may be derivationally complex, as in Figure 4.2b.



(a) Previous systems **(b)** My system

Figure 4.2 Root-containing configurations

The complex heads in both trees in Figure 4.2 result from external Merge. So, the question of how to merge a root-supported head is essentially the same that of how to merge a complex satellite. And inasmuch as the merger of complex satellites is not a problem, that of root-supported heads is not problematic either. For concreteness, I adopt the *layered derivation* technique proposed by Zwart (2007, 2009, 2011). Other more or less similar techniques include label-based "bookmarking" (Nunes & Uriagereka 2000, Fowlie 2013), renumeration (Johnson 2003), post-syntactic root insertion (De Belder & Van Craenenbroeck 2015), and so on. I adopt Zwart's proposal because it can straightforwardly fit into a proof-theoretic conception of derivation.[13]

The gist of Zwart's proposal is that complex noncomplements (e.g., subjects) are constructed in separate *derivational layers* before they merge onto the derivational spine. Each derivational layer is defined by a separate numeration (or LA), and the output of one layer may be included as an atom into the numeration of another layer. For example, the derivation of the sentence in (11a) involves two derivational layers, the numerations for which are given in (11b) and (11c).

(11) a. The man kicked the ball.

b. $\text{numeration}_{\text{layer}_1}$ = {the, man}

[13] Several of the above-cited authors (including Zwart) use unary Merge. I stick to binary Merge but acknowledge that the arity of Merge is an issue of ongoing debate. See Collins & Stabler (2016) for a strong argumentation for binary Merge and against unary- or self-Merge, and see Collins (2017) for a softened position.

c. $\text{numeration}_{\text{layer}_2}$ = {[the man], kicked, the, ball} (Zwart 2011: 48)

Apart from complex subjects, other cases of layered derivation include simple and complex words (12a), coordination (12b), gapping (12c), and grafting (12d). Constituents from separate derivational layers are displayed in square brackets.

(12) a. [v-$\sqrt{\text{DANCE}}$], [book-shops], [jack-of-all-trades]

b. [I saw John the other day] [and Bill].

c. [John kissed Mary], [and Bill Sue].

d. John left for [I think Budapest]. (Zwart 2007)

According to Zwart (2009: 173), separately derived elements "have a dual nature," for "they are complex in the sense that they have been derived in a previous derivation [but] single items in that they are listed as atoms in the numeration for a subsequent derivation." This perfectly matches the nature of the prederived objects I used in Chapter 3 (e.g., the modifier materials in MHCs), hence my conception there that those are constructed in separate layers. In Chapter 3 I characterized prederived objects as proof-theoretic lemmas; the same characterization applies to Zwart's theory. Thus, the syntactic tree in Figure 4.2b can be converted to the proof tree in Figure 4.3 (suppose X, Y, and Z are simple mergeme-bearing categories).

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\overline{Z}^{\,\text{axiom}} \quad \dfrac{}{\sqrt{\text{ROOT}_2}}\,\text{axiom}}{\{Z, \sqrt{\text{ROOT}_3}\}}\,\text{merge}}{{}_Z\{Z, \sqrt{\text{ROOT}_3}\}}\,\text{label} \quad \dfrac{\dfrac{\dfrac{\overline{Y}^{\,\text{axiom}} \quad \dfrac{}{\sqrt{\text{ROOT}_2}}\,\text{axiom}}{\{Y, \sqrt{\text{ROOT}_2}\}}\,\text{merge}}{{}_Y\{Y, \sqrt{\text{ROOT}_2}\}}\,\text{label} \quad \dfrac{\dfrac{\dfrac{\dfrac{\overline{X}^{\,\text{axiom}} \quad \dfrac{}{\sqrt{\text{ROOT}_1}}\,\text{axiom}}{\{X, \sqrt{\text{ROOT}_1}\}}\,\text{merge}}{{}_X\{X, \sqrt{\text{ROOT}_1}\}}\,\text{label} \quad \text{Comp}}{\{{}_X\{X, \sqrt{\text{ROOT}_1}\}, \text{Comp}\}}\,\text{merge}}{{}_X\{{}_X\{X, \sqrt{\text{ROOT}_1}\}, \text{Comp}\}}\,\text{label}}{\{{}_Y\{Y, \sqrt{\text{ROOT}_2}\}, {}_X\{{}_X\{X, \sqrt{\text{ROOT}_1}\}, \text{Comp}\}\}}\,\text{merge}}{{}_Y\{{}_Y\{Y, \sqrt{\text{ROOT}_2}\}, {}_X\{{}_X\{X, \sqrt{\text{ROOT}_1}\}, \text{Comp}\}\}}\,\text{label}}{\{{}_Z\{Z, \sqrt{\text{ROOT}_3}\}, {}_Y\{{}_Y\{Y, \sqrt{\text{ROOT}_2}\}, {}_X\{{}_X\{X, \sqrt{\text{ROOT}_1}\}, \text{Comp}\}\}\}}\,\text{merge}}{{}_Z\{{}_Z\{Z, \sqrt{\text{ROOT}_3}\}, {}_Y\{{}_Y\{Y, \sqrt{\text{ROOT}_2}\}, {}_X\{{}_X\{X, \sqrt{\text{ROOT}_1}\}, \text{Comp}\}\}\}}\,\text{label}
$$

Figure 4.3 Root-supported heads in proof tree format (cf. Figure 4.2b)

In Figure 4.3 the derivational spine (i.e., the main derivational layer) is in black (the derivational detail of Comp is omitted), while the prederived lemmas together with their separate-layer derivations are in blue. Note that only the conclusion (i.e., final line) in each blue chunk is relevant for the spine, hence their head status in the main derivational layer. The proof tree clearly shows that both input slots in a Merge step may be occupied by derivationally complex objects.

4.1.3 Implications

GRS offers a new angle to view two general issues in human language: vocabulary item typology and (PF-oriented) head movement.

4.1.3.1 Vocabulary item typology

GRS treats semifunctional items as an independent type of vocabulary item[14] alongside the conventional lexical/functional division. If we use two boolean features [±LEXICAL] ("having a root") and [±FUNCTIONAL] ("having a non-big-X, nonempty syntactic category") to cross-classify vocabulary items, then semifunctional items are [+LEXICAL, +FUNCTIONAL] as in Table 4.1.

Table 4.1 Cross-classification of vocabulary items

	Lexical	**Functional**	**Semifunctional**
[LEXICAL]	+	−	+
[FUNCTIONAL]	−	+	+

Lexical and semifunctional vocabulary items, being rooted, both qualify as root-supported heads; by contrast, functional vocabulary items are rootless and merely abstract categories with overt exponents. See (13) for some examples in English.

(13) a. Lexical: *dog*, *run*, *good*

b. Functional: *the dog*, *plays*, *to run*

c. Semifunctional: *two dogs*, *I will stay*, *at home*

An important claim of GRS is that items like *two*, *will*, and *at* are rooted constructions; that is, they are prederived in separate layers before joining the main derivation as lemmas. As to the distinction between functional and semifunctional items, I use the criterion that semifunctional items show encyclopedic idiosyncrasies—which I take to be hallmarks of roots—whereas functional items do not. Thus, the infinitive *to* and the articles *the*/*a* are functional rather than semifunctional, even though they are free morphemes;[15] by comparison, quantifiers like *two*/*three*/*several*, auxiliaries like *will*/*have*/*can*, and prepositions like *at*/*in*/*on* do have idiosyncratic lexical semantics, as reflected in the obvious difference between *two* and *three*, the subtle difference between *I will stay* / *I shall stay*, and the choices of prepositions in *at home* / *in school* / *on the table*. Such idiosyncrasies are attributed to LIs in standard minimalism, but as we push the analytical granularity to the root level, their derivational source becomes an explican-

[14] I say "vocabulary item" instead of "lexical item" to avoid clumsy terms like "lexical/functional lexical item."
[15] So morphophonological status is not an efficient criterion to distinguish functional and semifunctional items.

dum; they also pose a challenge for DM, where roots (and hence idiosyncratic lexical semantics) are confined to traditional lexical categories. GRS provides a plausible answer to this "semifunctionality problem."

Let us examine the vocabulary items in (13) more closely. I illustrate their feature specifications in (14) following the template in (7).[16]

(14) a. *dog* ≔ [[PHON: /dɔg/], [SEM: 'a type of animal'], [SYN: ⟨CAT: N, POS: 1⟩], OP]

b. *the* ≔ [[PHON: /ðə/], [SYN: ⟨CAT: N, POS: 5⟩], OP]

c. *two* ≔ [[PHON: /tuː/], [SEM: 'one more than one'], [SYN: ⟨CAT: N, POS: 4⟩], OP]

Insofar as the examples in (14) are concerned, the three types of vocabulary item are very similar at PF, since they all have filled PHON features; moreover, their phonological values are all free-standing (i.e., nonaffixal). Thus, the only way to identify their types is to check their syntacticosemantic values, which indeed have different patterns, as summarized in (15).

(15)

Vocabulary item	**Syntacticosemantic characteristic**
dog	filled SEM + big X SYN
the	empty SEM + non-big-X SYN
two	filled SEM + non-big-X SYN

By hypothesis, if a phenomenon Φ only hinges on the phonological characteristics of the relevant vocabulary items, it would treat *dog*, *the*, and *two* alike; and if the typology in (15) does not affect Φ, then Φ is probably a PF phenomenon. I will discuss such a phenomenon in the next section.

4.1.3.2 PF-oriented head movement

Following the foregoing discussion, if a phenomenon is sensitive to the phonological nature of vocabulary items but insensitive to their semantic nature, then what we have is probably a PF phenomenon. To illustrate, consider three previous insights on head movement:

- Rizzi & Roberts (1989): I-to-C movement is impossible when C is filled by an *independent morpheme*; for example, in Québec French subject-clitic inversion is restricted to clauses with no overt C head, as in (16).

- Borer (2005a): A functional head realized by an abstract feature requires the movement of a lower head to phonologically support it, whereas a head realized by an *independent f-morph* (i.e., a free-morpheme formative) blocks such movement; for example, the plurality-encoding head is realized as an abstract feature in English but as an f-morph in Kraho (an

[16] The two POS values in (14b, c) are merely expository. For the current purpose it suffices to know that they respectively pick out the categories D and Q (at a certain granularity level).

Amazonian-Indian Creole), as in (17).

- Huang (2015): A null light verb[17] attracts the head of its complement to move up, whereas an *overt light verb* blocks such head movement; for example, in (18) the nominal head is incorporated into the light verb in English but stays in situ in Chinese.

(16) a. [$_{CP}$ *Qui* [$_{C}$ *as*$_i$-] [$_{IP}$ *tu* [$_{I}$ t$_i$] [$_{VP}$ *vu*]]]? [Québec French]
who have you seen
'Who have you seen?'

b. [$_{CP}$ *Qui* [$_{C}$ *que*] [$_{IP}$ *tu* [$_{I}$ as] [$_{VP}$ *vu*]]]?
who that you have
'Who have you seen?'

c. *[$_{CP}$ *Qui* [$_{C}$ *qu'-as*$_i$-] [$_{IP}$ *tu* [$_{I}$ t$_i$] [$_{VP}$ *vu*]]]?
who that-have you seen
(adapted from Rizzi & Roberts 1989: 4)

(17) a. [*woman*$_i$-PL [$_{NP}$ t$_i$]] → *women* [English]

b. [*me* [$_{NP}$ *kakay*]] [Kraho]
PL woman
'women'

c. *[*kakay*$_i$*-me* [$_{NP}$ t$_i$]]
woman-PL
(adapted from Borer 2005a: 33)

(18) a. [$_{VP}$ [$_{V}$ *telephone*$_i$- DO] [$_{NP}$ t$_i$]] 'telephone (someone)' [English]

b. [$_{VP}$ [$_{V}$ *dǎ*] [$_{NP}$ *diànhuà*]] [Mandarin]
hit.DO telephone
'telephone (someone)'

c. *[$_{VP}$ [$_{V}$ *diànhuà*$_i$- *dǎ*] [$_{NP}$ t$_i$]]
telephone hit.DO
(adapted from Huang 2015: 11)

The three head movement phenomena above are all sensitive to the phonological nature of the potential host. Are they also sensitive to the host's syntacticosemantic properties? Let us check the three hosts—namely, French *que* 'that (C)', Kraho *me* 'PL', and Mandarin *dǎ* 'hit (LV)'—against Table 4.1. I summarize the results in (19).

[17] By a "light verb" Huang means a semantically light lexical verb like *take* in *take a seat*. While this is the original sense of *light verb* as it was first proposed in Jespersen (1965), nowadays it is often mixed with another more technical sense of the term; i.e., a VP shell in the sense of Larson (1988). I will return to this point in §4.2.

(19)

Host head	Syntacticosemantic characteristic	Type
que, *me*	no idiosyncratic lexical semantics, non-big-X category	Functional
dǎ	idiosyncratic lexical semantics, non-big-X category	Semifunctional

Since lexical vocabulary items (i.e., lexical nouns, verbs, and adjectives) are never null, contrasts like those in (16)–(18) are irrelevant to them even if they can host head movement;[18] therefore, we can take the summary in (19) to be comprehensive, which shows that the head movement phenomena in question indeed disregard the vocabulary item typology, so they are PF phenomena by our earlier hypothesis.[19]

An even more interesting contrast emerges if we compare the blocked head movement above and the failed head movement in §3.2.3.4 (West Germanic immobile verbs). In the phenomena here, head movement is not needed when the host is filled by a free morpheme, be it functional or semifunctional, whereas in the case of immobile verbs, head movement is needed since the host is empty. The need just cannot be fulfilled for syntactic reasons. In short, in the phenomena here head movement is not blocked but simply unmotivated, whereas with immobile verbs head movement is truly blocked. The upshot is that unmotivated and blocked head movements have different behaviors.

4.2 Semifunctional items in Chinese

In the previous section I introduced the GRS theory and discussed its implications. In this section I present a case study, that of Mandarin Chinese semifunctional items, which has a natural analysis under GRS and brings along two further implications (concerning analyticity and grammaticalization). I begin with the verbal/sentential domain (§4.2.1) and then move on to the nominal domain (§4.2.2), the prepositional domain (§4.2.3), and the acategorial domain (§4.2.4). I briefly summarize the data in §4.2.5.

4.2.1 Verbal/sentential domain

In this subsection I present Mandarin semifunctional items from the verbal/sentential domain, more specifically from five grammatical categories: "light verb"[20] (§4.2.1.1), voice (§4.2.1.2), as-

[18] A possibly relevant phenomenon is N-to-V incorporation (see Baker 1988).

[19] I am not claiming that all head movements are PF phenomena. Head movement may well be a nonhomogeneous category, and we should research it case by case.

[20] I use quotation marks because I do not think such "light verbs" are really a separate syntactic category from that of nonlight verbs. The term *light verb* have two senses in the literature. The first sense applies to the items introduced in §4.2.1.1, which appear in the so-called light verb constructions; i.e., verb-object phrases like *take a shower*, *have a nap*, etc. Such light verbs are merely semantically light; syntactically they are ordinary verbs. On the other hand, *light verbs* in the second sense—i.e., items overtly realizing Larsonian VP shells—are syntactically different from nonlight verbs. I will suggest in §4.2.1.2 and §4.2.3 that Chinese voice elements and prepositions are light verbs

pect (§4.2.1.3), modal (§4.2.1.4), and sentence-final particle (§4.2.1.5).

4.2.1.1 "Light verbs"

In §4.1.3.2 we have already seen a light verb *dǎ*, which literally means 'hit, beat' and is frequently used as a placeholder for the activity predicate DO. See Table 4.2 for more items like *dǎ* in Mandarin, all of which are in productive use.

Table 4.2 Overt DO light verbs in Mandarin

Light verb	Literal meaning	Register	Example
dǎ	'hit, beat'	neutral	*dǎ diànhuà* 'hit telephone; make a phone call'
zuò	'do, make'		*zuò yóuxì* 'do game; play (nonelectronic) games'
bàn	'do, handle'		*bàn gōngchǎng* 'do factory; run a factory'
gǎo	'do, make, get'	casual	*gǎo weìshēng* 'do cleaning'
nòng	'fetch, fiddle'		*nòng chīde* 'fiddle food; prepare food'
zhěng[21]	'tidy, fix'		*zhěng diǎnr jiǔ* 'fix some alcohol; get some alcohol'
gàn	'do, engage in'		*gàn shìyè* 'do career; build a career'
jìnxíng	'conduct, carry out'	formal	*jìnxíng cǎifǎng* 'conduct interview'
cóngshì	'engage in'		*cóngshì xuéshù gōngzuò* 'engage in academic work'
jǐyǔ	'give, offer'		*jǐyǔ kǎolǜ* 'give consideration; consider'
jiāyǐ	'impose on'		*jiāyǐ fúchí* 'impose aid; aid'

The items in Table 4.2 show four typical characteristics of semifunctionality. First, they all have lexical origin in history and synchronically can still be used as lexical verbs, as in (20).

(20) *dǎ wénzi* 'beat mosquito'
zuò wǎnfàn 'make dinner'
zhěng yá 'fix teeth; go to the orthodontics'

Actually we can fairly classify the light verbs above as main verbs and treat the light verb constructions in Table 4.2 as V-O idioms. This is evidenced by the fact that they can all be separated by grammatical elements just like ordinary V-O phrases, as in (21).[22]

in this second sense, the former realizing Voice and the latter various applicative categories.

[21] This item originated in Northeastern Mandarin but quickly entered Common Mandarin (Putonghua).

[22] See Butt (2010: 75) for a similar argument: "The idea that light verbs and their corresponding main verbs are derived from one and the same underlying representation accounts for the fact that light verbs are always form identical to a main verb counterpart in the language and that they are stable with respect to historical change."

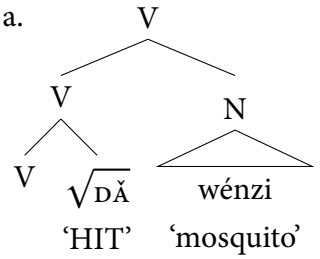
(21) a. *dǎ-le liǎng-tōng diànhuà*
hit-PERF two-CL telephone
'made two phone calls'

b. *gǎo-le yí-xià jiàoshì wèishēng*
do-PERF one-CL classroom cleaning
'had a cleaning of the classroom'

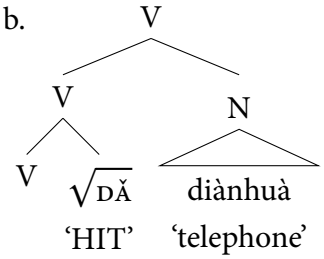
c. *jìnxíng-guò duō-cì cǎifǎng*
conduct-EXP many-CL interview
'have (ever) done many interviews'

Such a main-verb analysis for Chinese light verbs is in line with GRS, whereby overt light verbs are derived in the same way as nonlight verbs, as in (22).

(22) a.



b.



The difference between *dǎ wénzi* and *dǎ diànhuà* is just that the former is compositional whereas the latter is idiomatic. Crucially, *dǎ* in *dǎ diànhuà* is not an exponent of the abstract category V but a rooted item, for it semantically selects its nominal object as in (23a) and lexically restricts the broad event expressed by the construction (e.g., gaming, fishing) as in (23b–c). Such lexical-semantic conditioning is the second typical characteristic of semifunctionality.[23] It is also typical of lexicality, of course, and hence can be considered a general hallmark of rootedness.

(23) a. *dǎ/*gǎo diànhuà* 'make a phone call'
**dǎ/gǎo wèishēng* 'do cleaning'

b. *zuò/*dǎ yóuxì* 'play (nonelectronic) games'
**zuò/dǎ yóuxì* 'play (electronic) games'

c. *dǎ yú* 'catch fish (in batch, from sea or river)'
gǎo yú 'get fish (quantity/source unspecified; e.g., one fish from the market)'

A third typical characteristic of semifunctionality reflected in Chinese light verbs is stylistically conditioned variation. When multiple light verbs are acceptable for a particular expression (which happens a lot), the choice of light verb often hinges on the register. See (24) for examples.

[23] Butt (2010: 48) makes a similar remark about English: "[Light] verbs are clearly not entirely devoid of semantic predicative content either: there is a clear difference between *take a bath* and *give a bath*."

(24) a. *zuò tóufà* 'make hair; go to the hair salon (colloquial, neutral)'
nòng tóufà 'fiddle hair; go to the hair salon (casual) '

b. *jìnxíng cǎifǎng* 'conduct interview (formal)'
zuò cǎifǎng 'do interview (neutral)'
gǎo cǎifǎng 'do interview (casual)'

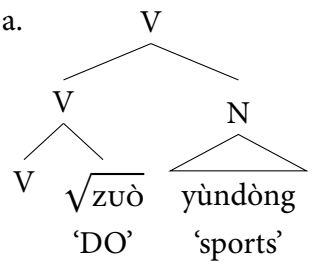
So, Chinese light verbs are quite different from the general-purpose light verbs in languages like Japanese and Korean. This crosslinguistic difference is reflected in errors made by Korean students learning Chinese, as in (25).

(25) a. **zuò xuéxí* 'do study' (correct form: *xuéxí* 'study')
cf. Korean *gongbu-ha-da* 'study-DO-TAM'

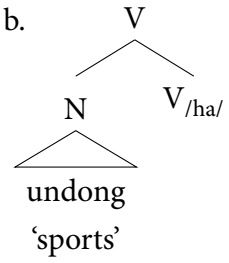
b. **dǎ yùndòng* 'hit sports' (correct form: *zuò yùndòng* 'do sports')
cf. Korean *undong-ha-da* 'sports-DO-TAM'. (Pak 2014)

Since there is only one DO light verb in Korean, no lexically conditioned selection occurs. Therefore, Korean *ha* may be a true exponent of the abstract category V (i.e., the verbalizer). I illustrate this difference between Chinese and Korean in (26).[24]

(26) a.



b.



Considering the syntactic similarity of light verbs with ordinary lexical verbs and their semantic similarity with verbalizer exponents, we can view them as lexical verbs in the process of becoming verbalizer exponents.

4.2.1.2 Voice elements

Table 4.3 are some commonly used voice auxiliaries[25] in Mandarin.

[24] I do not think head direction is encoded in Merge; the different V-N ordering in (26) is merely expository.
[25] Here and below I reserve the term *auxiliary* for preverbal free-morpheme elements.

Table 4.3 Voice auxiliaries in Mandarin

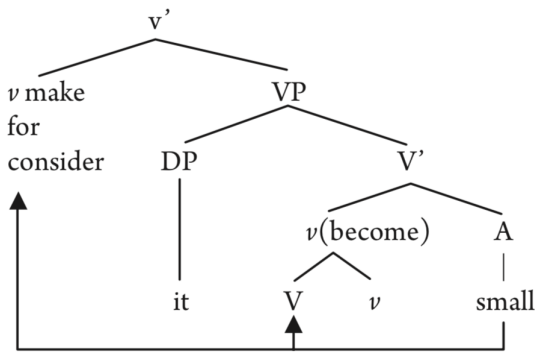
Auxiliary	Literal meaning	Voice	Example
shǐ	'use, employ'	causative	*shǐ rén chéngzhǎng* 'cause sb. to grow up'
ràng	'let, give'		*ràng rén gǎndòng* 'cause sb. to feel touched'
lìng	'order'		*lìng rén xiànmù* 'cause sb. to envy'
jiào	'call'		*jiào rén huáiniàn* 'cause sb. to be nostalgic'
bèi	'cover, suffer'	passive	*bèi rén zhuā-le* 'by sb. catch-PERF; be caught by sb.'
ràng	'let, give'		*ràng rén mǎi-le* 'by sb. buy-PERF; be bought by sb.'
gěi	'give'		*gěi rén piàn-le* 'by sb. deceive-PERF; be deceived by sb.'
jiào	'call'		*jiào rén dǎ-le* 'by sb. beat-PERF; be beaten by sb.'

These items also show the characteristics of semifunctionality mentioned in §4.2.1.1.[26] First, they are all lexical in origin and can still be used as lexical verbs, as in (27).

(27) *shǐ huàféi* 'use chemical fertilizer'
ràng lù 'give way; make way for sb./sth.'
bèi nàn 'suffer from a disaster'
jiào jǐngchá 'call the police'

Since the above voice auxiliaries are verbal in origin and occupy a VP-shell position, they qualify as light verbs in the second sense in note 20, p. 109. In literature on Chinese, *light verb* is used in both senses, with the first sense typically used for Mandarin light verb constructions and the second for Old Chinese empty VP-shell categories and their overt counterparts in Mandarin (which look just like voice auxiliaries or prepositions). The tree below is from Feng (2015: 243).

(28)



Feng calls both the lower *v* and the higher *v* "light verbs." In line with note 20, p. 109 I call the lower *v* a verbalizer and the higher *v* a VP shell. Observe that among the three readings Feng lists for the higher *v*, *make* is a (causative) voice auxiliary and *for* a preposition.

[26] Some voice auxiliaries (e.g., the passive ones in Table 4.3) are most naturally translated into English as prepositions. This is not a coincidence but reflects a deeper connection between VP-shell items and prepositions in Chinese.

Second, where a single voice can be expressed by several vocabulary items or a single vocabulary item can express several voices, there is lexically or stylistically conditioned variation in the choice of auxiliary, as in (29)–(30).

(29) a. *qiānxū shǐ/ràng/?lìng/?jiào rén jìnbù*
modesty CAUS people progress
'Modesty helps one to make progress. (formal, slogan)'

b. *zhè shìr zhēn ?shǐ/ràng/lìng/jiào wǒmen yìwài*
this matter so CAUS we feel unexpected
'This matter makes us feel so unexpected! (colloquial; e.g., in an office chat)'

(30) a. *fànrén zuótiān bèi/?ràng/??gěi/?jiào jǐngchá dàibǔ-le*
criminal yesterday PASS police arrest-PERF
'The criminal was arrested by the police yesterday. (formal; e.g., in news)'

b. *nǐ-de jiā-qiǎor bèi/?ràng/gěi/??jiào māo chī-le*
you-POSS house-sparrow PASS cat eat-PERF
'Your sparrow was eaten by a cat! (colloquial; e.g., between neighbors)'

In the above examples some ?-marked auxiliaries are less acceptable as a result of s-selectional restrictions (e.g., *lìng* and *jiào* do not go well with 'progress') while others are so as a result of stylistic infelicity (e.g., *ràng*, *jiào*, and especially *gěi* are too colloquial for news reports). So, the distribution of voice auxiliaries in Chinese is determined not only by their grammatical (i.e., voice-marking) functions but also by their encyclopedic information.

In addition, Mandarin also has a derivational suffix *huà* 'lit. change, transform', which creates causative/inchoative verbs from adjectives or nouns as in (31).

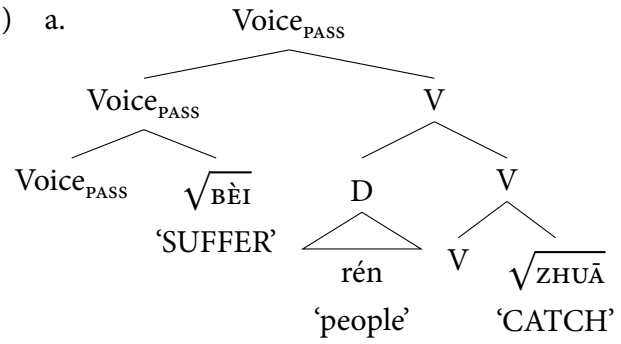
(31) *měi-huà* 'beautiful-CAUS/INCH; beautify'
lǜ-huà 'green-CAUS/INCH; make/turn green (by planting trees)'
gōngyè-huà 'industry-CAUS/INCH; industrialize'
chéngshì-huà 'city-CAUS/INCH; urbanize'

Being a suffix, *huà* is more inflectional than the auxiliaries in Table 4.3; in DM it may be analyzed as an exponent of the "flavored" verbalizer $v_{\text{CAUS/INCH}}$ (in the sense of Harley 1995 et seq.). Yet *huà* can also be used as a lexical verb, as in (32).

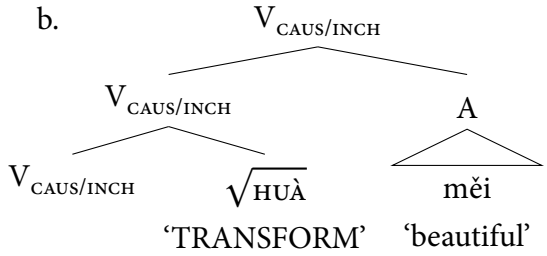
(32) *bīng huà-le* 'ice melt-PERF; the ice has melted'
huà-chéng huī 'transform-become ash; turn into ash'

Also, albeit highly productive, *huà* cannot be equally felicitously attached to all adjectives/nouns. For example, one can say *lǜ-huà* 'make green' but not ?*dà-huà* 'large-CAUS/INCH; enlarge', though *jùdà-huà* 'huge-CAUS/INCH; make/become huge' is fine; similarly, one can say *shī-huà* 'poem-CAUS/INCH; turn (sth.) into poem' but not ?*gē-huà* 'song-CAUS/INCH; turn (sth.) into song', though *gēqǔ-huà* 'song-CAUS/INCH; turn (sth.) into song' is fine. Such contrasts are hard to explain from a grammatical angle; they are more likely due to the lexical idiosyncrasies in the relevant roots instead. Since derivational affixes like *huà* are not completely bleached, Lü (1979: 48) proposes a separate class "quasi affix" (*lèicízhuì* 類詞綴) for them and comments that the existence of quasi affixes (i.e., affixes that have independent lexical uses) is a typical characteristic of Chinese morphology. Under GRS quasi affixes can be analyzed as root-supported (traditional lexical) categorizers. I illustrate the derivation of voice auxiliaries and quasi affixes in (33).[27]

(33) a.



b.



A question here is how to distinguish root-supported categorizers and categorized roots; for instance, how does the C-I interface know the {$V_{CAUS/INCH}$, $\sqrt{HUÀ}$} unit in (33b) is a semifunctional item instead of a lexical item? I do not have a final answer yet but have two speculations:

1. Perhaps the interface cannot make a decision for the unit alone but must reference its ambient configuration (e.g., whether there are arguments associated with it, whether its complement is a bare lexical category). This is plausible because out of context speakers cannot decide whether *huà* is a verb or a verbalizer.

2. Perhaps $V_{CAUS/INCH}$ is not an atomic category but a Voice-V span, which categorizes $\sqrt{HUÀ}$ into a "semi-semifunctional item"; by comparison, the lexical verb *huà* is categorized by V alone (and if someday $\sqrt{HUÀ}$ develops a purely semifunctional usage it would be categorized by Voice alone). This might be appealing because it maintains a one-to-one correspondence between structure and meaning and reflects the fact that *huà* as a quasi affix is in a preliminary stage of grammaticalization.

[27] I set aside linearization issues and assume that the verbalizer *huà* becomes suffixal by some independent procedure.

4.2.1.3 Aspectual elements

Aspectual elements in Chinese can be either preverbal or postverbal, as in Table 4.4.

Table 4.4 Aspectual elements in Mandarin

(a) Preverbal

Element	Literal meaning	Aspect
yǒu	'have'	perfective
zài	'be at, be present'	progressive

(b) Postverbal

Element	Literal meaning	Aspect
le	'(<*liǎo*) end'	perfective
zhe	'(<*zhuó*) attach, touch'	stative/imperfective
guò	'pass'	experiential
wán	'finish'	completive
hǎo	'good'	completive
diào	'drop'	completive
shàng	'ascend'	completive/inchoative
qǐ(-lái)	'rise(-come)'	inchoative
kāi	'be open'	inchoative

As Table 4.4 shows, postverbal aspect is the predominant pattern in Chinese, presumably due to the frequent grammaticalization of resultative complements since MidC (see Shi 2003). Below are some examples of Mandarin aspectual elements. A sentence may have a preverbal aspectual element (34), a postverbal one (35), or both (36).

(34) a. *wǒ yǒu chī-fàn* (preverbal)
I have eat-meal
'I have eaten.'

b. *wǒ zài chī-fàn*
I be at eat-meal
'I'm eating.'

(35) a. *tā huì shā-le nǐ de* (postverbal)
he will kill-PERF you SFP
'He'll kill you!'

b. *wǒmen shàng-zhe kè ne*
we be.engaged.in-PROG class SFP
'Hey, we're in a lecture (so stop chatting)!'

(36) a. *wǒ yǒu qù-guò běijīng* (pre- and postverbal)
I have go-EXP Beijing
'I have been to Beijing before.'

b. *shìjiè zài gǎibiàn-zhe wǒmen*
world be at change-PROG us
'The world is changing us.'

There are still more aspectual elements in Mandarin, but the examples in Table 4.4 sufficiently reveal that aspectual elements show the same semifunctional characteristics as light verbs and voice elements. First, they all have lexical origins and can still be used lexically, as in (37).

(37) a. *tā yǒu hěn duō qián*
he have very much money
'He has a lot of money.'

b. *fēijī zhuó-lù le*
plane touch-land PERF
'The plane has landed.'

c. *tā xiǎng shàng diànshì*
he want ascend TV
'He wants to be on TV.'

Second, where one vocabulary item can express several aspects or one aspect can be expressed by several vocabulary items, the variation is conditioned by lexical semantics or register, as in (38).

(38) a. *wǒ zuòyè xiě-wán/hǎo/??diào le*
I homework write-COMP SFP
'I have finished my homework. (neutral~colloquial; e.g., child to parent)'

b. *fàng-xué hòu xiān bǎ zuòyè xiě-wán/?hǎo/diào*
dismiss-school after first DISP homework write-COMP
'After school, first finish off your homework. (colloquial; e.g., parent to child)'

c. *fàn zuò-?wán/hǎo/??diào le*
meal make-COMP SFP
'The meal is ready. (colloquial, positive; e.g., wife to husband)'

d. *wǒ-de bīngqílín bèi māma chī-ʔwán/ʔʔhǎo/diào le*
I-POSS ice cream PASSIVE mom eat-COMP SFP
'My ice cream was eaten by mom. (neutral~colloquial, negative; e.g., child to father)'

To anticipate the content in Chapter 5, I assume two separate grammaticalization routes for aspectual elements in Chinese: preverbal ones are grammaticalized from a generic verbal category (V) and postverbal ones from a stative verbal category (V_{BE}). There may accordingly be two aspect categories as well, a generic one (Asp) and a stative one (Asp_{BE}).[28] The two classes of aspectual elements can thus be respectively derived as follows.

(39) a. [Asp [Asp $\sqrt{\text{YǑU}}$]] 'HAVE' b. [Asp_{BE} [Asp_{BE} $\sqrt{\text{GUÒ}}$]] 'PASS'

The constructions in (39) yield semifunctional items; Asp/Asp_{BE} corresponds to X in Figure 4.1. The double-route grammaticalization hypothesis for Chinese aspectual elements makes several predictions; for instance, it predicts that the three grammatical aspect suffixes *le* 'PERF', *zhe* 'PROG', and *guò* 'EXP' are lower than their usually assumed position (i.e., in *v*P instead of IP)[29] and that the CRS *le* may also have a position lower than usually assumed (i.e., around *v* instead of C). I will elaborate on the two grammaticalization routes in Chapter 5.

4.2.1.4 Modal elements

Modal elements in Chinese are also either preverbal or postverbal, as in (40); though unlike aspectual elements, they are predominantly preverbal. The only salient postverbal modal element in Standard Mandarin is *de*, which is grammaticalized from *dé* 'get, obtain'.[30]

(40) a. *wǒ néng líkāi ma* (preverbal)
I can leave SFP
'Can I leave?'

b. *zhè shìr zuò-de* (postverbal)
this matter do-MOD
'This matter is doable.'

[28] I will promote a slightly different view in Chapter 5, but this possibility is theoretically legitimate. See Song (2018b) for a relevant discussion of Asp_{BE}.

[29] What is relevant here is not a distinction between grammatical and lexical (or viewpoint and situation) aspects, since all three aspect markers are viewpoint-based and fully productive. As we will see in Chapter 5, the distinction is one of extended projection lines instead.

[30] There is another postverbal modal element *liu* 'POTENTIAL' in Northern Mandarin; see Lamarre (1995, 2001) and Song (2016b) for discussions (see also Chapter 5).

c. *zhè-zhǒng shuǐguǒ yìbān-rén néng chī-de-qǐ* (pre- and postverbal)
this-CL fruit ordinary-people can eat-MOD-up
'Ordinary people can afford to eat this fruit.'

Table 4.5 contains examples of Mandarin modal auxiliaries (Chao 1968, Li & Thompson 1981).

Table 4.5 Modal auxiliaries in Mandarin

Auxiliary	Meaning	Modal type
néng(-gòu)	'able(-enough); can, be able to'	dynamic
huì	'understand; know (how to)'	
xiǎng/yào/xiǎng-yào	'desire-want; would like to'	
ài	'love; love to, tend to'	
yuàn(-yì)	'wish to, be willing to'	
kěn	'agree; agree to '	
xǐhuān	'like; would like to'	
lèyì	'glad; be glad to'	
gāoxìng	'happy; be happy to'	
gǎn	'dare'	
pà	'fear; be afraid to'	
néng(-gòu)	'able(-enough); can'	deontic
kě(-yǐ)	'approve; may, be permitted to'	
xǔ	'permit; be permitted to, may'	
zhǔn	'allow; be allowed to, may'	
(yīng-)gāi	'(should-)ought; should, ought to'	
(yīng/gāi-)dāng	'(should/ought-)should.properly; should, ought to'	
děi	'must'	
bì-xū	'necessary-need; must, have to'	
xū-yào	'need-want; need to, should'	
yòng	'use; need to, be required to'	
huì	'understand; could possibly, would probably'	epistemic
kěnéng	'may, possibly'	
(yīng-)gāi	'(should-)ought; should probably'	
(yīng/gāi-)dāng	'(should/ought-)should.properly; ought to, likely'	

As Table 4.5 shows, quite a few modal elements in Mandarin are rather like lexical verbs. Indeed, the grammatical category "modal auxiliary" is disputed among scholars; some prefer treating it as a subtype of lexical verb (e.g., Zhu 1982, Huang et al. 2009). The dispute suggests that Chinese modal elements are not purely functional but either semifunctional or fully lexical; they indeed also show the typical semifunctional/lexical characteristics. On the one hand, the

modal elements in Table 4.5 all have a lexical origin and can still be used lexically, as in (41).

(41) a. *tā huì zhōngwén*
he know Chinese
'He knows Chinese.'

b. *néng-zhě duō láo*
able-person more work
'Able people should do more work.'

c. *tā gāi wǒ qián*
he owe me money
'He owes me money.'

On the other hand, where a single vocabulary item can express multiple modal meanings or a single modal meaning can be expressed by multiple vocabulary items, we see lexical-semantically or stylistically conditioned variation, as in (42).

(42) a. *bù xǔ líkāi*
not may leave
'Don't leave! (softer)'

b. *bù zhǔn líkāi*
not may leave
'Don't leave! (stronger and may sound bossy)'

(43) a. *wǒ děi gàosu nǐ yí-jiàn shì*
I must tell you one-CL thing
'I must tell you something. (colloquial; may or may not be serious)'

b. *wǒ bìxū gàosu nǐ yí-jiàn shì*
I must tell you one-CL thing
'I must tell you something. (formal; usually serious)'

I derive the preverbal and postverbal modal elements as follows.

(44) a. [Mod [Mod √NÉNG]] 'ABLE'

b. [Mod_{BE} [Mod_{BE} √DÉ]] 'GET'

The double-route grammaticalization hypothesis for aspectual elements can also be extended to

modal elements (though again see Chapter 5 for a slightly different view).

4.2.1.5 Sentence-final particles

Finally, I briefly repeat the SFP-related content in §3.3. We have seen that Chinese SFPs are closer to a lexical category, since they form an open class and encode idiosyncratic (emotional) information. I proposed a category Emotion (E) for both auditory and visual SFPs, which directly merges with a root in the same way a big X category does, as in (45).

(45) a. $\begin{array}{c} \text{E} \\ \text{E} \quad \sqrt{\text{YA}} \end{array}$ b. $\begin{array}{c} \text{E} \\ \text{E} \quad \sqrt{\text{😳}} \end{array}$ (adapted from Chapter 3, (72))

In hindsight, this is just an application of the generalized root categorization schema, so SFPs are semifunctional items. They indeed also show the typical characteristics of semifunctionality. First, SFPs are recruited from independently available lexical materials in the language, either verbs/adverbials or interjections, which can still be used lexically, as in (46).

(46)

SFP	Lexical source	Example of lexical use
ba 'tentativity'	*bà* 罷 'stop, finish'	*bà-gōng* 'stop-work; go on strike'
ma 'yes/no'	*wú* 無 'not have' (+*a*)[31]	*wú-qù* 'not.have-interest; boring'
a 'general emotionality'	*a* 'ah'	*a, hǎo ba* 'ah good SFP; ah, okay'
ei 'attracting attention'	*ei* 'eh'	*ei, děng-děng* 'eh wait-wait; eh, wait'

In internet language visual SFPs also have lexical uses, either as interjections or as traditional lexical category items. This situation is not limited to Chinese. See (47) for some English examples.

(47)

Visual SFP	Lexical use	Example
😊 'jolly emotion'	interjection	—*Are you okay?* —😊
😮 'surprised emotion'	interjection	—*You won the lottery!* —😮
😸 'happy/cute emotion'	'cat (N)'	*I saw a* 😸.
👯 'friendly/pleased emotion'	'dance (V)'	*Let's go* 👯.

Under GRS these lexical uses have the same derivational path, as in (48).

(48) a. $\begin{array}{c} \text{E} \\ \text{E} \quad \sqrt{\text{😊}} \end{array}$ b. $\begin{array}{c} \text{E} \\ \text{E} \quad \sqrt{\text{😮}} \end{array}$ c. $\begin{array}{c} \text{N} \\ \text{N} \quad \sqrt{\text{😸}} \end{array}$ d. $\begin{array}{c} \text{V} \\ \text{V} \quad \sqrt{\text{👯}} \end{array}$

[31] The yes/no usage of *wú* is preserved in Min Chinese (pronounced as *bo*); e.g., *li bat i bo?* 'you know him or-not; do you know him?' (retrieved from `https://twblg.dict.edu.tw`). See Lian (2015: 177, note 15) for a comment that this particle is structurally lower than Mandarin *ma*, which indirectly supports the hypothesis that *ma* (among other Mandarin SFPs) is decomposable (see §3.3.1).

Second, the semantic contribution of SFPs is mainly encyclopedic rather than functional, as in (49)–(50). In (47) *a* and *ya* are phonological variants of the same general-emotionality SFP, but they have different emotional strengths; similarly, the three emojis in (48) give a single sentence three quite different readings.

(49) a. *nǐ hǎo a* (auditory SFP)
you good SFP
'Hello! (friendly but more reserved)'

b. *nǐ hǎo ya*
you good SFP
'Aloha! (friendly and enthusiastic)'

(50) a. *nǐ zhēn yǒu-cái* 🤗 (visual SFP)
you so have-talent SFP
'You are so talented! (sincere)'

b. *nǐ zhēn yǒu-cái* 🙄
you so have-talent SFP
'You are so "talented"! (sarcastic)'

c. *nǐ zhēn yǒu-cái* 💅
you so have-talent SFP
'Okay, okay, you are very talented. (nonchalant)'

Third, the use of SFPs is also subject to register conditioning. A well-observed fact about auditory SFPs is that they are largely restricted to informal, conversational registers. Compare the sentences in (51) for example.

(51) a. *nǐ qù-guò běijīng ma?*
you go-EXP Beijing SFP
'Have you ever been to Beijing? (colloquial; e.g., between colleagues)'

b. *nǐ kě céng qù-guò běijīng?*
you EMPH ever go-EXP Beijing
'Have you ever been to Beijing? (elegant; e.g., in literary works)'

The same is true for visual SFPs; for example, in an online conversation between school staff and student parents 🙂 is much more felicitous than 🙃, as in (52a), whereas in a chat between peer friends (especially young people) 🙂 is often considered insincere and hence less preferred than 🙃 or other alternatives (e.g., 😀, 😸), as in (52b).

(52) a. *gè-wèi jiāzhǎng, wǒmen míngtiān bā-diǎn jíhé* 🙂 / ??🙃
every-CL parent we tomorrow eight-o'clock congregate SFP SFP
'Dear parents, we'll meet at 8 tomorrow. (between teacher and parents)'

b. *lǎoshī shuō bā-diǎn jíhé* ?🙂 / 🙃
teacher say eight-o'clock congregate SFP SFP
'The teacher says meeting at 8. (between students)'

4.2.2 Nominal domain

In this subsection I present semifunctional items in Mandarin Chinese from the nominal domain, including two grammatical categories: classifier (§4.2.2.1) and pronoun (§4.2.2.2).

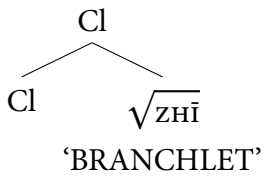
4.2.2.1 Classifiers

Classifiers in Chinese divide mass nouns by defining well-formed portions of designated shapes based on universally or culturally given semantic features (Borer 2005a, del Gobbo 2014). For example, *zhī* in (53a) divides the abstract concept 'pen' into quantifiable units of a long, thin shape (which a prototypical pen can fit in), and *duǒ* in (53b) divides the abstract concept 'flower' into quantifiable units of a prototypical flower shape.

(53) a. *yì-zhī bǐ*
one-CL pen
'a pen'

b. *liǎng-duǒ huā*
two-CL flower
'two flowers'

Classifiers are a semifunctional category par excellence: they are [+FUNCTIONAL] in having a syntactically significant mass-dividing functionality and [+LEXICAL] in defining idiosyncratic conceptual "molds." In GRS the former is encoded in some functional category (e.g., Cl) and the latter in various roots, with a classifier being a merger of the two, as in (54).

(54) a. 

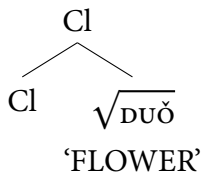
b. 

Table 4.6 presents a small subset of classifiers in Mandarin (see Chao 1968 for more). Since

some classifier subtypes (e.g., container) are open classes, classifier itself is an open class too.

Table 4.6 Classifiers in Mandarin

Subcategory	Classifier	Literal meaning	Example nouns
Individual	*jiàn*	'item, article'	clothing, matter, gift
	zhī$_1$	'single'	fowls, sheep, hand, foot
	zhī$_2$	'branchlet'	pen, candle, troop, music
	duǒ	'flower'	flower, cloud
	tóu	'head'	cattle, donkey, pig, garlic
	pǐ	'(cloth) bolt'	cloth, horse, wolf
	tiáo	'twig'	dog, snake, dragon, ribbonfish
	wěi	'tail'	goldfish, brocaded carp
	gè	'individual'	general purpose
Group	*qún*	'(animate) crowd'	sheep, people, cattle
	duì	'couple'	pigeons, lovers, earrings, twins
	shuāng	'pair'	shoes, socks, chopsticks, eyes
	pī	'batch'	goods, students, tourists, medicine
Partition	*fèn*	'portion, set'	meal, gift, newspaper, contract
	tuán	'lump, mass'	clay, dough, wool, loose hair
	dī	'drop'	water, rain, tear, blood
	xiē	'some'	general purpose
Container	*hé*	'(small, plain) box'	match, medicine, rouge, cigarette
	xiá	'(small, fine) box'	jewelry, treasure
	bāo	'package'	books, sugar, cigarette, crisps
	pén	'basin, tub, pot'	water, flour, fish, mutton

The three typical characteristics of semifunctionality are well manifested in classifiers. First, classifiers all have a lexical origin and can still be used lexically, as in (55).

(55) a. *jīntiān jì-le sān-gè jiàn*
today send-PERF three-CL item
'Today I sent three items (via courier).'

b. *tā xīn jiàn-le yí-gè qún*
he newly create-PERF one-CL group
'He has newly created a (chatting) group.'

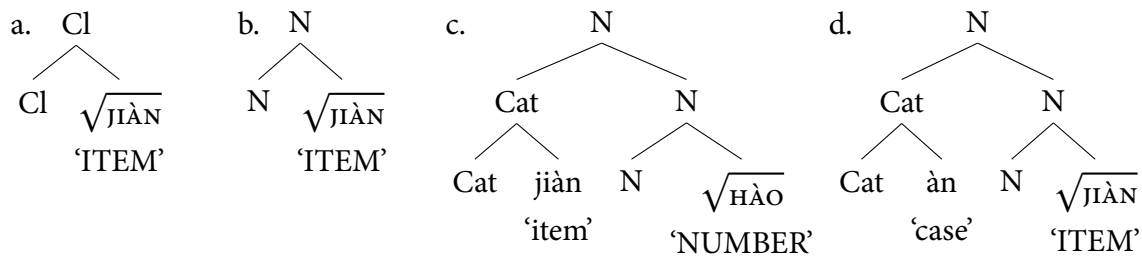
Classifiers can also be recruited as MHC components, either as modifiers or as heads. (56) are some examples (see also Loke 1997, Zhang 2013).

(56) a. *jiàn-hào* 'item-number' (as modifier)
gè-wèi 'individual-digit; (math) the unit'
dī-guàn 'drop-irrigate'

b. *àn-jiàn* 'case-item; (legal) case' (as head)
yún-duǒ 'cloud-flower.shape; cloudlet'
niú-qún 'cattle-crowd; herd'

Below I illustrate the different derivational scenarios of classifier roots with the example $\sqrt{\text{JIÀN}}$, which is derived as a classifier in (57a), a noun in (57b), a nominal-origin compound modifier in (57c), and a compound head in (57d).

(57)



Second, the classifier-noun pairing is a matter of lexical-semantic selection; the selection is often idiosyncratic and must be memorized (Li & Thompson 1981). For example, the classifier for donkeys is *tóu* 'lit. head' while that for horses is *pǐ* 'lit. cloth bolt', and the classifier for paintings is *fú* 'lit. cloth/silk textile' while that for photos is *zhāng* 'lit. sheet'. This situation is reminiscent of gender assignment in languages like German, where the noun-gender pairing is often idiosyncratic too (e.g., *der Zug* 'the.MASC train', *das Ende* 'the.NEU end').

Third, when multiple classifiers are acceptable for a single noun, the choice of classifier is often conditioned by register, as in (58).

(58) a. *yí-gè* *lǎoshī*
one-CL teacher
'a teacher (neutral∼casual, not always polite)'

b. *yí-wèi* *lǎoshī*
one-CL teacher
'a teacher (respectful)'

c. *yì-míng* *lǎoshī*
one-CL teacher
'a teacher (professional; e.g., in self-introduction)'

(59) a. *yì-zhī* *zhū*
one-CL pig
'a pig (neutral, sometimes affectionate; e.g., in *Three Little Pigs*)'

b. *yì-tóu zhū*
one-CL pig
'a pig (usually domestic, sometimes pejorative)'

4.2.2.2 Pronouns

The three textbook pronouns in Mandarin are *wǒ* 'I', *nǐ* 'you', and *tā* 'he, she, it'. However, these are far from the only pronominal items in contemporary use. Table 4.7 presents some alternative pronouns. I give literal meanings where available; all examples are interpreted singularly even when their original/literal meanings are plural (e.g., *ǎn*, *nín*).

Table 4.7 Pronouns in Mandarin

Person	Pronoun	Description
1st	*wǒ*	standard
	ǎn 'lit. we, our'	from Northern Mandarin
	ālā 'lit. we, our'	from Ningbo/Shanghai Wu
	ǒu	*wǒ* in Southern accent, popular on internet
	lúnjiā 'lit. other people'	distorted pronunciation of *rénjiā*, cute
	wú	archaic, formal
	zài-xià 'lit. be.at-below'	archaic, humble
	zhèn	archaic (since Qin), used by emperors
	gū 'lit. lone'	archaic (pre-Qin), used by kings and dukes
	guǎ-rén 'lit. lacking-person'	archaic (pre-Qin), used by kings and dukes
	āi-jiā 'lit. mourn-er'	archaic, used by empress dowagers
	běn-gōng 'lit. this-palace'	archaic, used by emperors' wives or children to inferiors
	chén-qiè 'lit. subject-concubine'	archaic, used by emperors' wives to superiors
2nd	*nǐ*	standard
	nín 'lit. you all'	standard, respectful
	qīn 'lit. dear'	colloquial, popular on internet
	jūn 'lit. monarch'	archaic, formal, respectful
	qīng 'lit. an official rank'	archaic, emperor to state officials or between lovers
3rd	*tā* 'lit. other'	standard
	tān 'lit. they, others'	from Beijing Mandarin, respectful
	yā 'lit. (low-status) girl'	from Beijing Mandarin, vulgar

Many of the alternative pronouns in Table 4.7 are absorbed from other synchronic or diachronic Chinese varieties. For example, the first-person dialectal pronouns *ǎn* (Northern Mandarin) and *ālā* (Wu) made their way into Common Mandarin via literary works and (mainly) the

mass media; similarly, many archaic or royal pronouns, which in the modern era (since the early twentieth century) had long been restricted to formal or literary registers, have reentered the colloquial language due to the wide popularity of historical TV dramas in the past few decades. There have been public debates as to whether some of these vocabulary items (e.g., *āijiā*, *chén-qiè*) were really used in history or merely coined by (ancient) novel/opera writers.[32] However, for ordinary speakers today the origins of the alternative pronouns are less important; the more interesting observation is that they seem to have somehow filled in the blank of first-person honorific pronouns in Modern Chinese, at least in internet language, where they are most often used. See (60) for examples.

(60) a. *cǐ lǜ̀-jìng shēn dé āijiā zhòng-yòng*
this filter-glass deeply get mourner heavy-use
'This (camera) filter is very much beloved by me. (humorously arrogant)'

b. *běn-gōng jīntiān xīnqíng bù jiā*
this-palace today mood not good
'I'm in a bad mood today. (humorously arrogant)'

c. *néng gěi chén-qiè diǎn gè shūcài ma*
can give subject-concubine order CL vegetable SFP
'Can you order a veggie dish for me? (humorously humble)' (from Weibo)

There are also alternative pronouns that have entirely been developed in internet language, and such items tend to catch on even more widely; for example, *ǒu* 'I (playful)' began its life as a (nonderogatory) mocking of Southern accent, and *qīn* 'you (affectionate)' originated as a greeting on the shopping website Taobao.[33] Below are some real-life examples.

(61) a. *lún-jiā xiǎng zài shuì yí huìr*
other-people want again sleep one while
'I wanna sleep for another while. (cute)'

b. *qīn juéde zěnmeyàng a*
dear feel how SFP
'How do you think? (affectionate)'

c. *qīng běn yìyuán, nàihé zuò-zéi*
you originally senator wherefore DO-thief
'You are a senator—why playing thief? (elegant, news title)'

[32] See Chen (2009) and especially Xia (2018) for historical evidence that they are real appellations.

[33] The pronominal/addressing *qīn* can be used in more than one grammatical category. It can also be used as a generic pronoun or common noun (which may have been influenced by the Sino-Korean word *cingu* 'friend'); e.g., *wǒmen xūyào xīnlǐxué zhuānyè de qīn* 'we need psychology major MODI friend; we need someone with a psychology degree'. Here I only focus on the pronoun usage of this vocabulary item.

d. *yā shuō yā méi qián*
he say he not have money
'That bastard says he has no money. (vulgar, mainly by Beijingers)' (from Baidu)

The examples in (60)–(61) show that the choice of pronouns in real-life Mandarin is subject to register-based conditioning. Thus, while *lúnjiā* is fine among family members or close friends, it is inappropriate in more formal circumstances (e.g., in official news); and while *qīn* is commonly used to show friendliness, it is not used when an affectionate tone is inappropriate (e.g., in police interrogations). Besides, when a particular pronoun is chosen, other parts of the sentence usually "agree" with it in register; for example, in (60a, b) the elegant/literary predicates *zhòngyòng* 'heavy-use; belove' and *bù jiā* 'not good' are used with the archaic-origin pronouns, and in (61c) the entire sentence is formulated in an archaic style to accompany *qīng*.[34] But such register agreement is not obligatory, as evidenced by the use of the archaic *chénqiè* in an otherwise completely modern and colloquial sentence in (60c).[35]

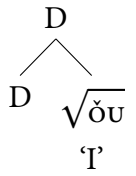
Also note that the alternative pronouns in Table 4.7 are distinct from another class of pronominal items known as (pseudo-)imposters (Collins & Postal 2012, Wang 2014), because they have lexically fixed φ values, whereas imposters get φ values from the context, as in (62).

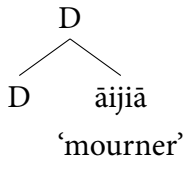
(62) a. *lǎoshī kuài yào shī-qù wǒ/tā de nàixìng le* (pseudo-imposter)
teacher almost will lose-COMP I/(s)he POSS patience SFP
'Teacher (=I/(s)he) will lose my/his/her patience.' (adapted from Wang 2014: 175)

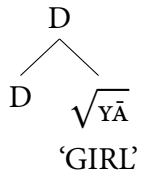
b. *āijiā kuài yào shī-qù wǒ/*tā de nàixìng le* (pronoun)
mourner almost will lose-COMP. I/(s)he POSS patience SFP
'I'm almost losing my/*his/*her patience.'

In (62a) *lǎoshī* may either refer to the speaker or a third party depending on the context, whereas in (62b) *āijiā* can only refer to the speaker whatever the context is. Imposters are widely used in Chinese, but the items in Table 4.7 all behave like authentic pronouns.

Since the alternative pronouns all have a lexical origin and are subject to stylistic licensing based on their lexical semantics, they qualify as semifunctional items. Assuming that the category of pronouns is D (or some span equivalents thereof), I derive them as follows.

(63) a. 

b. 

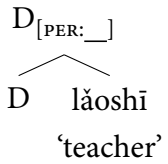
c. 

[34] Strictly speaking the usage of *qīng* in (61c) is wrong, because *qīng* is the way an emperor addresses officials (i.e., from superior to inferior), and the journalist producing (61c) is surely not an emperor. This demonstrates that the contemporary use of archaic-origin pronouns does not always abide by their original senses.

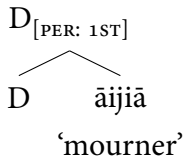
[35] See Feng (2010, 2012, 2019) for further discussion on the "register grammar" in Chinese.

I take the recruited material for *ǒu* as a root even though it originated as a distorted *wǒ* (i.e., the standard 'I'), because the recruited material is no longer an exponent of [+1ST, +SG]—if it were it would be indistinguishable from *wǒ*—but has developed its own encyclopedic content. On the other hand, the recruited material in *āijiā* is a more complex construction (an MHC); this is allowed in GRS because in Figure 4.1 ω may be either a root or a root-like chunk. Finally, imposters—being pronominal—may also be given a GRS-based analysis. They differ from the alternative pronouns in that they lack lexicalized or idiomatized meanings (i.e., they have no special encyclopedic entries), as in (64); in other words, categorization need not always be accompanied by lexicalization. Most of the semifunctional items we have seen are indeed lexicalized, so imposter items may represent a special subtype of semifunctional item.

(64) a. Nonidiomatic



b. Idiomatic



4.2.3 Prepositional domain

In §4.2.1.2 I mentioned that some Chinese voice auxiliaries resemble English prepositions. This is not a coincidence but reveals a deeper intercategorial connection: voice auxiliaries and prepositions in Chinese both have lexical origins, and they both can still be used lexically; in other words, they are both rooted items.[36] I have illustrated this situation for voice auxiliaries in §4.2.1.2; see (65)–(66) for examples of prepositions.[37]

(65) a. *tā* *cóng* *nán-fāng* *lái* (prepositional usage)
he from south-side come
'He comes from the south.'

b. *wǒ* *bǎ* *tāng* *hē-diào* *le*
I DISP soup drink-COMP SFP
'I ate the soup.'

c. *fànrén* *xiàng* *chá-lǐ* *xià-le* *dú*
criminal to tea-inside put in PERF
'The criminal poisoned the tea.'

[36] See Mardale (2011) for a similar proposal that prepositions are a "semilexical category."

[37] Chinese also has "postpositions"; e.g., *jiàoshì-lǐ* 'classroom-in; in the classroom', *zhuōzi-shàng* 'table-on; on the table'. However, items like *lǐ* 'lit. inside' and *shàng* 'lit. upper side' are intuitively more nominal than adpositional, so I treat them as location nouns (i.e., lexical category items). See Huang et al. (2009) and Song (2017a) for discussions.

(66) a. *tā cóng-guò jūn* (verbal usage)
he join-EXP army
'He once joined the army.'

b. *xiǎo-gǒu jià-chē qīnzì bǎ-zhe fāngxiàng-pán*
small-dog drive-car personally hold-PROG direction-wheel
'The puppy is driving and holding the wheel by itself.'

c. *kuíhuā xiàng-zhe tàiyáng*
sunflower face-PROG sun
'The sunflower is facing towards the sun.'

In their lexical uses, *cóng*, *bǎ*, and *xiàng* respectively mean 'join, follow', 'hold by hand, guard', and 'face, favor'. The lexical origin of voice auxiliaries is crosslinguistically common (e.g., English *get*, *be*), but that of prepositions is a special characteristic of Chinese (and some other analytical languages like Vietnamese). For this reason, Li & Thompson (1981) call Chinese prepositions *coverbs*. Song (2017a) analyzes Chinese prepositions as VP-shell items, more exactly applicative heads in the sense of Pylkkänen (2008), which are used to introduce additional arguments to verbal predicates;[38] see Aldridge (2012) for a similar view.

That Chinese prepositions are rooted is also evidenced by the fact that their uses are subject to lexical-semantic and stylistic conditioning, as in (67)–(68).

(67) a. *dǎbàn yí-xià gēn /??yǔ wǒ zǒu*
dress up one-CL with with me go
'Spruce up and come with me. (colloquial; e.g., two friends going to a party)'

b. *wǔhuì-shàng wángzǐ ??gēn /yǔ gōngzhǔ gòng-wǔ*
ball-at prince with with princess together-dance
'The prince danced with the princess at the ball. (literary; e.g., in a fairy tale)'

(68) a. *wǒ dǎ /??zì zuórgè yè-lǐ yìzhí dǎ-hāqiàn*
I since since yesterday night-inside constantly DO-yawning
'I've been yawning since last night. (colloquial; e.g., between friends)'

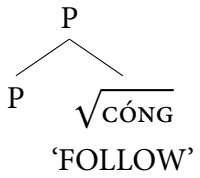
b. *tā ??dǎ /zì shào-shí zuì-xīn wénxué*
he since since young-time drunk-heart literature
'He has been engrossed in literature since young. (elegant; e.g., in a biography)'

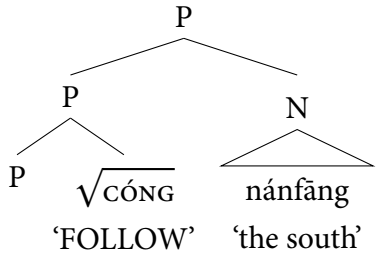
In (67) both *gēn* and *yǔ* mean 'with', but *gēn* is more colloquial; in (68) both *dǎ* and *zì* mean 'since, from', but *zì* is more elegant. Such many-to-one correspondences between Chinese and English

[38] Svenonius (2014) uses the term "generalized applicative" for all non-main-verb argument-introducing heads.

prepositions are common; see Table 4.8 for more examples (based on Chao 1968).

Under GRS I derive Chinese prepositions by merging an argument-introducing category with various roots, as in (69). The label P is expository; I remain agnostic about the exact functional category (or span) for prepositions (any proposal would be compatible).

(69) a. 

→ b. 

4.2.4 Acategorial domain

Finally, remember from §4.1 that the generalized categorizer X is not confined to mergeme-bearing categories—all the derivational schema requires is that X be a non-Root; namely, that it has some (model-theoretically) composable denotation. So, rooted or semifunctional logical operators should be possible in human language too. This is borne out in Chinese. See Table 4.9 for a nonexhaustive list of AND and OR conjunctions in Mandarin.

We can observe several patterns in Table 4.9. For instance, the AND items tend to be specialized for particular categories, and most OR items come from adverbials. But the most important observation is that the vast majority of logical operators in Mandarin are recruited from lexical categories, and many of them can still be used lexically, as in (70).

(70) a. *èr yǔ sān de hé shì wǔ*
two and three POSS union is five
'The sum of two and three is five.'

b. *bǎ liǎng zǔ bìng zuò yì zǔ*
DISP two group merge as one group
'merge two groups into one'

c. *tāmen míngtiān huò kě dàodá*
they tomorrow perhaps can arrive
'Perhaps they can arrive tomorrow.'

That the conjunctions are subject to lexical or stylistic conditioning is evidenced in (71)–(72).

Table 4.8 Prepositions in Mandarin

Preposition	Literal meaning	English translation
zài	'be present'	at, on, in (neutral)
dāi	'stay'	at, on, in (colloquial)
āi	'be close to'	at, on, in (colloquial)
gēn$_1$	'follow'	at, on, in (colloquial)
hé	'harmony, union'	with (neutral)
gēn$_2$	'follow'	with (colloquial)
tóng	'be the same'	with (formal)
yǔ	'get along'	with (formal)
dào	'arrive'	to, toward, till (neutral)
shàng	'ascend'	to, toward (colloquial)
wàng	'look over'	to, toward (colloquial)
wǎng	'go, pass'	to, toward (colloquial)
xiàng	'face, turn towards'	to, toward (neutral)
cháo	'face, look'	to, toward (colloquial)
chòng	'collide, dash against'	to, toward (colloquial)
duì	'reply, be directed at'	to, toward
gěi$_1$	'give'	to, for
cóng	'follow, join'	from (neutral)
dǎ	'hit'	from, since (colloquial)
yóu	'pass through'	from, by, through (formal)
zì	'(archaic) origin, source'	from, since (formal, literary)
yán	'follow, continue'	along (neutral)
shùn	'obey'	along (colloquial)
yòng	'use'	with, by means of (neutral)
ná	'take, hold'	with, by means of, DISPOSAL (colloquial)
bǎ	'hold'	DISPOSAL (neutral)
guǎn	'manage, govern'	DISPOSAL (colloquial)
gěi$_2$	'give'	DISPOSAL (colloquial)
jiāng	'support with hand, hold'	DISPOSAL (formal, literary)

Table 4.9 Conjunctions in Mandarin

Conjunction	Literal meaning	Logical function
hé	'harmony, union'	and (general purpose)
yǔ	'get along'	and (formal, literary)
tóng	'be the same'	and (for nouns, colloquial, old-fashioned)
gēn	'follow'	and (for nouns, colloquial)
jí	'(archaic) accompany'	and (for nouns, formal, objective)
jì	'(archaic) and'	and (for nouns, highly formal, solemn)
bìng	'juxtapose, merge'	and (for verbs)
qiě	'just, even'	and (for adjectives, formal~written)
ér	'(archaic) and'	and (for adjectives, formal~spoken)
(ér-)yòu	'(and-)again'	and (for adjectives, colloquial)
huò(zhě)	'(archaic) perhaps'	or (general purpose)
huò-shì	'or-is'	or (colloquial)
yì-huò(-shì)	'(archaic) perhaps-perhaps(-is)'	or (for alternatives)
hái-shì	'still-is'	or (for alternatives)

(71) a. *jīròu hé /?yǔ niúròu dōu hěn hǎochī*
chicken and and beef both very delicious
'Chicken and beef are both delicious. (colloquial)'

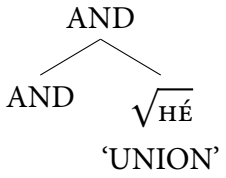
b. *hālì bōtè ?hé /yǔ mófǎ-shí*
Harry Potter and and magic-stone
'Harry Potter and the Philosopher's Stone (literary, elegant)'

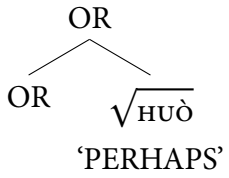
(72) *wúlùn shì zìránjiè háishì /??yìhuò shèhuì, ??háishì/yìhuò rénshēng…*
no matter is nature or or society or or human life
'Whether in nature or in society, or in human life…'

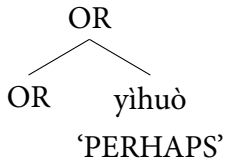
Between the two AND items in (71) *yǔ* is more literary, and the felicity of the two OR items in (72) hinges on their relative order. There are also ordering restrictions on AND items. See (73).

(73) *měi-nǚ yǔ /??jí yě-shòu ??yǔ /jí qítā*
pretty-girl and and wild-beast and and others
'Beauty and the Beast and Other Stories (book title)'

In (73) *yǔ* and *jí* can hardly be swapped, which may have to do with the fact that *yǔ* creates equal conjunction whereas *jí* creates unequal conjunction (Xue 2009, He 2010). In sum, logical operators in Chinese do behave like semifunctional items. I illustrate their derivation in (74).

(74) a. 

b. 

c. 

4.2.5 Interim summary

The foregoing data reveal two characteristics of Chinese function words:

1. They have a lexical origin and synchronically still have independent lexical usage.
2. They are not completely bleached but retain enough encyclopedic content to manifest lexical-semantically conditioned or register-sensitive distribution.

These match the typical characteristics of semifunctional vocabulary items predicted by GRS. Therefore, I analyze Chinese function words as root-supported heads. Root-supported heads are derived separately from the derivational spine and are hence similar to complex specifiers or adjuncts (i.e., satellites), though they differ in that they are labeling heads on the spine.

4.3 Further implications

Next I discuss two further implications of GRS based on the data in §4.2, respectively concerning analyticity (§4.3.1) and grammaticalization (§4.3.2). Whether these have wider crosslinguistic significance awaits future investigation, but I tentatively assume that root support is another "free rider" strategy in human language syntax made available by UG-allowed categorial flexibility.

4.3.1 Root support and analyticity

4.3.1.1 Deducing the analyticity typology

The free-rider status of root support means that it as a syntactic strategy is universal, just like adjunction. Both the underspecified mergeme (which underlies the defective category) and the unspecified mergeme (which underlies the Root category) are automatically available once the first mergeme-based categories become available, as in (75).

(75) The availability of [MERGEME:⟨CAT: C, POS: k⟩] entails
the availability of [MERGEME:⟨CAT:_, POS:_⟩] and
the availability of [MERGEME:_].

The universality of root support is trivially evidenced by the universality of lexical category items. Content words like nouns and verbs are derived by combining a mergeme-based category and a root, with the former syntactically licensing the latter and the latter interpretively supporting the former. Beyond the trivial universality of the root support mechanism, what is more interesting is its application outside the lexical domain. There are three possible scenarios:

- **Scenario I:** Root support is confined to the traditional lexical domain.
- **Scenario II:** Root support is moderately extended to the functional domain.
- **Scenario III:** Root support is heavily or even fully extended to the functional domain.

The three scenarios largely correspond to the three conventional morphological types: scenario I corresponds to polysynthetic languages, scenario II to moderately synthetic/analytic languages, and scenario III to highly analytic or even isolating languages.

In §4.1.3.2 we saw that root support has a bearing on PF-oriented head movement, which is motivated by the need to morphophonologically support an abstract head (Borer 2005a). For nonabstract heads Borer only considers f-morphs, which are independent grammatical functional formatives like *the* and *will*. F-morphs are by definition rootless in Borer's system, because they are stored in the functional lexicon while roots (aka "listemes") are stored in the "encyclopedia" (Borer 2005a: 30–31). However, as we saw in §4.1.3.2, the (un)availability of PF-oriented head movement is only dependent on the morphophonological nature of a potential host head but not on the source of that nature—that is, whether the head happens to have a morphophonologically independent exponent, as in French *que* 'that (C)', or it has a derived freestanding status via a supporting root, as in Chinese *dǎ* 'hit (V.DO)'. Since neither *que* nor *dǎ* can host head movement, a purely f-morph-based theory would misclassify the latter as functional formatives when it is really a semifunctional item; for example, Borer (2005a: 95) treats Chinese-style classifiers as f-morphs. In this sense, GRS is an extension of Borer's idea, as it further distinguishes two types of freestanding heads. The extension, combined with the three scenarios of root support, predicts the following typology:

(76) Morphological typology based on head movement

a. Polysynthetic languages make heavy use of head movement.

b. Moderately synthetic/analytic languages make moderate use of head movement.

c. Highly analytic languages make sparse or even no use of head movement.

This largely matches the actual typology.[39]

[39] See Roberts (2019: Chapter 4) for a potentially similar idea.

4.3.1.2 Criterion of analyticity

Above I reinterpreted Borer's insight about (PF-oriented) head movement as one about analyticity. Head movement creates morphosyntactically complex heads, which contain more than one syntactic category and perhaps also more than one morpheme, and the analyticity/synthesis distinction in its conventional sense is precisely defined by a category/morpheme-per-word ratio (Sapir 1921, Greenberg 1960). More specifically, we have the following criterion:

(77) Criterion of analyticity
An analytic language has a very low category/morpheme-per-word ratio.

Here *word* is understood as a morphophonologically freestanding unit. Since head movement creates complex heads and hence complex words, we expect an analytic language to only use it sparsely. Importantly, what analyticity repels is not head complexity per se but the syntactic creation thereof for morphophonological reasons. This is what distinguishes head-movement-created complex heads from root-supported heads, which strictly speaking are also complex. I call the former *synthetic heads* and the latter *analytic heads*. Root support differs from head movement in that it does not add extra syntacticosemantic categories to the host head (since roots are categoryless) and therefore only increases the morpheme-per-word ratio of a language but not its category-per-word ratio—and even this morpheme-per-word ratio increase is of a much smaller scale than that which results from head movement, because head movement tends to be roll-up, as in (78a), while root support is usually one-off, as in (78b).[40]

(78) a. *Yes p'amidoR-a* [$_{\text{V-Voice-Asp-T}}$ *čor-a-ts-r-ets-i*] [Eastern Armenian]
I tomato-ACC dry-INCH-CAUS-ASP-AOR-1SG
'I dried the tomato.' (adapted from Megerdoomian 2002: 213)

b. [$_{\text{Q}\sqrt{}}$ *yì*] [$_{\text{Cl}\sqrt{}}$ *zhāng*] [$_{\text{N}\sqrt{}}$ *zhǐ*] [Mandarin]
one piece paper
'a piece of paper'

The one-off nature of root support may be due to lexicalization, for a root-supported head has its own semantic-encyclopedic entry. While complex heads may be created on the fly (e.g., via head movement), lexicon entries may not. In more conventional theories (including standard minimalism) all three words in (78b) are viewed as LIs and hence atomic heads; by contrast, in GRS only their sounds and meanings are lexicalized, while their categorial structure (i.e., {X, √}) is completely regular (i.e., derivable) and hence syntactic. So, GRS is part of an enterprise that syntactically represents the lexicon: it extends the lexical decomposition method to the nonlex-

[40] An apparent exception is derivational morphology, where the output of one categorization cycle can be input to another one, as in [$_{\text{N}}$ [$_{\text{V}}$ [$_{\text{A}}$ [$_{\text{N}}$ *institution*]-*al*]-*iz*]-*ation*]. However, such "recategorization" is not really roll-up, since each cycle of categorization presumably happens at a separate derivational layer.

ical domain and helps maintain the tenet that the language faculty only needs one generative engine (i.e., is syntax-all-the-way-down in DM terms; Harley & Noyer 1999, Siddiqi 2010).

Recall from §4.2 that Chinese semifunctional items are predominantly composed of a null syntactic category (X) and a root (√), so they have a category/morpheme-per-word ratio of 1 : 1 and may be called *analytic heads par excellence*. When X is overt, the semifunctional item has a category-per-word ratio of 1 : 1 but a morpheme-per-word ratio of 2 : 1. No such item exists in the Chinese data in §4.2, but English provides some potential examples; namely, modal auxiliaries like *should*, *would*, and *could*. These can be analyzed as root-supported Modal heads, where the Modal category is overt and affixal (realized irregularly). Since each root-supported head is accompanied by a lexical entry, *should*/*would*/*could* are not merely inflections of *shall*/*will*/*can* but also separate lexical items. So, "inflected" auxiliaries are another type of analytic head (call them *inflected analytic heads*), though they are less analytic than the Chinese type in terms of morpheme-per-word ratio. Now we can reformulate the typology in §4.3.1.1 as follows:

(79) Morphological typology based on analytic head

a. Highly analytic languages make heavy use of analytic heads, most of which are analytic heads par excellence.

b. Moderately analytic/synthetic languages make moderate use of analytic heads, which may be analytic heads par excellence or inflected analytic heads.

c. Polysynthetic languages make sparse use of analytic heads.

4.3.1.3 Is there an analyticity parameter?

Huang (2015) studies the high analyticity of Mandarin in comparison with English, Old Chinese, and some Modern Chinese dialects. He proposes an analyticity macroparameter that consists of a cluster of microparameters, each responsible for an analytic property. Huang's analytic properties include the following (among others).

Table 4.10 Huang's (2015) analytic properties

V domain	N domain	T/C domain
light verb construction	numeral classifier	wh-in-situ
pseudo noun incorporation	localizer	discontinuous *wh-the-hell* construction
verbal atelicity	discontinuous preposition	no negative quantifier
no verbal coercion	positive degree marker	no reciprocal pronoun
		Kaynean word order par excellence
		absence of canonical gapping

Huang explains these properties by two formal features, [+EPP] and [+strong], with the former triggering phrasal movement and the latter head movement, as in the quote below:

> [T]he difference in the degree of analyticity can be characterized … within the Probe-Goal system on the basis of the nature of a probing head: a head with [+EPP] or a [+strong] affixal feature leads to movement and synthesis, while [−EPP] and a [−strong] head preserves analyticity by leaving elements in situ. While each case of the variation may be described in terms of a micro-parameter associated with a particular probing head, the clustering of parametric values across the board gives rise to … the analytic-synthetic macro-parameter. (Huang 2015: 34)

Huang's proposal relates analyticity to movement in an even tighter way than my proposal. On the one hand, the [+strong] feature is more or less a syntacticized version of my PF-oriented reasoning (and that of Borer). For instance, Huang explains the lack of noun-to-light-verb incorporation in Mandarin as follows:

> [T]he light verb is phonetically non-null, thus blocking noun incorporation. In particular, the light verb is phonetically independent, thus at least not an affix, and not [+strong] in the relevant sense. (Huang 2015: 11)

Huang does not specify the correlation between "affix" and "strong." In one place he seems to equate the two:

> An F feature may be "strong" (or affixal) or "weak." A [+strong] feature F marks a strongly defective head that requires licensing by overt head movement (Move), while a [−strong] feature may be licensed simply under an appropriate matching configuration (Agree). (Huang 2015: 10)

However, if strong means affixal, then what is the ground for the syntacticization of such an apparently morphophonological property? After all, Borer's work demonstrates that it is feasible to let the affixal property remain morphophonological. In addition, the ontological distinction between [+strong] and [+EPP] and its consequences are unclear;[41] we know from Huang's description that the former is responsible for head movement and the latter for phrasal movement, but what information must the syntax be granted to implement this division of labor? And how feasible it is for children to acquire two movement features? Overall, resorting to purely formal features to account for a highly lexically oriented phenomenon like analyticity seems to bring about more puzzles than it resolves.

On the other hand, [+EPP] or phrasal movement is entirely absent from the traditional perspective on analyticity, which treats it as a dimension of morphological typology. Huang is more

[41] This is not only Huang's problem but a general problem in the field.

interested in "syntactic analyticity," though again the term is not explicitly defined. Insofar as the properties in Table 4.10 (especially those in the T/C domain) are concerned, it seems quite different from the traditional perspective, and whether those properties are all indicators of analyticity is debatable. For instance, wh-in-situ also exists in Japanese and Korean, and the Japanese/Korean counterpart of *wh-the-hell* is discontinuous too, as in (80). Since Japanese and Korean are quintessentially agglutinative in the T/C domain, properties like wh-in-situ and discontinuous *wh-the-hell* may not be good indicators of high analyticity.[42]

(80) a. *ittai* *kare-wa* *nani-o* *katta* *no?* [Japanese]
the hell he-TOP what-ACC bought QUESTION
'What the hell did he buy?' (adapted from Huang & Ochi 2004)

b. *daeche* *neo* *nugu* *ya?* [Korean]
the hell you who QUESTION
'Who the hell are you?'

Faced with issues like the above, I choose not to parameterize analyticity in terms of formal features, nor do I treat it as a cluster property. Instead, I view morphological typology in the traditional sense as a matter of tendency; under GRS it can be conceived as a tendency in the application of a universally available structure-building strategy, root support. Being a tendency, analyticity may spread or shrink within a language and wield areal influence across languages at its own pace (i.e., independently from other dimensions of language change, such as the change in word order). Since root-supported heads are lexicalized, this is ultimately a lexicon-centered variation, though it differs from the classical BCC in not being encoded in formal features. In relation to this, I find the following quote from Borer (2005a) appealing:

> [A]ll variation, both within a language and across languages, is reducible not only to the properties of range assigners to functional open values, but [also] to their morpho-phonological properties. (Borer 2005a: 264)

Based on this I tentatively propose an *extended Borer-Chomsky conjecture*:

(81) Extended Borer-Chomsky conjecture
All parameters of variation are attributable to the lexicon, either to differences in the formal features of functional heads or to differences in their morphophonological nature.

Thus, if analyticity were parameterizable, it would be more appropriately parameterized along the morphophonological path and independently from the formal feature path.

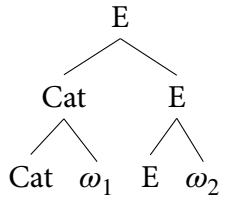
[42] Ian Roberts (p.c.) points out that Huang's properties are not meant to be considered indicators of analyticity in isolation but only codefine it when they cluster. While I am sympathetic to this perspective (which is essentially based on Biberauer & Roberts' 2015 parameter hierarchy), to what extent the empirical phenomena in Huang (2015) lend themselves to this perspective requires more careful evaluation.

4.3.2 Root support and grammaticalization

4.3.2.1 Grammaticalization routes

In §3.3.4 I proposed two grammaticalization routes for Chinese SFPs based on the defective category schema (repeated in (82)): (i) via direct recruitment of some available material as ω_2 (e.g., the general-emotionality *a* comes from the interjection *a*); (ii) via fusion of some material at the end of ω_1 into ω_2 (e.g., the yes/no-signaling *ma* is a fusion of the sentence-final adverbial *mu* 無 'not have' and *a*).

(82) (defective category schema for SFP)



Route (ii) resembles grammaticalization in the conventional sense; in particular, it instantiates the grammaticalization type where the source and the target categories are not in the same projection line but are only linearly adjacent. Route (i), on the other hand, is less conventional, as it does not follow any projection line at all; perhaps a better term for this route is *recategorization* or *relexicalization*, since it assigns a new syntactic category and a new lexicon entry to any material that happens to be chosen. In §4.2.1.5 I further identified {E, ω_2} as a case of root support, so the grammaticalization of SFPs also has implications for that of other root-supported heads. Root-supported heads include lexical category items and semifunctional items; since the former are not results of grammaticalization, below I restrict my discussion to semifunctional items.

Most of the semifunctional items in §4.2 (all but the SFPs) have nothing to do with the defective category, so route (ii) is less relevant. As for route (i), no categorial change is involved in the establishment of an interjection-origin SFP, which simply recruits a root-supported head as the head of a Cat-E construction. Meanwhile, the grammaticalization of most other semifunctional items does involve categorial change. Thus, when a noun like *tóu* 'head' is grammaticalized into a classifier, its category changes from N to Cl. I assume this to be a two-step process: (i) the noun is recruited to support Cl; (ii) the recruited material loses its nominal status. Schematically:[43]

(83) $\{\text{N}, \sqrt{\text{TÓU}}\} \rightarrow \{\text{Cl}, \{\text{N}, \sqrt{\text{TÓU}}\}\} \rightarrow \{\text{Cl}, \sqrt{\text{TÓU}}\}$

When I specify some semifunctional item's grammaticalization path as $\{X, \omega\} \rightarrow \{Y, \omega\}$, I take this as a shorthand for $\{X, \omega\} \rightarrow \{Y, \{X, \omega\}\} \rightarrow \{Y, \omega\}$. Here the relation between X and Y is not fixed—they could be either in the same projection line (in which case Y is hierarchically above X, à la Roberts & Roussou 2003) or from different projection lines (e.g., $that_D > that_C$). Alternatively,

[43] Alternatively the recruited material could be a root-like chunk (i.e., a flattened sound-meaning pair; see §3.2.1) with no syntactic category N but merely a lexicalized semantic type `entity`. Under this assumption the second step of change in (83) is simply a matter of semantic bleaching.

either or both of them may not be from any projection line at all (but be acategorial); this is the situation in, say, the grammaticalization of the noun *hé* 'union' into the conjunction *hé* 'and'. In sum, I list the following grammaticalization paths for semifunctional items:

(84) a. $\{X, \omega\} \rightarrow \{Y, \omega\}$, where X, Y are from the same projection line

b. $\{X, \omega\} \rightarrow \{Y, \omega\}$, where X, Y are from different projection lines

c. $\{X, \omega\} \rightarrow \{Y, \omega\}$, where X or Y is acategorial (i.e., a logical operator)

d. $\{X, \omega\} \rightarrow \{\{Cat, \psi\}, \{X, \omega\}\}$, mainly for interjection-origin SFPs

e. $\psi + \{X, \omega\} \rightarrow \{X, \psi\omega\}$, mainly for fused SFPs

Note that in (84e) when ψ and ω are fused, they are very likely to form a new root. For example, when the MidC sentence-final adverbial *mu* got fused with *a*, the new SFP *ma* gained an independent lexical entry as a separate SFP $\{E, \sqrt{\text{MA}}\}$, which retains certain lexical semantic content of both *mu* and *a* but is not a simple composition of the two. While *mu* 'not have' can only be attached to a positive proposition ('lit. … or not?', Sun 1999), *ma* is not subject to this restriction. The sentence-final usage of *mu* is preserved in Min; observe in (85) that *bo* (=*mu*) is infelicitous after a negative proposition, while the same construction is perfectly fine with *ma*.

(85) a. *i u / *bo ki bo?* [Taiwan Southern Min]
he have not have go not have
'Did he (*not) go?' (adapted from Cheng et al. 1996: 74)

b. *tā yǒu / méi-yǒu qù ma?* [Mandarin]
he have not-have go SFP
'Did he (not) go?'

With the above grammaticalization paths in hand, let us reexamine the Chinese semifunctional items. For expository convenience I only consider items whose supporting materials are roots instead of root-like chunks. I illustrate the grammaticalization situations of the nine grammatical domains in Table 4.11, each with a representative example.

4.3.2.2 Multifunctionality

In the grammaticalization of a semifunctional item, since neither the root nor its original categorization is lost, we effectively obtain a multifunctional vocabulary item. And since the semifunctional item may participate in further grammaticalization, its multifunctionality may end

[44] I view "light verbs" in light verb constructions as lexical verbs (see §4.2.1.1), so no grammaticalization qua categorial change happens for them.
[45] I assume different grammaticalization routes for preverbal and postverbal aspectual elements (see §4.2.1).
[46] Or (84a) if the Chinese prepositions are in fact VP-shell items (see §4.2.3).

Table 4.11 Grammaticalization of Mandarin semifunctional items

Domain	Example	Grammaticalization path
Light verb	no grammaticalization[44]	-
Voice	*shǐ* 'use (V)' > *shǐ* 'use (CAUS)'	{V, $\sqrt{\text{SHǏ}}$} → {Voice, $\sqrt{\text{SHǏ}}$}, (84a)
Aspect[45]	*yǒu* 'have (V)' > *yǒu* 'have (PERF)'	{V, $\sqrt{\text{YǑU}}$} → {Asp, $\sqrt{\text{YǑU}}$}, (84a)
	guò 'pass (V)' > *guò* 'have (EXP)'	{V_{BE}, $\sqrt{\text{GUÒ}}$} → {Asp_{BE}, $\sqrt{\text{GUÒ}}$}, (84a)
Modal	*néng* 'able (A)' > *néng* 'can (MOD)'	{A, $\sqrt{\text{NÉNG}}$} → {Mod, $\sqrt{\text{NÉNG}}$}, (84b)
	dé 'obtain (V)' > *de* 'can (MOD)'	{V_{BE}, $\sqrt{\text{DÉ}}$} → {Mod_{BE}, $\sqrt{\text{DÉ}}$}, (84a)
SFP	*a* 'interjection' > *a* 'general emotionality'	{E, $\sqrt{\text{A}}$} → {{Cat, ω_1}, {E, $\sqrt{\text{A}}$}}, (84d)
	m(u) 'not' + *a* > *ma* 'yes/no'	*m* + {E, $\sqrt{\text{A}}$} → {E, $\sqrt{\text{MA}}$}, (84e)
Classifier	*tóu* 'head (N)' > *tóu* 'head (CL)'	{N, $\sqrt{\text{TÓU}}$} → {Cl, $\sqrt{\text{TÓU}}$}, (84a)
Pronoun	*qīn* 'dear (A)' > *qīn* 'you (D)'	{A, $\sqrt{\text{QĪN}}$} → {D, $\sqrt{\text{QĪN}}$}, (84b)
Preposition	*zài* 'be present (V)' > *zài* 'at (P)'	{V, $\sqrt{\text{ZÀI}}$} → {P, $\sqrt{\text{ZÀI}}$}, (84b)[46]
Acategorial	*hé* 'union (N)' > *hé* 'and (CONJ)'	{N, $\sqrt{\text{HÉ}}$} → {AND, $\sqrt{\text{HÉ}}$}, (84c)

up being strikingly rich. The shared root is what links all the categorizations (i.e., functions) of a multifunctional item together. See Table 4.12 for two examples.

Table 4.12 Grammaticalization of Mandarin multifunctional items

Root	Grammaticalization path
$\sqrt{\text{LIǍO}}$	{V, $\sqrt{\text{LIǍO}}$} (*liǎo*) 'end, finish' → {V_{BE}, $\sqrt{\text{LIǍO}}$} (*liǎo*) 'finished, completed' →
了	{Asp_{BE}, $\sqrt{\text{LIǍO}}$} (*le*) 'PERF' → ({Mod_{BE}, $\sqrt{\text{LIǍO}}$} (*liu*) 'can'[47]) → {Persp, $\sqrt{\text{LIǍO}}$} (*le*) 'CRS'[48]
$\sqrt{\text{DÉ}}$	(post-V) {V, $\sqrt{\text{DÉ}}$} (*dé*) 'get, obtain' → {V_{BE}, $\sqrt{\text{DÉ}}$} (*dé*) 'manage to' → {Mod_{BE}, $\sqrt{\text{DÉ}}$} (*de*) 'can'
得	(pre-V) {V, $\sqrt{\text{DÉ}}$} (*dé*) 'get, obtain' → {Voice, $\sqrt{\text{DÉ}}$} (*dé*) 'PASS (positive)' →
	{Modal_1, $\sqrt{\text{DÉ}}$} (*dé*) 'can' / {Modal_2, $\sqrt{\text{DÉ}}$} (*děi*) 'need, must' → {E, $\sqrt{\text{DÉ}}$} (*dé*) 'okay (INTERJ)'

In Table 4.12, there are only two/three pronunciations for the five functions of $\sqrt{\text{LIǍO}}$ (i.e., *liǎo*, *le*/*liu*) and also only two/three for the eight functions of $\sqrt{\text{DÉ}}$ (i.e., *dé*/*de*, *děi*). This kind of "systematic homophony" (Biberauer's 2017a term) makes Chinese a less "well-behaved" language in terms of grammaticalization compared with the familiar Indo-European languages. Consider Hopper & Traugott's (2003: 7) "cline of grammaticality":

[47] This usage of $\sqrt{\text{LIǍO}}$ is absent from Standard Mandarin but present in Northern Mandarin (see note 30, p. 118).

[48] I suggested in §4.2.1.3 that the CRS *le* might be located around (Chomskyan) *v*. According to Biberauer (2018, p.c.) there is a speaker-perspective-related category in this region, which I borrow and label as Persp. See Alcázar & Saltarelli (2014) and Ramchand (2018) for similar ideas.

(86) content item > grammatical word > clitic > inflectional affix

Insofar as the data in §4.2 are concerned, it seems that grammaticalization in Chinese hardly ever makes its way beyond the grammatical word stage. Even the three so-called aspect suffixes *le* 'PERF', *zhe* 'PROG', and *guò* 'EXP' are much less inflectional than the inflectional suffixes in fusional languages like English and Spanish or even those in agglutinative languages like Japanese and Korean. Among others, suffixes in synthetic languages are obligatory, whereas "suffixes" in Chinese are not. Thus, in a past/perfect sentence, the finite verb must inflect in English (87a), Spanish (87b), and Japanese (87c); by contrast, the perfective suffix *le* is only optional in Chinese, where past/perfect semantics may alternatively be expressed by lexical means (88).

(87) a. Yesterday I went/*go to the library. [English]

b. *Ayer fui / *voy a la biblioteca.* [Spanish]
yesterday went.1SG go.1SG to the library
'Yesterday I went to the library.'

c. *kinō toshokan-ni ikimashita / *ikimasu* [Japanese]
yesterday library-to went go
'Yesterday I went to the library.'

(88) a. *zuótiān wǒ qù-le túshūguǎn* [Mandarin]
yesterday I go-PERF library
'Yesterday I went to the library.'

b. *zuótiān wǒ yǒu qù túshūguǎn*
yesterday I have go library
'Yesterday I went / did go to the library.'

c. *zuótiān wǒ qù túshūguǎn le*
yesterday I go library CRS
'Yesterday I went to the library (for you information).'

It is clearly shown in (88) that the Mandarin perfective *le* is nonobligatory, so whether it is really an inflectional suffix or just a clitic or grammatical word is debatable; such classificatory difficulty is probably why functional morphemes in Chinese, including *le*, are often simply referred to as "particles." Under GRS the various particles in Chinese are just root-supported heads.

To conclude this section, I tentatively draw a connection between two facts:

1. Grammaticalization in Chinese seldom goes beyond the grammatical word stage.

2. Grammaticalization in Chinese is mainly via the recategorization of root-supported heads.

So, what "inhibits" the full manifestation of Hopper and Traugott's cline in Chinese is the systematic employment of root support, or in other words, its high analyticity.

4.4 Chapter summary

The mergeme-based template for syntactic categories allows for two types of flexibility: that characterized by an underspecified mergeme (i.e., the defective category) and that characterized by an unspecified mergeme (or the lack of a mergeme). In this chapter I studied the latter in more detail, which I have referred to as the acategorial category. There are two subtypes of acategorial category: the logical operator category, defined by purely compositional nonmergeme features, and the Root category, defined by the lack of any SCDP. A root, being radically underspecified, cannot appear without being categorized. Considering the stipulative nature of the marriage of roots and traditional lexical categories in previous root-based theories, I proposed a generalized categorization assumption, which redefines a potential categorizer as any non-Root category. Building on this assumption, I further proposed a generalized root categorization schema, $[_{X}$ X $\omega]$, and a generalized root syntax. The generalized root categorization schema defines a root-supported head, which may be either a traditional lexical category item or a semifunctional item; I have focused on the latter. Semifunctional items are constructed in separate derivational layers before joining the main layer (or spine); they are a third type of vocabulary item alongside the conventional lexical/functional division. GRS provides a solution to the semifunctionality problem; namely, how to fit half-lexical–half-functional items into a root-based theory.

Chinese provides a quintessential case of semifunctionality. My comprehensive investigation of Mandarin reveals two typical characteristics of semifunctional items: (i) they all have a lexical origin and retain their lexical usage synchronically; (ii) they all manifest lexical-semantically conditioned and register-sensitive distribution. These match the typical characteristics of root-supported heads predicted by GRS. The Chinese data have two further implications. First, root-supported heads can be identified as analytic heads because of their low category/morpheme-per-word ratio, which in turn gives us a root-based reformulation of the traditional morphological typology. Second, semifunctional items can be obtained via a number of grammaticalization paths, which fall in two major routes: reanalysis via direct recruitment (predominant) and fusion under adjacency (mainly for a subtype of SFP). GRS also provides a new angle to look at the rich multifunctionality in Chinese and similar languages.

Chapter 5

Incomparable categories and category flavoring

5.1 Climbing up the ladder of abstraction

Recall the ladder of abstraction in Chapter 2 as repeated below.

Table 5.1 Templates and flexibility types in the syntactic category system (= Table 2.1)

Level of concern	Abstraction layer	Template	Flexibility
Individual	single category	mergeme-based feature set	incomplete mergemes
	single c-sequence	scope-based partial ordering	incomparable categories
Global	single granularity level	universal spine	varied c-sequences
	all granularity levels	category subtyping	varied granularity levels
Supraglobal	entire SCS	syntactic rules	parametric variation

In Chapters 3–4 I studied two types of categorial flexibility, the defective category and the Root category, both of which are at the first abstraction layer. In the rest of this dissertation let us climb up the ladder of abstraction. I will examine the second abstraction layer in this chapter—which is also the end of the individual level of concern—and explore the global-level layers in Chapter 6 (the supraglobal level is not the focus of this dissertation). In the remainder of §5.1 I introduce a type of categorial flexibility at the single c-sequence layer (i.e., incomparable categories). In §5.2 I link this flexibility to the "flavored" categories (e.g., v_{DO}, v_{CAUS}) often seen in the minimalist (especially DM) literature. In §5.3 I present a case study on Chinese resultative compounds and argue based on diachronic evidence that the emergence of a flavored category (v_{BE} in the case of Chinese) is subject to more complicated conditions and has more widespread ramifications than previously thought; hence, flavored categories should be used parsimoniously as an analytical tool. §5.4 summarizes the chapter.[1]

[1]The material in this chapter is partly built on Song (2017b,c).

5.1.1 The categorial sequence poset

In §2.3.2 I characterized a categorial sequence or c-sequence as a set of syntactic categories immediately dominated by a granularity level set. For instance, if (1) is a granularity level, then {V $\leq v \leq$ T $\leq$ C} and {N $\leq n \leq$ Num $\leq$ D} (among others) are its c-sequences.

(1) {{V $\leq v \leq$ T $\leq$ C}, {N $\leq n \leq$ Num $\leq$ D}, …}

Above is a top-down definition of c-sequence. We can also define it in a bottom-up fashion:

(2) A *categorial sequence* or *c-sequence* is a complete sequence of mergeme-bearing categories at a given granularity level ordered by their mergeme values (i.e., first-Merge positions).

A c-sequence defined as such is a mergeme-based reformulation of the more familiar concept "hierarchy of projection" or "extended projection." I do not use these more conventional terms here because they are only meaningful in a derivational discourse, which we do not have when reasoning about the SCS at a purely ontological level. See (3) for more example c-sequences.

(3) a. {V $\leq v \leq$ Asp $\leq$ C}

b. {V $\leq$ Voice $\leq$ Asp $\leq$ T $\leq$ Mod $\leq$ C}

c. {N $\leq$ Num $\leq$ D $\leq$ K}

d. {N $\leq$ Cl $\leq$ Num $\leq$ Q $\leq$ D}

The binary relation $\leq$ reads "has a first-Merge position lower than or equal to" or "has a combinatorial scope smaller than or equal to." Since $\leq$ is reflexive, antisymmetric, and transitive, it is a partial order, so a c-sequence is a partially ordered set or poset in the mathematical sense, as defined in (4) (adapted from Schröder 2016: 1).

(4) A *partially ordered set* or *poset* is an ordered pair $\langle P, \leq \rangle$ of a set P and a binary relation $\leq$ contained in $P \times P$, called the *partial order* on P, such that

- $\leq$ is reflexive; namely, $\forall p \in P: \ p \leq p$.
- $\leq$ is antisymmetric; namely, $\forall p, q \in P: \ [(p \leq q) \wedge (q \leq p)] \Rightarrow (p = q)$.
- $\leq$ is transitive; namely, $\forall p, q, r \in P: \ [(p \leq q) \wedge (q \leq r) \Rightarrow (p \leq r)]$.

Take (3a) for example. The combinatorial-scope-induced relation $\leq$ is reflexive because each of the four verbal categories has a combinatorial scope (smaller than or) equal to itself; $\leq$ is antisymmetric because for each pair of distinct categories in the set, the smaller-than-or-equal-to direction is unique (e.g., we do not have Asp $\leq$ C and C $\leq$ Asp simultaneously); and $\leq$ is transitive because if a category is lower than another in terms of first-Merge position, then the former is also lower than all categories higher than the latter. In short, the concept of poset provides an

accurate formal description of c-sequences. Note that the reflexivity axiom is implicit in the $\leq$ notation; it becomes explicit if we display c-sequences in an alternative arrow format, as in (5).

(5) $\mathrm{V} \rightarrow v \rightarrow \mathrm{Asp} \rightarrow \mathrm{C}$

I assume each c-sequence member to be equipped with a loop arrow. The arrow notation will be used a lot in Chapter 6. Also note the word *complete* in (2). A c-sequence is complete if it contains all and only those syntactic categories that are in the same partial order. Thus, even though (3a) and (3c) may be from the same granularity level, they do not belong to the same c-sequence. Likewise, the following are not c-sequences either.

(6) a. $\{\mathrm{N}, \mathrm{V} \leq v \leq \mathrm{T} \leq \mathrm{C}\}$

b. $\{\mathrm{V} \leq v\}$ (suppose the granularity level also contains Asp, C)

c. $\{v_{\mathrm{A}} \leq \mathrm{Asp}_{\mathrm{A}}, \mathrm{V}_{\mathrm{B}} \leq \mathrm{Voice}_{\mathrm{B}}\}$ (A and B represent different granularity levels)

In sum, each mergeme-bearing category at a given granularity level is a member of a unique c-sequence, and all mergeme-bearing categories at a given granularity level are partitioned into disjoint c-sequences. I take this to be the basic ontological organization of the SCS.

5.1.2 Incomparable categories

There is a further commonality among the c-sequences in (2)–(3): every two categories in a c-sequence makes a comparable pair. This additional characteristic makes the c-sequences in question not only posets but in fact tosets (= totally ordered sets) or chains, as defined in (7) (adapted from Davey & Priestley 2002: 3).

(7) Let $\boldsymbol{P}$ be an ordered set. Then $\boldsymbol{P}$ is a *totally ordered set* or *chain* if, for all $x, y \in \boldsymbol{P}$, either $x \leq y$ or $y \leq x$ (i.e., if any two elements of $\boldsymbol{P}$ are comparable).

Again take (3a) for example. It is a chain since every two categories in it are comparable. That hierarchies of projections are chains is usually simply assumed rather than argued for in the linguistics literature, but from a formal-logical perspective we may legitimately ask: Are c-sequences necessarily chains? Insofar as the definition in (2) is concerned, the answer is no. All that a c-sequence needs to be is a complete set of syntactic categories in a single partial order. That the c-sequences in (3) are all chains is merely an observation (based on a small sample) instead of a definitional requirement. For a nonchain c-sequence see (8).

(8) $\{\mathrm{V} \leq \begin{matrix} v_1 \\ v_2 \end{matrix} \leq \mathrm{T} \leq \mathrm{C}\}$

In (8) all categories are in the same partial order, which is nontotal since v_1 and v_2 are incompa-

rable. The well-formedness of (8) suggests that the received assumption that hierarchies of projections are chains is not a hard-and-fast rule. I claim that incomparable categories are precisely the flexibility type at the single c-sequence layer, where the UG template is the combinatorial-scope-induced partial ordering of syntactic categories and a point that UG does not care about is how this ordering is implemented—whether it is plainly partial or furthermore total.

What does the potential existence of incomparable categories mean for the first abstraction layer? Consider what it means for one category to be smaller than or equal to another in featural terms. As aforementioned, the c-sequence ordering of mergeme-bearing categories references their mergeme values, which in turn reflect their first-Merge positions. Thus, a category with the mergeme $\langle$CAT: **v**, POS: 3$\rangle$ is first-merged above two other categories in the same c-sequence (called **v**). If the c-sequence in question is (3a), then the category is simply Asp. More generally, we can represent the entire c-sequence in mergeme terms as in (9).

(9) $\{\langle \text{CAT: } \mathbf{v}, \text{POS: } 1\rangle \leq \langle \text{CAT: } \mathbf{v}, \text{POS: } 2\rangle \leq \langle \text{CAT: } \mathbf{v}, \text{POS: } 3\rangle \leq \langle \text{CAT: } \mathbf{v}, \text{POS: } 4\rangle\}$

So, the partial order on a specific c-sequence is equivalent to the numerical ordering of its members' POS values. Now consider the featural representation of (8):

(10) $\{\langle \text{CAT: } \mathbf{v}, \text{POS: } 1\rangle \leq \begin{matrix} \langle \text{CAT} : \mathbf{v}, \text{POS} : 2\rangle_1 \\ \langle \text{CAT} : \mathbf{v}, \text{POS} : 2\rangle_2 \end{matrix} \leq \langle \text{CAT: } \mathbf{v}, \text{POS: } 3\rangle \leq \langle \text{CAT: } \mathbf{v}, \text{POS: } 4\rangle\}$

Clearly the question is how to maintain the incomparability of the two categories with identical mergeme values.[2] If we solely let the ordering reference the mergeme value, the two identically valued categories would collapse into one in the set-theoretic context. For a potential solution let us recall the mergeme-based definition for syntactic categories as repeated below:

(11) syntactic category$_{\text{LANG, GRAN}}$:= [$\langle$CAT: `cat`, POS: `nat`$\rangle$, NM]

Apart from the mergeme, each syntactic category also has a nonmergeme feature slot. Recall from Chapter 2 that nonmergeme features are also SCDPs and hence can also serve to distinguish one category from another. More specifically, nonmergeme features are formal features irrelevant to combinatorial scoping; they may encode either purely taxonomic properties recruited for category-defining purposes (e.g., [ELEGANT] in a language where register has a syntacticized status) or taxonomically instantiated combinatorial properties (e.g., [EPP]). In a word, two distinct syntactic categories may well share a single mergeme value and only differ in their nonmergeme features. Based on this reasoning, I modify (10) to (12).[3]

(12) $\{\langle \text{CAT: } \mathbf{v}, \text{POS: } 1\rangle \leq \begin{matrix} [\langle \text{CAT} : \mathbf{v}, \text{POS} : 2\rangle, \text{NM}_1] \\ [\langle \text{CAT} : \mathbf{v}, \text{POS} : 2\rangle, \text{NM}_2] \end{matrix} \leq \langle \text{CAT: } \mathbf{v}, \text{POS: } 3\rangle \leq \langle \text{CAT: } \mathbf{v}, \text{POS: } 4\rangle\}$

Now let us reexamine the order-theoretic relation between the two stacked categories (call them

[2]The subscript indices are merely for expository purposes and not part of the SCS ontology.
[3]The other three categories in (12) may also have nonmergeme features. I leave them out to save space.

c_1 and c_2). Suppose c_1 and c_2 are comparable. Since they have identical mergeme values, they are either the same category (i.e., $c_1 = c_2$) or mutually related by $\leq$ (i.e., $c_1 \leq c_2 \wedge c_2 \leq c_1$). But according to the definition of partial order, if $c_1 \leq c_2 \wedge c_2 \leq c_1$ then $c_1 = c_2$ (antisymmetry), so both possibilities point to the same conclusion; namely, c_1 and c_2 are the same category. However, we know that c_1 and c_2 are different categories due to their different nonmergeme features. Therefore, our initial assumption that c_1 and c_2 are comparable must be false (for it leads to a contradiction). This proves that nonmergeme features are indeed an appropriate tool to define and distinguish incomparable categories.

5.1.3 Characteristics of incomparable categories

Above I deduced the potential existence of nonchain c-sequences and incomparable categories. What would incomparable categories look like? I make three predictions about their behavior.

First, since incomparable categories have identical mergeme values, they also have identical combinatorial scopes. This means that no two incomparable categories can co-occur in the same projection line, for they would compete for the same first-Merge position. For example consider a single-projection-line LA based on the c-sequence in (8):

(13) {V, v_1, v_2, T, C, …}

As the derivation proceeds, upon the completion of VP, the next object to be selected and merged is v. However, a contradiction results no matter which v we choose to merge with VP, because once a v is merged in, the other v can only be merged somewhere above it, and the two vs will end up having different combinatorial scopes. But this contradicts the premise that they have identical combinatorial scopes. Hence, the initial selection of v_1 and v_2 into the same single-projection-line LA is inappropriate.[4]

Second, since incomparable categories have identical mergeme values, we expect there to be no c-selectional relation affecting them like that between C and T or D and Q (i.e., no c-sequence-induced selection). This means that if incomparable categories co-occur in different projection lines of the same derivation, there is no rigid requirement on which category occurs in which projection line. Again take (8) for example and consider a multi-projection-line derivation like (14), where a V takes a TP complement.

(14) [… T [… v_m [… V [… T [… v_n [… V …]]]]]],

In such a derivation m and n may freely take values from 1 and 2. Different value assignments will yield different interpretive effects, but they are all syntactically licit. All the syntax requires

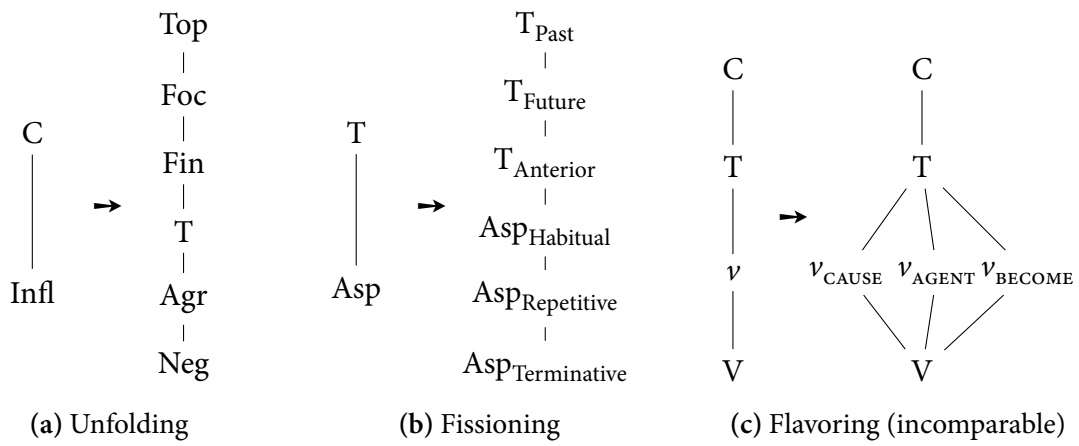
[4]This does not mean incomparable categories can never co-occur in the same LA. The prediction is merely that they cannot co-occur in the same projection line. Thus, if an LA contains multiple projection lines—e.g., a CP embedded in another CP—then it may contain both v_1 and v_2, as long as those are distributed into different CPs.

is that the category merged between T and V is a *v* but not, say, C or D. This lack of c-selectional restriction becomes even clearer if the incomparable categories are V_1 and V_2 and the embedded projection line is a very short one (e.g., containing only V), as in (15).

(15) [… T [… *v* [… V_m [… V_n …]]]]

Here again *m* and *n* may freely take values from 1 and 2. As we will see in §5.2.1, this is exactly what happens in Chinese resultative V-V compounds on a popular analysis.

Third, since incomparable categories have identical mergeme values and mergeme values are associated with ontologically fixed interpretive units, we expect the interpretations of incomparable categories to be highly similar, to the extent that they can be considered variants or subtypes of the same category. In other words, incomparable category formation is essentially category subtyping. However, incomparable categories are different from ordinary subtype categories. To prefigure the content in §5.2, I distinguish three patterns of category subtyping: unfolding (e.g., Infl → Neg, Agr, T), fissioning (e.g., T → $T_{Anterior}$, T_{Future}, T_{Past}), and flavoring (e.g., *v* → v_{CAUSE}, v_{AGENT}, v_{BECOME}).[5] And among the three patterns incomparable categories are formed by flavoring. Unfolding and fissioning put the resulting subtype categories in a hierarchical organization (i.e., a chain), whereas flavoring does not. See Figure 5.1 for a comparative illustration. I will elaborate on the distinctions between the three patterns in §5.2.3.1.



(a) Unfolding (b) Fissioning (c) Flavoring (incomparable)

Figure 5.1 Three patterns of category subtyping

Considering it is a basic tenet of cartography that categories are hierarchically organized, we can describe incomparable categories as "anticartographic." To sum up, incomparable categories have three typical characteristics:

1. They cannot co-occur derivationally in the same projection line.
2. They do not c-selectionally affect each other.
3. They are subtype categories but do not fall in the ordinary pattern of category subtyping.

[5] What I call unfolding and fissioning have been frequently used in generative syntax since the 1990s, especially since the advent of cartography, while flavoring has mainly been used in DM (especially after 2000).

These characteristics make incomparable categories look rather bizarre in the SCS, more bizarre than defective and acategorial categories. My position is that incomparable categories are possible in human language, though they are a deviation rather than a norm due to non-UG factors; see §5.3 for elaboration.

5.2 Category flavoring

In §5.1.3 I claimed that incomparable categories are instantiated by flavored categories. Next I back up this claim by examining flavored little *x*s. I first introduce the proposed flavors for little *x*s in the literature (§5.2.1), then demonstrate that their behavior matches that of incomparable categories (§5.2.2), and finally make two further comments on the place of flavored categories in the SCS (§5.2.3).

5.2.1 Flavored little *x*s

As a most important (or even the only) representative of flavored categories, flavored little *x*s are commonly seen in the DM literature. There are both flavored *v*s (e.g., Marantz 1997, Cuervo 2003, Folli & Harley 2005, 2007, in press, Harley 2009b) and flavored *n*s (e.g., Déprez 2006, Lowenstamm 2008, De Belder 2013).[6] See (16) for examples.

(16) a. Flavored *v*s:[7]
Marantz (1997): v_1 (agent-projecting), v_2 (no agent)
Cuervo (2003): v_{DO} (activity), v_{GO} (change), v_{BE} (state)
Folli & Harley (2005): v_{DO} (agentive), v_{CAUSE} (causative), v_{BECOME} (inchoative)
Chomsky (2013): v (unaccusative/passive), v^* (transitive/unergative)[8]

b. Flavored *n*s:
Déprez (2006): n_{WHAT} (number individuation), n_{WHERE} (spatial individuation)
Lowenstamm (2008): n_{I} (masculine), n_{II} (feminine), n_{III} (neuter), n_{IV} (other)[9]
De Belder (2013): n_{ATOMIC} (for collective mass nouns)

Flavored little *x*s are subtypes of the relevant unflavored categories. For *v* the subtyping is usually based on argument or event structure, and for *n* it is usually based on number or gender. As such,

[6] I have not seen proposals of *a* flavors, but for practitioners the flavoring method may conceivably be extended to other categories; e.g., Folli & Harley (in press) analyze English auxiliaries as v_{AUX}, an AUX flavor of *v*.
[7] DO has different definitions in Cuervo's theory and in Folli & Harley's theory. In my own presentation I consistently use it in Cuervo's sense and adapt Folli & Harley's DO to AGENT when necessary.
[8] The distinction between *v* and v^* is already made in Chomsky (1995), but the explicit recognition of these as "flavors" only happens later; see Chomsky (2013: 44, note 29).
[9] Lowenstamm does not use the term flavor but calls these different "versions" of *n*.

a "flavor" is more precisely a syntacticosemantic property. See (17)–(19) for examples of flavored-*x*-based analyses.

(17) a. [$_{v_{\text{GO}}}$ *Llegaron*] *dos cartas.* [Spanish]
arrived.GO two letters
'Two letters arrived.'

b. [$_{v_{\text{BE}}}$ *Faltan*] *dos velitas en la torta.*
lack.BE two little candles in the cake
'Two birthday candles are missing from the cake.' (adapted from Cuervo 2003: 25)

(18) *ok-i-ru* 'get.up-V.BECOME-TAM; get up' [Japanese]
ok-os-u 'get.up-V.CAUSE-TAM; wake sb. up' (adapted from Harley 2008: 29)

(19) a. [$_{\text{NumP}}$ [$_{\text{Num}}$ *s* [$_{n\text{P}}$ … n_{WHAT} [$_{\text{NP}}$ *book*]]]] [English]

b. [$_{\text{NumP}}$ [$_{\text{Num}}$ ∅ [$_{n\text{P}}$ … n_{WHERE} [$_{\text{NP}}$ *liv*]]]] 'book(s)' [Haitian Creole]
(adapted from Déprez 2006: 76, 78)

In (17) the two Spanish verbs *llegaron* and *faltan* differ in eventuality types, which Cuervo encodes in two *v* flavors. In (18) the two Japanese verbs *okiru* and *okosu* differ in voice types and verbal suffixes, and Harley attributes both to a difference in *v* flavors. In (19) Déprez identifies the difference between English *books* (plural) and Haitian Creole *liv* (number underspecified) as one of *n* flavors or "nominal aspects." Note that Cuervo and Harley understand *v* differently: for Cuervo *v* is just a verbalizer, whose sole semantic function is introducing eventuality, whereas for Harley *v* is both a verbalizer and a voice head. In addition, the ontological status of flavored categories is highly controversial. For instance, opponents of *v* flavors argue that argument/event structural properties come from configurations or roots instead, and that only one type of *v* is needed. See below for two proposals:

> [T]he semantic "flavors" v may adopt arise structurally: i.e., for instance, a little v taking a DP specifier and a DP or root complement is interpreted as DO, while one taking a DP specifier and a Small Clause Result … complement is interpreted as CAUSE. (*the configurational approach*, Acedo-Mattelán & Mateu 2014: 15)

> [T]he grammar does not make reference to annotated v heads, or flavors of v. … [W]e maintain that all v heads are verbalizers. The semantics of the constructions result from the combination of v heads and different types of roots …. In particular, the combination of v with a particular type of root (result) or a small clause gives rise to causative semantics. (*the root-centered approach*, Alexiadou & Lohndal 2017: 101)

Inasmuch as such alternative theories are tenable, flavored *v*s are not a conceptual necessity but

more like an expository device.[10] That said, if little *x* flavoring is just category subtyping, I see no formal-logical ground to ban it from the minimalist system, where category subtyping as a category-forming strategy is anything but illegitimate. So, if flavored categories are to be disfavored, as the above-cited authors suggest, then the reason cannot be from UG but must be from somewhere else. In §5.3 I will argue that it is "third factors" in the sense of Chomsky (2005) that leave flavored categories in disfavor.

5.2.2 Valued categories vs. flavored categories

In §2.4.2 I made a distinction between valued and flavored categories, claiming that they are both pseudo-antidecompositional but are pseudo in different ways. See (20) for examples.

(20) a. Valued categories: $\text{Asp}_{\text{Repetitive}}$, $\text{Asp}_{\text{Habitual}}$, $\text{T}_{\text{Anterior}}$, T_{Past}, $\text{Mod}_{\text{Possibility}}$, $\text{Mod}_{\text{Necessity}}$

b. Flavored categories: v_{DO}, v_{BE}, v_{CAUSE}, v_{BECOME}, n_{WHAT}, n_{WHERE}, $n_{\text{MASCULINE}}$, n_{FEMININE}

The categories in (20a) are taken from the cartographic Infl domain (via Shlonsky 2010: 422), and those in (20b) are repeated from (16). Valued categories—so called because their subscripts are attributive values of their bases viewed as interpretable features (e.g., [ASP: HABITUAL], [T: PAST], [MOD: NECESSITY])—result from fissioning. I have chosen the term *fissioning* because in this pattern of category subtyping, the supertype labels (Asp, T, etc.) are carried over to the subtype categories, and the subtyping process merely splits or *fissions* their attributive values syntagmatically[11] in a way analogous to the splitting of an atom nucleus for energy release or the division of a cell for reproduction, as illustrated in Figure 5.2.

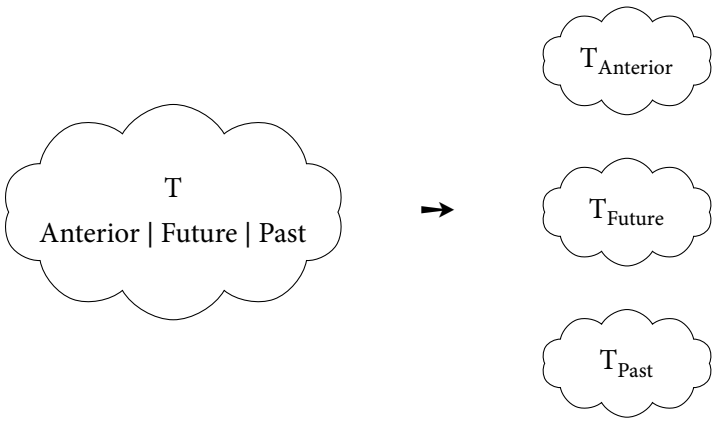


Figure 5.2 A vivid depiction of fissioning

Upon fissioning the supertype label can be safely omitted (e.g., $\text{T}_{\text{Past}} \equiv$ Past), for it merely records the category subtyping history; the omitted supertype can be uniquely deduced from the

[10] Pace Folli & Harley (2007: 210), for whom *v* flavors are not "simply a descriptive convenience" but reflect "structurally distinct primitives of the v inventory."

[11] See §2.2.4 for the difference between syntagmatic and paradigmatic granularities.

subtype labels. By contrast, the base and subscript of a flavored category do not enter any appropriate attribute-value relation (e.g., ?[v: DO], ?[v: CAUSE], ?[n: WHAT], ?[n: MASCULINE]), nor do different flavors of a category have a rigid hierarchical ordering. As such, we cannot safely omit the base labels of flavored categories without causing confusion. That the base-subscript relation in a flavored category is not one of attribute-value is due to the fact that the little *x*s cannot be directly viewed as interpretable features. Rather, their interpretations must be separately defined. Unfortunately, the interpretive definitions of the little *x* categories are highly varied, and consequently different flavors of a little *x* may turn out not to be attributive values of the same grammatical feature at all. Take *v* for example. It is understood as a pure verbalizer in Cuervo (2003) and in works influenced by Cuervo (e.g., Lin 2004, Kan 2007) but as a hybrid head bundling two grammatical functions (verbalizer and voice) in Harley (2008) among other works by Harley and colleagues.[12] Meanwhile, in standard minimalism *v* is neither a verbalizer nor a hybrid head but a Larsonian VP shell, which introduces the external argument at its specifier position and is more recently (Chomsky 2008 et seq.) identified as a phase head.[13]

With this much inconsistency in the definition of *v*, its various flavors ought not to be taken as subtypes of the same supertype category. Thus, the four flavored *v*s in (20b) involve two supertypes: eventuality for $v_{\text{DO}}/v_{\text{BE}}$ and voice for $v_{\text{CAUSE}}/v_{\text{BECOME}}$.[14] Similarly, the three flavored *n*s in (20b) involve three supertypes: `individuation` for n_{WHAT}, `gender` for $n_{\text{MASCULINE}}$, and `number` for n_{ATOMIC}. The *n* flavors appear even less homogeneous; researchers take whatever property they see fit as a flavor to subtype the nominalizer.

On the other hand, that flavored categories have no rigid hierarchical ordering is reflected in their usage. Again take flavored *v*s for example. There are two major camps regarding their usage, which I respectively call the alternation camp and the stacking camp:

(21) a. The alternation camp (e.g., Folli & Harley 2005, 2007, Harley 2008, 2009b):
There is only one *v* position in a single event structure, which may assume one of the flavors CAUSE, DO, BECOME, and BE.

b. The stacking camp (e.g., Miyagawa 1998, Cuervo 2003, Lin 2004, Kan 2007):
There may be multiple *v* positions in a single event structure, which together compose a complex event; for example (> indicates hierarchical positioning):
- Miyagawa (1998): CAUSE > BECOME
- Cuervo (2003): DO > DO/GO/BE, GO > BE

[12] More recently, Harley (2017) proposes a "bundling hypothesis" to crystallize this line of thought.

[13] Marantz (2001) also identifies the verbalizer *v* as a phase head, though the purpose of this phase is word formation, which differs from the purpose of Chomsky's *v* phase. Nevertheless, Chomsky (2013: 43) suggests that the VP-shell *v* may simultaneously "[introduce] a verbal phrase" (i.e., be a verbalizer), which leaves the phase-theoretic status of the "doppelgänger" *v* nebulous. Adding to the nebulousness is another question of whether roots can take complements, for if they cannot—as argued in Alexiadou (2014), Van Craenenbroeck (2014), Alexiadou & Lohndal (2017), inter alia—then the identification of the VP-shell *v* and the verbalizing *v* becomes questionable.

[14] Pace Harley (2009b), who treats causative and inchoative as subtypes of eventuality.

- Lin (2004): δ > BE, DO > δ > BE, BE > DO > δ > BE, BE > δ > BE[15]
- Kan (2007): DO > DO/GO/BE, BE > BE, GO > BE

In the alternation camp, since different *v* flavors compete for the same position, they do not have a hierarchical ordering; in the stacking camp, although flavored *v*s may stack, their hierarchical positioning is quite flexible, and a flavor's hierarchical position does not affect its intrinsic interpretation. Thus, v_{BE} denotes a stative eventuality whether it is above or below v_{DO}. This situation is different from that of the occasional position-flexible categories in cartography; for instance, Top is assumed to be iterable in the C domain (Rizzi 1997 et seq.), but the different Top positions host different types of topic (Benincà & Poletto 2004, Frascarelli & Hinterhölzl 2007). To illustrate the flexible stacking of *v* flavors, consider Kan's (2007) analysis of Chinese resultative compounds; that is, V_1-V_2 constructions with V_1 leading to V_2 as a result. I will critically review this analysis in §5.3, but for now let us focus on the lack of rigid interflavor ordering in Kan's theory. Kan presents the following examples among others:

(22)

Example	**Meaning**	**Structure**	[Mandarin]
dǎ-pò	'hit-broken; break sth.'	v_{DO}-v_{BE}	
tī-jìn	'kick-enter; kick sth. in sw.'	v_{DO}-v_{GO}	
zuì-dǎo	'drunk-fallen; sb. is so drunk that they are fallen'	v_{BE}-v_{BE}	
là-kū	'spicy-cry; sth. is so spicy that sb. cries'	v_{BE}-v_{DO}	

(adapted from Kan 2007: 60–61, 191)

We can easily find more examples to complete the free-stacking picture initiated by Kan:

(23)

hē-tù	'drink-vomit; sb. drinks so much that they vomit'	v_{DO}-v_{DO}
bìng-jìn	'ill-enter; sb. is so ill that they enter sw. (e.g., hospital)'	v_{BE}-v_{GO}
shēng-gāo	'ascend-high; arise'	v_{GO}-v_{BE}
shēng-jìn	'ascend-enter; ascend into sw.'	v_{GO}-v_{GO}
jiàng-kū	'descend-cry; sth. (e.g., salary) lowers so much that sb. cries'	v_{GO}-v_{DO}

The examples in (22)–(23) exhaust all binary combinations of v_{DO}, v_{BE}, and v_{GO} and show that their hierarchical ordering is not rigid. A converging insight is seen in Ramchand (2008: 41–42). Ramchand adopts a different framework, where there is no categorizer or acategorial root, but her decomposition of V into Init–Proc–Res is similarly based on the flexible stacking of two basic eventuality types: stative and dynamic. A stative eventuality merged above a dynamic one is interpreted as an initiating or causing subevent (labeled Init), while a stative eventuality merged below a dynamic one is interpreted as a result subevent (labeled Res).[16] The intrinsic interpretation of the stative eventuality does not covary with its position; the initiating and result readings

[15] Lin's (2004) δ is equivalent to other researchers' GO; it probably stands for *dynamic*.

[16] Ramchand (2008: 42): "[I]t has not been claimed that the causing subevent is a state. [I] assume this position … partly because it gives a simpler ontology, and also because it allows a simpler analysis of stative verbs."

both emerge from the configuration instead. If we notate Ramchand's basic eventuality types as BE (stative) and δ (dynamic), we obtain a picture similar to that in (21b); namely, a complex event may either be composed of BE $>$ δ, δ $>$ BE, or BE $>$ δ $>$ BE.

We can observe a similar lack of rigid hierarchical ordering for flavored *n*s. Déprez's (2006) n_{WHAT} and n_{WHERE} as well as Lowenstamm's (2008) four gender-oriented *n* flavors are obviously in complementary distribution; De Belder's (2013) n_{ATOMIC} presumably cannot co-occur with another *n* in the same nominal projection line either, for a noun is either collective-mass or not. So, previous proposals of flavored *n*s seem to all fall in the alternation camp.

To sum up, flavored categories are unlike valued categories in that the flavors are not attributive values of their bases and in that they do not have a rigid hierarchical ordering. In addition, the lack of rigid ordering implies the lack of c-sequence-based selectional relations between flavored categories. These characteristics match those of incomparable categories listed in §5.1.3. Therefore, flavored categories can be viewed as instantiations of incomparable categories.

5.2.3 Further implications

5.2.3.1 Split-X categories and the limit of cartography

Apart from fissioning and flavoring, there is still a third pattern of category subtyping, which I call *unfolding* (since it is analogous to expanding a folder in a computer file system). Unfolding is partly fissioning-like and partly flavoring-like: as with fissioning, subtype categories that result from unfolding are organized in a chain; as with flavoring, the relation between a "folder category" and its "contents" is not one of attribute and values. Typical folder categories are the coarse categories used in early days of generative syntax, such as Infl and C. Their contents are the split-X categories proposed in the 1980s and 1990s, as exemplified in (24).

(24) a. Pollock (1989): Infl → Neg, Agr, T

b. Rizzi (1997): C → Fin, Foc, Top, Force

The split-Infl/C categories are hierarchically ordered[17] and not attributive values of Infl/C. This is not to say that we cannot put unfolded categories in the [ATT: VAL] format—we can put anything in this format, be it [V: V], [INFL: C], or [❀: ❄], but most such combinations cannot be interpreted by the C-I interface if they are part of some syntactic representation. For example, while one can readily define a conceptually-intentionally grounded value (or value template) for ⟦T⟧, it is difficult to do so for ⟦Infl⟧ without redefining the category as one of its subtypes. For clarity I exemplify valued, flavored, and split-X categories in the [ATT: VAL] format side by side:

[17] Though as Theresa Biberauer (p.c.) points out, there has always been a lot of debate on the ordering.

(25) a. Valued categories: [ASP: REPETITIVE], [ASP: HABITUAL], [T: ANTERIOR], [T: PAST]

b. Flavored categories: [V: DO], [V: CAUSE], [N: MASCULINE], [N: ATOMIC]

c. Split-X categories: [INFL: T], [INFL: AGR], [C: FIN], [C: FOC]

Two observations emerge from (25). First, the attributes of flavored categories resemble those of split-X categories since they have no fixed interpretations, while the values of flavored categories resemble those of valued categories since they correspond to atomic sort-types in some type hierarchy. Second, the values of split-X categories but not those of flavored categories may become attributes of valued categories. The two observations together suggest an overarching pattern of category subtyping consisting of three levels, as illustrated in Figure 5.3.

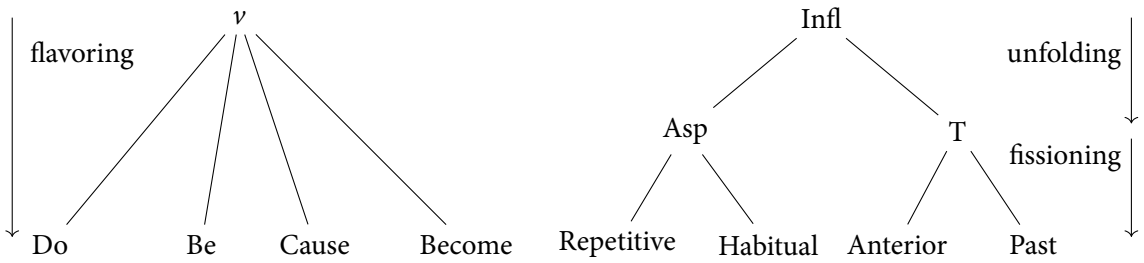


Figure 5.3 Three levels of category subtyping in Chomskyan syntax

So, the bizarreness of flavoring as a pattern of category subtyping is that it skips the intermediate level (i.e., that of the split-X categories). Filling in the skipped level, we get Figure 5.4. I use Event and Voice to notate the eventuality- and voice-introducing split-*v* categories.

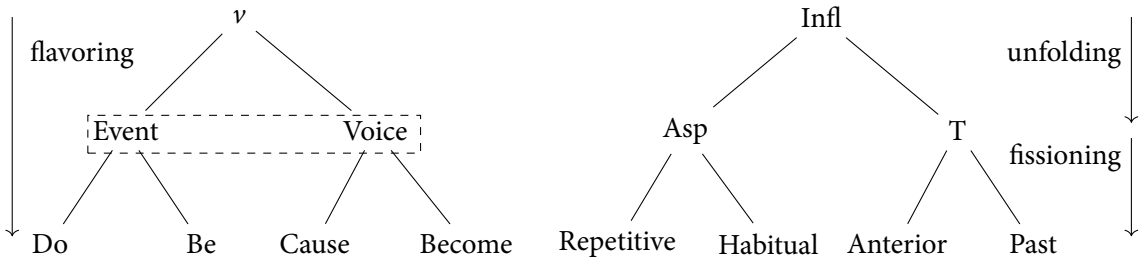


Figure 5.4 Split-*v* categories

Little *x*s are folder categories just like Infl and C.[18] Splitting them stepwise gives us a quasi-

[18]There is a subtle difference between folder categories and Chomskyan core functional categories (CFCs). A folder category is not a derivationally usable primitive (though it may be an ontological primitive, as a "universal spine category"; see Chapter 6) but merely a cover term used by linguists; it is a "domain name," so to speak, and feeding it to a derivation is like committing a type error, for derivations take specific categories as inputs, not domains. By comparison, a CFC is a specific category and hence a legal derivational input. Folder categories and the CFCs they contain are often homonymous; e.g., C is simultaneously a domain and a CFC name. However, we should not confuse the two concepts. An example of nonhomonymous folder vs. CFC names is Infl vs. T. A folder category is also different from a low-granularity category; i.e., a not-yet-subtyped category assumed by a speaker at a certain stage of acquisition. Since a low-granularity category is not yet subtyped, the speaker does not possess knowledge

cartographic[19] expansion of the lexical domain, while skipping the intermediate level deprives the subtype categories of a cartographic organization, which in turn become "flavors" of the folder categories. Similarly, if we omit the split-Infl categories and directly subtype Infl into attributive values, we also get "flavored Infls" as in Figure 5.5.

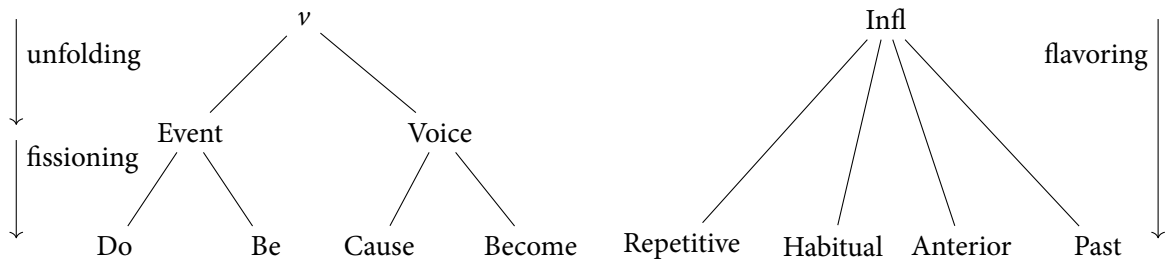


Figure 5.5 Flavored Infls

Flavored *v*s like $v_{\text{DO}}/v_{\text{CAUSE}}$ and flavored Infls like $\text{Infl}_{\text{REP}}/\text{Infl}_{\text{ANT}}$ are alike in that both demand some behind-the-scenes substitution from the reader/linguist in order to yield reasonable interpretations; for example, to make sense of v_{CAUSE} (or Infl_{PST}) we need to understand *v* (resp. Infl) as voice (resp. tense). Such substitutions are needed precisely because *v* and Infl, being folder categories, define collections of grammatical categories rather than individual ones.

The unified pattern of category subtyping has a prima facie striking consequence: the distinction between flavored and valued categories is an artificial one. Valued categories are all hierarchically ordered (i.e., cartographic) not because all attributive values of all grammatical categories are order-theoretically comparable but because cartographers, when designing the cartographic hierarchies, have only chosen the comparable ones. If values like DO/BE and CAUSE/BECOME were given head status—namely, if there were cartographic categories like Event_{Do}, Event_{Be}, $\text{Voice}_{\text{Cause}}$, and $\text{Voice}_{\text{Become}}$—these would manifest the same oddities observed in flavored categories. Most importantly, since incomparable categories have no rigid hierarchical ordering or c-selectional relation, their very existence challenges the foundation of cartography. In order to maintain the cartographic tenet that hierarchies of projections are chains, effort must be made to only "categorify"[20] attributive values (i.e., "features") that do form chains, discarding those that do not. This reveals a limit of the cartographic method, which maps syntactic categories to chains via transitivity. Cinque formulates this method as follows:

> I recapitulate … the evidence for the pairwise order of the different functional heads. As noted, such relative orders give rise, by transitivity, to a single overall order of

of its subtypes-to-be; by contrast, a linguist does possess knowledge about the subtypes of a folder category. Thus, folder categories reflect a low level of *analytical granularity*, whereas low-granularity categories reflect a low level of *mental granularity*. The theory of granularity levels I will develop in Chapter 6 takes both into account.

[19] I say *quasi* because v_{DO}, v_{BECOME}, etc. are anticartographic (see §5.2.2) whether they are viewed as flavored or valued categories. See below for elaboration.

[20] Pending a better term, I use this to mean "make sth. into a category." It should not be confused with *Categorify* (capital *C*), which means "make sth. into a mathematical Category" (see Chapter 6).

heads … which seems to be cross-linguistically consistent. (Cinque 1999: 71)

Let us translate Cinque's formulation into order-theoretic terms. A "pairwise order" is a binary relation on the set of functional heads, and "a single overall order" is a total order. Since Cinque only considers the pairwise order of "different functional heads," this binary relation is irreflexive (i.e., a strict order), though apparently adding a reflexivity condition (whereby the binary relation becomes a weak order) is innocuous. In sum, Cinque's cartographic binary relation is a strict total order. It is precisely this assumption that makes cartography incompatible with flavored categories, hence the following limit.

(26) The limit of cartography:
Not all attributive values correspond to categories in totally ordered c-sequences, or in more conventional terms, not all (atomic) interpretable features can be included in extended projections as heads.

The categorial flexibility at the c-sequence abstraction layer effectively frees cartography from the above limit. All we need to do is modify a stipulation in the cartographic method; namely, that "such relative orders give rise, by transitivity, to a single overall order." This clause is problematic because, mathematically speaking, transitivity does not entail total ordering; transitivity and totality (or linearity) are two separate axioms in order theory. Once the totality restriction is removed, cartographers can freely categorify any attributive values of any grammatical features without worrying that those might not fit in a hierarchical ordering. And under this new setting, valued and flavored categories no longer need to be segregated. Both are just part of the same sort of (high-granularity) c-sequences qua posets.

5.2.3.2 Pseudo-antidecompositionality and the limit of flexibility

Finally, recall from §2.4.2 that the whole issue of flavored categories is contextualized in the topic of pseudo-antidecompositionality (PAD). *Pseudo-antidecompositional* or *PAD categories* are syntactic categories ($\neq$ LIs)[21] that ostensibly have more than one mergeme-based interpretive unit but turn out to be monomergemic under scrutiny. Accordingly, an *antidecompositional category* is a category with a multimergemic specification; namely, a category that has more than one mergeme-typed value in its feature set. (27) displays a bimergemic category.

(27) $^{*}[\langle \textsc{cat}\text{: } \texttt{cat}, \textsc{pos}\text{: } \texttt{nat}\rangle, \langle \textsc{cat}\text{: } \texttt{cat}, \textsc{pos}\text{: } \texttt{nat}\rangle, \textsc{nm}]$

When a mergeme-based category appears to have more than one minimal interpretive unit, it either is not a truly minimal category (but can be decomposed) or has a mixture of mergeme-based and non-mergeme-based interpretive units. The first situation is irrelevant to PAD, because the

[21] I strictly distinguish syntactic categories and LIs in this dissertation; see note 82, p. 37 and §4.1.1.

term is only used to describe single syntactic categories (i.e., proof-theoretic axioms). Derived multimergemic categories more naturally occur in LI-based analyses (e.g., standard minimalism), where the primitives of syntactic derivation are full-fledged LIs. It is common in human language for a single LI to spell out a series of categories.

The second situation (i.e., PAD categories with mixed interpretive units) describes flavored categories. For instance, v_{DO} seemingly encodes two interpretive units, v and DO, but the foregoing examination reveals that the flavor DO, being an incomparable-category-defining property, is nonmergemic. Hence, the correct feature specification of v_{DO} is not (28a), where ⟨CAT: v, POS: 3⟩ is an expository mergeme value for DO, but (28b), or more exactly (28c), considering DO is a value of the split-v category Event (identified with the DM verbalizer) rather than Voice.

(28) a. *[⟨CAT: v, POS: 2⟩, ⟨CAT: v, POS: 3⟩]

b. [⟨CAT: v, POS: 2⟩, DO]

c. [⟨CAT: v, POS: 1⟩, DO]

So, however many flavors are posited for Event, there is only one ontological position for it in the SCS, that picked out by ⟨CAT: v, POS: 1⟩. As a side note, as we update the feature specification for v_{DO} from (28b) to (28c), the ontological c-sequence we assume is also updated from (29a) to, say, (29b). This is essentially a change in analytical granularity level.

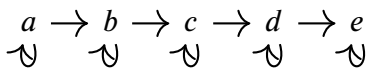
(29) a. {V, v, …} b. {Event, Appl, Voice, …}

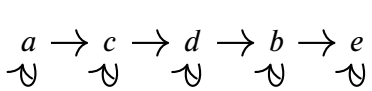
In Chapter 2 I also listed valued categories (e.g., $\text{Asp}_{\text{Habitual}}$, T_{Past}) as PAD categories, though their PAD nature has more to do with notation than feature specification. As mentioned in §5.2.2, the bases of valued categories, such as Asp in $\text{Asp}_{\text{Habitual}}$ and T in T_{Past}, are safely omissible, so the most important type of PAD category is still the flavored category. PAD categories reveal a limit in the kind of categorial flexibility studied in this dissertation; namely, multimergemic simplex categories are a disallowed type of flexibility in human language. This ban is not stipulated but deduced, because by antisymmetry no element in a poset can be simultaneously smaller-than-or-equal-to and greater-than-or-equal-to another element distinct from itself, but allowing one category to have two or more distinct mergeme values precisely leads to this precluded situation. For concreteness, consider a potential c-sequence $F \coloneqq \{a \leq c \leq d \leq e\} \cup \{b\}$, which consists of a chain plus a bimergemic category b. Since b has two mergeme values, it has two combinatorial-scope-induced positions in F. Suppose the two positions are respectively between a and c and between d and e. Is this c-sequence well-defined? Since a c-sequence is essentially a poset, all we need to check is whether F is a poset. Based on the given information about F, we can specify it in ordered pair format (30a) (the two positions of b are underlined) or in arrow format (30b).

(30) a. $\{\underline{\langle a,b\rangle, \langle b,c\rangle}, \langle c,d\rangle, \underline{\langle d,b\rangle, \langle b,e\rangle}, \langle a,c\rangle, \langle a,d\rangle, \langle a,e\rangle, \langle b,d\rangle, \langle b,e\rangle, \langle c,b\rangle, \langle c,e\rangle, \langle d,e\rangle,$
$\langle a,a\rangle, \langle b,b\rangle, \langle c,c\rangle, \langle d,d\rangle, \langle e,e\rangle, \langle d,c\rangle, \langle d,a\rangle, \langle b,a\rangle, \langle c,a\rangle\}$

b.

This binary relation is reflexive and transitive but obviously not antisymmetric; hence, it is not a partial order and therefore not a well-defined c-sequence. In particular, all the nonreflexive loops in (30b) result from the dual positioning of *b*, and once this dual positioning is removed, we should obtain a well-defined c-sequence. There are two possibilities as in (31).

(31) a. 

b. 

In conclusion, the aforementioned limit of categorial flexibility (i.e., the ban on multimergemic categories) is entailed by the definition of c-sequence. Recall that the legitimacy of PAD categories, especially that of flavored categories, is also entailed by the definition of c-sequence. I integrate these two entailments below:

(32) A c-sequence is a poset. ⇒
Flavored/incomparable categories are legitimate. ∧
Multimergemic/antidecompositional categories are illegitimate.

5.3 Revisiting Chinese resultative compounds

A recurring issue in our foregoing discussion of flavored categories is the "conflict" between legitimacy and bizarreness: flavored categories are in principle allowed but in practice bizarre (with regard to projection hierarchy, c-selection, etc.), and this bizarreness is partly due to potential improper application or overuse. What might be conceptually inappropriate with the overuse of categorial flavors? What should we take into account when positing flavored categories as primitives? In the rest of this chapter I tentatively address these questions from a *neo-emergentist* perspective (Biberauer & Roberts 2015, 2017, Biberauer 2016, 2017a, Roberts 2019). Following an introduction of neo-emergentism in relation to category flavoring (§5.3.1), I present a case study on Chinese resultative compounds. I will first critically review some previous analyses (§5.3.2) and then revisit the historical development of the resultative compound construction (§5.3.3). As it turns out, the establishment of the construction indeed has to do with a flavored category (v_{BE}), but the emergence of this flavor is a quite unpredictable incident produced by a series of language-specific conditions (§5.3.4). Besides, the impact of the new verbalizer on Chi-

nese grammar goes far beyond its motivating construction (§5.3.5). I claim that global impact should accompany each new categorizer, flavored or not, and that this can be used as a gauge of adequacy of theories utilizing flavored categorizers.

5.3.1 A neo-emergentist view on category flavoring

Neo-emergentism[22] is the perspective on parametric variation and feature/category formation developed in Biberauer (2011 et seq.), especially in Biberauer (2016, 2017a) (see also Biberauer & Roberts 2015, 2017, Roberts 2019). Its central idea is that parameters are not prespecified in the innate endowment of the language faculty but emerge from the interaction of three factors in the sense of Chomsky (2005); namely, UG, PLD, and general cognitive (or "third") factors. Third factors are "non-domain-specific cognitive optimization principles" such as principles of data analysis, structural architecture, efficient computation, and so forth (Biberauer & Roberts 2017: 114). The core third factor promoted in neo-emergentism is *maximize minimal means* (MMM), conceived as a general cognitive bias:

- **Maximize minimal means** (linguistic manifestation): Postulate as few formal features as possible to account for the intake and maximize use of already postulated features (adapted from Biberauer 2017a: 47–48).

Biberauer (2016: 8) views MMM as a *minimax algorithm*, whose *min* and *max* parts respectively have the more specific reflexes *feature economy* (FE) and *input generalization* (IG) in natural language syntax and perhaps also in phonology. These more domain-specific reflexes are defined as follows (see Biberauer 2017a: 47–48):

- **Feature economy:** Postulate as few formal features as possible to account for the input (= intake) (cf. Roberts & Roussou 2003: 201).

- **Input generalization:** Maximize already-postulated features (cf. Roberts 2007: 275).

The result of the minimax algorithm is an increasingly detailed taxonomy of LIs, with each detected feature defining a natural class and the whole inventory of (ranked) features successively dividing the syntactic categorial space. This is essentially the same effect as that created by Dresher's (2009) successive division algorithm in phonology, which I briefly introduced in §2.2.1.1. I will return to division of the categorial space in Chapter 6.

A key insight from neo-emergentism is that formal features should be posited parsimoniously, only when the PLD cannot be otherwise accounted for. This has a direct consequence for the use of flavored categories in syntactic analyses—inasmuch as flavors are understood as syntactically

[22] Biberauer (2017a: 47) motivates *neo-* as follows: "That grammars [are] characterized in terms of the distribution of formal features and the way … these regulate the operations of Merge and Agree … crucially distinguishes the present approach from 'standard' emergentist approaches, e.g. those in the Construction Grammar tradition."

visible formal features, they should not be just assumed but need to be argued for. Specifically, two questions should be asked at the introduction of each flavored category:

1. Why *must* this flavor be postulated? (i.e., Is it really necessary?)

2. Why *can* this flavor be postulated? (i.e., Is it properly cued by the PLD?)

Under the neo-emergentist view, we need to seriously consider the alternative, non-flavor-based theories each empirical phenomenon may have, such as the configurational and the root-centered approaches to argument/event structure mentioned in §5.2.1. If a flavored theory and a flavorless theory are both empirically adequate—namely, if they both make correct predictions—then the flavorless theory is conceptually more advantageous because it obeys FE. Roberts (2019: 100–101) remarks that the neo-emergentist approach is firmly minimalist in spirit and that it can possibly take us closer to an "intensional" characterization of formal features. Thus, I take neo-emergentism as a promising research paradigm in general and a well-motivated source of guidance on the appropriate usage of formal features in particular.

5.3.2 Previous studies

5.3.2.1 A note on glossing

Recall from §5.2.2 that a resultative compound in Chinese is a two-component verbal compound with a causative-resultative reading. The following examples are repeated from (22)–(23).

(33)	*dǎ-pò*	'hit-broken; break sth.'	[Mandarin]
	tī-jìn	'kick-enter; kick sth. in sw.'	
	hē-tù	'drink-vomit; sb. drinks so much that they vomit'	
	zuì-dǎo	'drunk-fallen; sb. is so drunk that they are fallen'	
	là-kū	'spicy-cry; sth. is so spicy that sb. cries'	

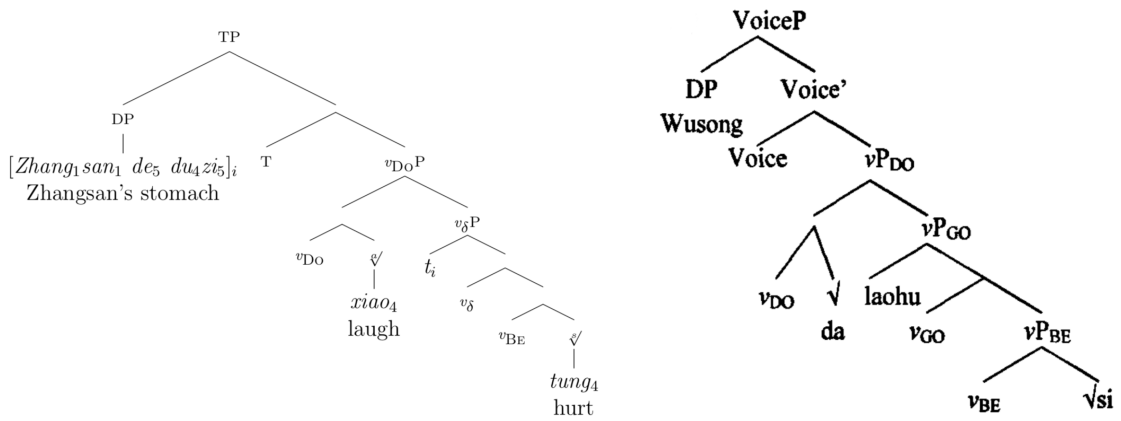
Note that English translations like 'broken' and 'fallen' are only approximate; there is no past participle inflection in Chinese. The result materials are all m-roots and only get interpreted as result states in the syntactic construction, so a more accurate way to gloss 'broken/fallen' is 'be in a broken/fallen state'. By analogy, 'enter', 'vomit', and 'cry' are also more accurately glossed as 'be inside', 'be in the state of vomiting', and 'be in the state of crying', for they, too, can be understood as result states. As such accurate glossing is verbose, I will keep using the shorthand style in (33) unless special clarification is needed.

In §5.2.2 I followed Kan's (2007) suit and described the cause and result materials by their usually associated eventuality types; for example, 'hit-broken' is v_{DO}-v_{BE} or activity-state, 'kick-enter' is v_{DO}-v_{GO} or activity-change, and 'spicy-cry' is v_{BE}-v_{DO} or state-activity. My purpose there was to show that the proposed *v* flavors in the literature had no rigid hierarchical ordering. Now that

our focus has shifted onto the resultative compounds per se, we can untie ourselves from Kan's particular analysis. The fact that anything filling in the result slot can be reasonably understood as a state suggests that perhaps Kan's eventuality-type-based glosses are not accurate after all, for what they describe are merely the eventuality types of the cause/result materials when those are used independently (i.e., outside the resultative compound construction). Do those eventuality types get carried over to resultative compounds? Perhaps not, especially in a DM-based theory like that of Kan, where a verb's eventuality type is determined by its categorizer rather than its root's intrinsic property. Thus, a more DM-y practice is reassessing each vocabulary item's interpretation in each new grammatical context it occurs in rather than letting one interpretation carry over across various contexts.

5.3.2.2 The dummy verbalizer pitfall

Lin (2004) and Kan (2007) are two representative DM-based studies of Chinese resultative compounds, both utilizing stacked little *v*s, as illustrated in Figure 5.6. The three *v* flavors Lin and Kan assume are basically the same as those in Cuervo (2003); namely, v_{DO} (activity), v_{BE} (state), and v_{GO} (change) (Lin uses δ to notate GO).



(a) Lin (2004: 134) (b) Kan (2007: 243)

Figure 5.6 Two DM analyses of Chinese resultative compounds

There is a common oddity in theories that stack flavored *v*s: *v* is consistently identified as the DM verbalizer, but it does not consistently do the verbalizing job. That the stacked *v*s are all taken to be verbalizers is reflected in the following quotes (emphases mine):

> I will assume … that verbs are formed in the syntax by the combination of a lexical root and a *verbalizing* head little *v*. (Cuervo 2003: 17)
>
> I take "verbs" as abstract categoryless concepts (verbal roots) that gain their verbal categorial status by association with a *verbalizing* head …. Verbal roots serve

> as event modifiers that elaborate on the basic eventive readings introduced by the *verbalizing* heads. I posit … three such elements: [v_{DO}, v_{δ}, and v_{BE}.] (Lin 2004: 32)
>
> I assume that verbs are formed in the syntax by the combination of a lexical root and a *verbalizing* head little *v*. … I will assume … three types of little *v* that correspond to three distinct basic types of simple events. (Kan 2007: 59)

That these authors' verbalizers do not consistently verbalize is reflected in Figure 5.6. Neither Lin's v_{δ} nor Kan's v_{GO} merges with a root that needs categorization. Namely, both are what I will call *idle* or *dummy verbalizers*. But if they do not verbalize, why call them verbalizers? This creates inconsistency in the theory. In my opinion, the nub of the matter is a confusion of verbalizer and VP-shell categories (see Myler 2016 for a similar view). Both may participate in event structure regulation, but in DM the defining property of the former is its root-categorizing role; if some category/head merely regulates the event structure but does not categorize any root, then strictly speaking it is a VP shell instead of a verbalizer. The confusion is a typical pitfall in the usage of flavored *v*s and presumably has to do with the all-purpose notation *v*, which in the terminology of §5.2.3.1 is a folder category.

To clarify, I do not object to the methodology of syntactic event decomposition per se; what I object to is merely its DM incarnation in the guise of stacked "verbalizers," which runs afoul of the very definition of verbalizer in DM. By comparison, there is no such problem in works like Hale & Keyser (2002) and Ramchand (2008), which also make use of event decomposition but do not use verbalizers. Accordingly, accounts of Chinese resultative compounds built on these works (e.g., Huang 2006, built on Hale & Keyser 2002) are free from the dummy verbalizer pitfall. Huang (2006) proposes the following template for resultative compounds, where the optional CAUSE and the obligatory BECOME/STATE are light verbs, to which lexical verbs may adjoin via a conflation process.

(34) [(CAUSE) [BECOME [STATE]]],

Kan's example in Figure 5.6b can be translated into Huang's theory as in (35). Huang assumes that the two lexical verbs *dǎ* 'hit' and *sǐ* 'die' form a compound via head movement.

(35) [$_{\text{CauseP}}$ *wǔsōng* [$_{\text{Cause}}$ *dǎ*-CAUSE] [$_{\text{BecomeP}}$ BECOME
Wusong hit-CAUSE BECOME

[$_{\text{StateP}}$ *lǎohǔ* [$_{\text{State}}$ *sǐ*-STATE]]]]
tiger die-STATE

'Wusong hit the tiger to death (lit. hit-dead the tiger).'

Since there are only two lexical verbs in a resultative compound, one of the three light verbs in (34) (either CAUSE or BECOME) remains null. Nevertheless, Huang's light verbs are not DM verbalizers, and so their null status does not cause any definitional problem.

5.3.2.3 Is the result slot adjectival?

There is a revived debate on the categorial status of the result slot in resultative compounds, which goes hand in hand with another long-standing controversy; namely, whether adjective is a separate syntactic category in Chinese at all. A widely accepted position since the 1960s (e.g., Chao 1968, Dixon 1977, Li & Thompson 1981, McCawley 1992) is that adjectives in other languages are verbs in Chinese, but some more recent works like Huang et al. (2009) and especially Paul (2015) argue for a separate adjective category. As a thorough investigation of adjectives will take us too far afield and the controversy is unlikely to be settled in just a few paragraphs, I will not go into the issue any deeper but simply acknowledge the controversy. For my purpose here what really matters is not the general categorial status of adjectives in the language but the categorial status of the result slot in a resultative compound; even if adjectives do form a separate category, it still cannot be concluded from there that the result slot must be adjectival.

Conventionally, both the cause and the result slots in resultative compounds are considered verbal. This is clearly reflected in handbook articles; for instance, Williams (2014: 311) states that "[causal VVs] imply a causal relation between two distinct events, those of the first and second *verb*" (emphasis mine, same below), and Tham (2015: 307) states that "a resultative verb compound comprises two *verbal* elements." Huang et al. (2009: 40)—though they endorse a separate adjectival category for Chinese—also state in the section on resultative compounds that "[t]he two *verbal* morphemes in each compound are in a causal relation."

Two works that explicitly treat the result slot as adjectival rather than verbal are Paul (2015) and Hu (2018).[23] Paul does not provide justification for this treatment but merely situates it in an overall argumentation for the independent categorial status of adjectives. Hu does provide several arguments for his position, which I review below. While I disagree with Hu on the adjectival status of the result slot, I do share his idea that the result slot is adjective-like; namely, that it is associated with a simple stative interpretation (i.e., a `state`-typed first-order property). Unlike Hu, however, I hold that this interpretation is syntactically carried by a (stative) verbal category instead of an adjectival one.

Hu divides result slot items (henceforth RSIs) into three types: (i) canonical adjectives (in the sense that their counterparts are adjectives in English-like languages), (ii) unaccusative-verb-like items, and (iii) unergative-verb-like items. See (36) for examples (adapted from Hu 2018: 105).

(36) a. Canonical adjectives: *píng* 'flat', *dà* 'big', *xiǎo* 'small', *kuān* 'wide', *zhǎi* 'narrow', *shī* 'wet', *gān* 'dry', *lèi* 'tired', *yǎ* 'hoarse', *hóng* 'red'

b. Unaccusative-verb-like items: *pò* 'broken', *wān* 'bent', *kāi* 'boiled', *shú* 'cooked, ripe'

c. Unergative-verb-like items: *kū* 'cry', *xiào* 'laugh'

[23] Chao (1968) also identifies the result slot as adjectival, but for Chao adjectives as a whole are a subclass of verbs.

Translating Hu's classification into that of Kan (2007), canonical adjectives are stative verbs (v_{BE}),[24] unaccusative-verb-like items are change-of-state (CoS) verbs (v_{GO}), and unergative-verb-like items are activity verbs (v_{DO}). Hu's separation of type-(i) and type-(ii) RSIs is mainly based on their English translations; namely, type-(ii) items are translated as past participles. In Chinese, however, the two classes are nondistinct either morphologically (there is no past participle inflection) or semantically (they all denote simple states), and syntactically all three types of RSIs can be naturally used in verbal contexts; for example, they can all be followed by the aspectual particle *le*[25] as in (37).

(37) a. *háizi yòu dà le yí suì* [Mandarin]
child again big ASP one year
'The child has grown another year older (lit. bigged another year).'

b. *shuǐ kāi le hǎn wǒ yì shēng*
water boiled ASP shout me one CL
'Give me a shout when the water is ready (lit. has become boiled).'

c. *báobao kū le bàn tiān, māma què bù lǐ tā*
baby cry ASP half day mom but not pay attention him
'The baby cried for a long time, but the mom ignored it.'

Hu's classification is incomplete, as it excludes non-adjective-like stative verbs (e.g., *dǒng* 'understand', *hòuhuǐ* 'regret', *pà* 'fear', *dānxīn* 'worry', *shuì* 'sleep', *xiàng* 'resemble', *xìn* 'believe') and change-of-location (CoL) verbs (e.g., *pǎo* 'run', *zǒu* 'walk', *jìn* 'enter', *chū* 'exit', *shēng* 'rise', *luò* 'drop'), which clearly can also be used in the result slot, as in (38)–(39).

(38) a. *wǒ tīng-dǒng le* (stative result)
I listen-understand ASP
'I get it (lit. have understood upon listening).'

b. *máma bǎ háizi hǒng-shuì le*
mom DISP child coax-sleep ASP
'The mother has put (lit. coaxed) the child to sleep.'

c. *běnlái bú xiàng, bèi nǐ gěi shuō-xiàng le*
originally not resemble PASS you give talk-resemble ASP
'They didn't look similar at first, but your words made them look similar.'

[24] So *píng*, *dà*, etc. should more accurately be glossed as 'be flat', 'be big', etc.
[25] In this section I gloss over the concrete functions of *le* to avoid unnecessary complexity. See Song (2018b) for an introduction of the multiple aspectual functions of *le* in two Mandarin varieties.

(39) a. *nǎinai bǎ è-gǒu dǎ-pǎo le* (change-of-location result)
grandma DISP ferocious-dog hit-run ASP
'The grandma hit the ferocious dog and (as a result) it ran away.'

b. *qiú-yuán bǎ qiú tī-jìn le zì-jiā qiú-mén*
ball-member DISP ball kick-enter ASP self-family ball-door
'The soccer player kicked the ball into the goal of his own team.'

c. *língling bǎ xiāng-cài dǒu-luò le jǐ gēn*
Lingling DISP fragrant-vegetable shake-drop ASP several stalk
'Lingling shook her hand and (as a result) a few stalks of cilantro dropped.'

Besides, contrary to Hu's claim that there are very few type-(iii) items, we can easily find more examples (e.g., *jiàohuan* 'shout', *tiào* 'jump', *chōu* 'twitch', *duōsuo* 'tremble', *tù* 'vomit', *wàng* 'forget', *niào* 'pee', *lā* 'poo', *zhāo* 'confess'), as illustrated in (40).

(40) a. *wǒ yòng kuàizi bǎ tā dǎ-jiàohuan le*
I use chopsticks DISP him hit-shout ASP
'I hit him so hard with chopsticks that he began shouting.'

b. *chúle kāitóu jǐ jù, hòumiàn quán xià-wàng le*
except beginning several sentence later all scare-forget ASP
'Except for the beginning few lines, I forgot all the remaining part out of fear.'

c. *jīntiān shuāi le yì jiāo, téng-niào le*
today tumble ASP one stumble hurt-pee ASP
'I tumbled today and (almost) peed out of pain.'

d. *yǒude fànrén hái méi wèn huà jiù yǐjīng xià-zhāo le*
some criminal still not have ask words just already scare-confess ASP
'Some criminals confessed out of fear even before the interrogation started.'

In Table 5.2 I expand Hu's classification of RSIs into five types. Following Hu, this expanded classification is based on the categories of the RSIs' English counterparts.

Since three out of the five types in Table 5.2 are verbs in English, and all RSIs (even the canonical-adjective ones) show verbal behavior (e.g., taking aspectual elements) in Chinese, we can legitimately ask, Why can't they just be treated as verbal? Hu gives three main arguments for an adjectival treatment, but none of them is sufficient. First, Hu observes that RSIs all have attributive and predicative uses and that the canonical-adjective ones can further be modified by *hěn* 'very', and he takes these as typical adjectival characteristics. However, a variety of categories (e.g., nouns, verbs) can be used attributively in Chinese (41), and stative verbs can be modified by *hěn* as well (42). And obviously all verbs have predicative usage.

Table 5.2 An expanded classification of result slot items

RSI type	English counterpart	Example (cause-result)
Canonical adjective	adjective	*yā-píng* 'press-flat'
Stative-verb-like item	stative verb	*shuō-xiàng* 'talk-resemble'
CoS-unaccusative-verb-like item	past participle	*dǎ-pò* 'hit-broken'
CoL-unaccusative-verb-like item	change-of-location verb	*tī-jìn* 'kick-enter'
Unergative-verb-like item	unergative verb	*là-kū* 'spicy-cry'

(41) *shuǐ*$_{\text{N}}$*-lù* 'water-road; waterway'
bǎiyóu$_{\text{N}}$*-lù* 'asphalt-road; blacktop'
jiāochā$_{\text{V}}$*-lù* 'intersect-road; crossroad'
[*qiú-zhī*]$_{\text{VP}}$*-lù* 'seek-knowledge-road; the road of knowledge seeking'

(42) a. *tā hěn hòuhuǐ nà jiàn shì*
he very regret that CL matter
'He regrets a lot over that matter.'

b. *tā hěn xǐhuān tián-shí*
he very like sweet-food
'He likes sweet food a lot.'

Since *hòuhuǐ* 'regret' and *xǐhuān* 'like' are clearly verbal, the fact that they can be modified by *hěn* is evidence that the possibility of degree modification does not guarantee an adjectival category.

Second, Hu argues that unaccusative-verb-like RSIs are adjectival because they entail no prior events, which makes them resemble primitive rather than derived adjectives in English; namely, like *flat* in *flat metal* and unlike *broken* in *broken vase*. Hu gives the example in (43a). However, adjectives are not the only class of property-denoting items with no prior event entailment. Stative verbs and nouns in the same construction entail no prior events either, as in (43b–c).

(43) a. *zhè tiáo niúzǎikù shèjì hěn tèbié, shì pò de*
this CL jeans design very special COP broken MODI
'The design of the jeans is very special: they are broken (e.g., have holes in them).'
(adapted from Hu 2018: 107)

b. *tā zuǐ shàng shuō nán-kàn, xīn lǐ què shì xǐhuān de*
he mouth on say difficult-looking heart in but COP like MODI
'He says it's ugly but secretly likes it.'

c. *zhè tiáo qúnzi shì sīchóu de*
this CL dress COP silk MODI
'This dress is made of silk.'

The fact that *pò* 'broken', *xǐhuān* 'like', and *sīchóu* 'silk' all denote `state`-typed first-order properties and all lack prior event entailment reveals that there is no necessary correlation between prior event entailment and syntactic categorial status. Hu gives two additional arguments for the adjectival status of unaccusative-verb-like RSIs: (i) they are incompatible with the passive marker *bèi*, as in (44a); (ii) they cannot be causativized, as in (44b–c).

(44) a. *yì bēi (*bèi) kāi shuǐ*
one glass PASS boiled water
'a glass of boiled water'

b. *yí gè huāpíng pò le*
one CL vase broken ASP
'A vase broke.'

c. **zhāngsān pò le yí gè huāpíng*
Zhangsan break ASP one CL vase
Intended: 'Zhangsan broke a vase.' (adapted from Hu 2018: 107–108)

However, incompatibility with *bèi* can only prove that the modifiers in question are not derived from passivization but cannot further prove that they are adjectives. Stative-verb-turned modifiers are also incompatible with *bèi*, as in (45).

(45) *yì sī (*bèi) pà yì*
one thread PASS fear sentiment
'a hint of fear'

And the lack of causative usage as an empirical observation is questionable. For example, while *pò* cannot be causativized in (44c), it can in (46b) and (47b). Whatever the difference between breaking a vase and breaking a dam / cracking a password is, the data suggest that the ill-formedness of (43b) is not simply a matter of category (or syntax).

(46) a. *nóngchǎng dìduàn de dī pò le*
farm area POSS dam broken ASP
'The dam at the farm area broke.'

b. *dà hóngshuǐ pò le dī, yānmò le shì-qū*
big flood break ASP dam inundate ASP urban-district
'The deluge broke the dam and inundated the urban area.' (from Baidu)

(47) a. *tàiyáng de hòuyì dōu wánjié le, nǐ de Office mìmǎ pò le ma*
sun POSS descendant all finish ASP you POSS Office password broken ASP SFP
'Descendants of the Sun has finished; has your Office password been cracked?'

b. *wǒ jūrán chénggōngde pò le mìmǎ*
I unexpectedly successfully break ASP password
'I can't believe I have successfully cracked the password.' (from Baidu)

Third, Hu argues that unergative-verb-like RSIs are independently verbal but are coerced into adjectives when used as RSIs, because they can be modified by volitional adverbials when used alone (48a) but not when used as RSIs (48b).

(48) a. *zhāngsān gùyì kū le*
Zhangsan intentionally cry ASP
'Zhangsan cried intentionally.'

b. *zhāngsān gùyì mà-kū le lǐsì*
Zhangsan intentionally criticize-cry ASP Lisi
'Zhangsan intentionally made Lisi cry by criticizing him.'
* 'Zhangsan made Lisi intentionally cry by criticizing him.'
(adapted from Hu 2018: 109–111)

However, a closer look reveals that the result slot cannot be modified by any adverbial, as in (49), whatever its semantic nature.

(49) a. *zhāngsān bùxiǎoxīn mà-kū le lǐsì*
Zhangsan accidentally criticize-cry ASP Lisi
'Zhangsan accidentally made Lisi cry by criticizing him.'
* 'Zhangsan made Lisi accidentally cry by criticizing him.'

b. *zhāngsān dàshēngde mà-kū le lǐsì*
Zhangsan loudly criticize-cry ASP Lisi
'Zhangsan made Lisi cry by loudly criticizing him.'
* 'Zhangsan made Lisi loudly cry by criticizing him.'

As we can see, not only the volitional adverb *gùyì* 'intentionally' but also the nonvolitional adverb *bùxiǎoxīn* 'accidentally' and the manner adverb *dàshēngde* 'loudly' can only be interpreted as modifiers of the cause slot item *mà* 'criticize, scold'. This suggests that perhaps structural separation is a more plausible explanation for the modification failure in (48d).

In sum, Hu's arguments are insufficient to prove that RSIs are adjectives. All his observations are also (and in fact more) compatible with a stative verb category; namely, a vocabulary item recruited for the result slot, whatever its original category, is "coerced" into a stative reading. Below I exemplify this for the five types of SRI in my expanded classification. As the third column shows, the interpretation of the result slot follows a templatic pattern; namely, 'be in a … state' or 'be in the state of … '.

Table 5.3 Readings of result slot items

RSI type	Example	Reading
"Adjective"	*yā-píng* 'press-flat'	'press sth. so that it ends up **being in a** flat **state**'
Stative	*shuō-xiàng* 'talk-resemble'	'talk so that sth. ends up **being in the state of** resembling sth.'
CoS unacc.	*dǎ-pò* 'hit-broken'	'hit sth. so that it ends up **being in a** broken **state**'
CoL unacc.	*tī-jìn* 'kick-enter'	'kick sth. so that it ends up **being in the state of** inside sw.'
Unergative	*là-kū* 'spicy-cry'	'sth. is so spicy that sb. ends up **being in the state of** crying'

The verbal rather than adjectival category of the result slot in Chinese resultative compounds is a crucial assumption in my case study; in a sense it *is* the case. Now that I have justified this assumption, in the upcoming sections I will simply hold it as a premise and point of departure. Everything else is built on it.

5.3.2.4 Interim summary

In §5.3.1 I argued for a parsimonious employment of flavored categories on a conceptual (i.e., neo-emergentist) basis; in §5.3.2.2 I further noticed a theoretical pitfall caused by the unrestricted use of flavored verbalizers (i.e., it creates inconsistency in DM). Based on these two points, an extreme option is not to use flavored verbalizers at all. However, it is clear from §5.3.2.3 that RSIs do acquire a special categorial status in the resultative compound construction: they acquire (sometimes ad hoc) stative denotations by virtue of being in the result slot. In DM this situation can be explained by endowing the result slot with a flavored verbal category v_{BE}. I formulate this hypothesis as follows:

(50) The result slot in Chinese resultative compounds has the syntactic category v_{BE}.

Two clarifications are in order. First, v_{BE} is associated with the result slot, not the materials recruited to fill the slot. In the terminology of Chapter 4, what we have is a root-support scenario {X, ω}, where the abstract syntactic category X is v_{BE} (or V_{BE})[26] and the recruited material ω can

[26] I keep using the little v notation in this chapter because it is the standard notation for the DM verbalizer; in my own SCS framework I prefer using the big V for this purpose (see Merchant 2017 for a similar practice) and reserve v for the folder category and/or the CFC (see §5.2.3.1).

be of any type in Table 5.3. This allows us to categorially tease apart RSIs and the result slot and maintain a unified category for the latter without having to impose a unified category on former.

Second, the postulation of v_{BE} does not have to be accompanied by that of other flavored verbalizers (e.g., v_{DO}, v_{GO}). This is a major difference between my conception of verbalizer flavoring and the conception in Cuervo (2003), Lin (2004), and Kan (2007). For those authors v flavoring is for the purpose of event decomposition, but for me it is merely for the purpose of syntactic categorization. I assume that event (de)composition can remain semantic when there is no evidence that it affects syntactic categories; thus, there is no problem for a verb to have a DO eventuality type and a generic, nonflavored syntactic category. In short, I assume that in Mandarin v_{BE} is the only flavored verbalizer in addition to a generic v and that the two are distinguished by a single (nonmergeme) boolean feature [±BE]. My goal in the next section is to investigate how this flavor has emerged in history.

5.3.3 The history of resultative compounds

Albeit fully productive in Modern Chinese, the resultative compound construction came into being quite late in history, only during the Middle Chinese period. Grammarians cannot agree on an exact beginning time for it, but the construction is already in stable use by the twelfth century (Shi 2003: 44). Table 5.4 is a historical periodization of Chinese (based on Wang 2004: 43–44).

Table 5.4 Periodization of Chinese

Period	Time
Old Chinese (OC)	before 3rd century
Middle Chinese (MidC)	4th–12th century
Early Modern Chinese (EMC)	13th–19th century
Modern Chinese (ModC)	20th century–present

The establishment of the resultative compound construction is considered a big event in the history of Chinese. The introduction I give in this section is based on Shi (2002, 2003) and Liu (2000). Both authors hold that resultative compounds developed from juxtaposed clauses via increasing erosion of the clause boundary. I schematize the process in Table 5.5, indicating each point of reanalysis by a blue line.

Table 5.5 Historical development of the resultative compound construction

Step	Schema (S = Subject, O = Object)	Description
1	$[_{\text{Clause}_1}\ V_1\ O_i\ ]\ [_{\text{Clause}_2}\ S_i\ V_2\ ]$	separate clauses
2	$[_{\text{Clause}_1}\ V_1\ O_i\ ]\ [_{\text{Clause}_2}\ \text{pro}_i\ V_2\ ]$	second clause with null subject
3	$[_{\text{Clause}}\ V_1\ [_{\text{Small Clause}}\ \text{O/S}\ V_2\ ]]$	single clause with small clause
4	$[_{\text{Clause}}\ V_1\ [_{\text{Small Clause}}\ \text{pro}\ V_2\ ]]$	small clause with null subject
5	$[_{\text{Clause}}\ V_1\text{-}V_2\ O\ ]$	resultative compound

Shi (2002: 59–61) illustrates the process by tracing the development of a number of RSIs (*chéng* 'completed, accomplished', *liǎo* 'finished, ended', *suì* 'smashed, shattered', etc.). Here I only cite one of them, *chéng*. To begin with, in step one *chéng* qua V_2 is in a separate clause, whose subject corefers with the object of the preceding clause, as in (51).[27]

(51) a. 支道林造卽色論，論成，示王中郎。 [MidC, 5c.]

$[_{\text{Clause}_1}\ [_S$ *zhī dàolín* $[_{V_1}$ *zào* $]\ [_O$ *jísè-lùn*$_i$, $]]$
Zhi Daolin compile Jise-theory

$[_{\text{Clause}_2}\ [_S$ *lùn*$_i$ $]\ [_{V_2}$ *chéng,* $]]$ *shì wáng zhōngláng*
theory complete show Wang Palace Guard

'Zhi Daolin compiled the Jise (lit. matter-as-such) theory. When the theory was completed, he showed it to Palace Guard Wang.'

b. 求紙筆作書，書成，救者多，乃得免。

$[_{\text{Clause}_1}$ *qiú zhǐ bǐ* $[_{V_1}$ *zuò* $]\ [_O$ *shū*$_i$, $]]$
request paper pen compose letter

$[_{\text{Clause}_2}\ [_S$ *shū*$_i$ $]\ [_{V_2}$ *chéng,* $]]$ *jiù zhě duō, nǎi dé miǎn*
letter complete save those who many so get exempt

'He (a prisoner) requested paper and pen to compose a letter. When the letter was completed (and sent out), many people came to his aid, so he got exempted.'

(SSXY, adapted from Shi 2002: 59)[28]

The two examples in (51) both consist of separate clauses, and in both examples the event in Clause_1 naturally leads to that in Clause_2; thus, Zhi Daolin's compiling leads to the completion of the Jise theory, and the prisoner's composing leads to the completion of the letter. Further-

[27] Since historical Chinese documents are written in ideographic characters, whose phonological reconstructions are often controversial, I simply present the examples in Pinyin (the official romanization system for Mandarin) according to their current pronunciation. I also provide the original characters.

[28] For expository convenience I give document sources in abbreviations; see Appendix B for full titles. Besides, I have provided more contexts in the data (here and below) than that provided by Shi and adopted the versions in the original documents where they differ from Shi's presentation.

more, in neither example is this cause-and-effect relation explicitly expressed; it is implied by the temporal sequencing of the events instead. Liu (2000) describes this phenomenon as a case of *metaphorical extension* following Traugott & König (1991) (see also Claudi & Heine 1986); namely, the semantic extension from temporal sequencing to cause-and-effect sequencing.

Next, in step two Clause_2 has a null subject, whose dropping is pragmatically conditioned. Since such "radical pro-drop" is a preexisting phenomenon in OC, the separation of steps one and two is merely taxonomic rather than chronological. This is reflected in historical data, where the two structures can be found in the same document. (52) is an example of step two.

(52) 王中郎令伏玄度、習鑿齒論青、楚人物。臨成，以示韓康伯。 [MidC, 5c.]

[$_{\text{Clause}_1}$ *wáng zhōngláng lìng fú xuándù, xí zuòchǐ* [$_{V_1}$ *lùn*]
Wang Palace Guard order Fu Xuandu Xi Zuochi discuss

[$_O$ *qīng, chǔ rénwù*$_i$.]] [$_{\text{Clause}_2}$ [$_S$ pro$_i$] *lín* [$_{V_2}$ *chéng,*]]
Qing Chu figure nearly complete

yǐ shì hán kāngbó
take show Han Kangbo

'Palace Guard Wang made Fu Xuandu and Xi Zuochi write essays on figures from Qing and Chu. The essays were nearly complete, and Wang showed them to Han Kangbo.'

(SSXY, adapted from Shi 2002: 59)

According to Shi, *lùn* 'discuss' and *chéng* 'be completed' in (52) are still in separate clauses, as the latter is modified by an adverbial *lín* 'just about' and has its own argument structure (or more exactly topic-comment structure). V_2 in step two may also be preceded by negative particles, such as *wèi* 'not have' in (53).

(53) 吾久欲注，尚未了。 [MidC, 5c.]

[$_{\text{Clause}_1}$ *wú jiǔ yù* [$_{V_1}$ *zhù*] [$_O$ pro,]
I long want annotate

[$_{\text{Clause}_2}$ [$_S$ pro] *shàng wèi* [$_{V_2}$ *liǎo.*]]
still not have complete

'I have long wanted to annotate (it), but (the annotation) has not yet been completed.'

(SSXY, adapted from Shi 2002: 62)

When no extra elements (e.g., adverbials, negative particles) intervene, V_1 and V_2 end up being separated by a mere nominal argument, which is simultaneously the object of V_1 and the subject of V_2; it qualifies as a "pivotal" argument in Li & Thompson's (1981: 607) terms. This pivotal configuration fostered the reanalysis from step two to step three, where the two clauses are collapsed into one, with the second clause becoming a small clause complement of V_1. Apparently

this reanalysis is already established by the fifth century, as the structure in step three is attested regularly in the same document as that in (51) and (52). See (54) for an example.

(54) 袁伯彥作名士傳成，見謝公。 [MidC, 5c.]

[$_{Clause}$ *yuán bóyàn* [$_{V_1}$ *zuò*] [$_{O/S}$ *míngshìzhuàn*] [$_{V_2}$ *chéng,*]]
Yuan Boyan compose Mingshizhuan complete

jiàn xiè gōng
see Xie Duke

'Yuan Boyan wrote *Mingshizhuan* (lit. Biography of Celebrities) up and met Duke Xie.'
(SSXY, adapted from Shi 2002: 60)

When the pivotal argument is pragmatically dropped, V_1 and V_2 become adjacent as in step four, which is the immediate predecessor of the resultative compound construction. Shi does not exemplify this with *chéng* but Liu (2000) gives an example from early MidC, as in (55).

(55) 佛聽比丘……捉婦女出……爾時比丘即便捉出。 [early MidC, late 3c.]

[$_{Clause}$ *fó tīng bǐqiū* … [$_{V_1}$ *zhuō*] [$_{O/S}$ *fùnǚ*$_i$] [$_{V_2}$ *chū*]] …
Buddha allow Bhiksu grab woman exit

[$_{Clause}$ *ěr-shí bǐqiū jíbiàn* [$_{V_1}$ *zhuō*] [$_{O/S}$ pro$_i$] [$_{V_2}$ *chū*]]
that-time Bhiksu immediately grab exit

'The Buddha allowed the Bhiksu to grab the woman out … At that time the Bhiksu immediately grabbed her out.' (AYWZ, adapted from Liu 2000: 254)

In (55) *zhuō* … *chū* 'grab sth./sb. out' occurs twice, once with an overt pivotal argument *fùnǚ* 'the woman' and once with a pro. The final step is the reanalysis of the null-pivotal-argument construction to the resultative compound construction, with the small clause structure being eliminated. Once V_1-V_2 becomes a syntactic unit, it acquires the ability to take a direct object as a whole. Shi exemplifies this final stage for *chéng* with the following example.

(56) 行至路邊，忽然染患，壘成方丈。 [MidC, 9c.]

xíng zhì lù biān, hūrán rǎn huàn, [$_{V_1\text{-}V_2}$ *lěi-chéng*] [$_{O}$ *fāngzhàng*]
walk to road side suddenly catch disease pile-complete abbot's chamber

'They walked to the road side, suddenly caught a disease, and piled up an abbot's chamber.'
(DHBW, adapted from Shi 2002: 61)

As Shi observes, the RSI usage of *chéng* appeared much (ca. three centuries) later than its small clause complement usage. This means that the reanalysis from step four to step five is not as easy as that from step two to step three. Shi attributes this to the existence of a competing analysis for immediately adjacent verbs inherited from OC, whereby when a cluster of verbs (two to four) precede a nominal, the nominal is interpreted as a direct object of each verb. That is, all verbs

in the cluster are necessarily transitive. Shi (2002: 45) calls this the verb coordination principle because it forces a coordination analysis for verb clusters; the phenomenon is also known as "multiple/serial-verb-common-object" among Chinese linguists (e.g., Zuo 2004, Xiao 2006, Jiang 2013). Crucially, the coordination analysis for verb clusters is possible because coordination structures in Chinese do not require overt coordinators. See (57) for examples.

(57) a. 而後陳卒亡，楚克有之。 [OC, late 4c. BCE]
ér hòu chén zú wáng, chǔ [$_{V_1}$ *kè*] [$_{V_2}$ *yǒu*] [$_O$ *zhī*]
and then Chen eventually perish Chu capture occupy it
'And then Chen (State) eventually perished. The Chu captured and occupied it.'
(ZZ, adapted from Xiao 2006: 112)

b. 姜原以爲神，遂收養長之。初欲棄之，因名曰棄。 [OC, 1c. BCE]
jiāng yuán yǐ wéi shén, suì [$_{V_1}$ *shōu*] [$_{V_2}$ *yǎng*] [$_{V_3}$ *zhǎng*]
Jiang Yuan take as god so collect nurture make grow

[$_O$ *zhī.*] *chū yù qì zhī, yīn míng yuē qì*
him first want discard him therefore name say discard
'Jiang Yuan thought he (i.e., her baby) was a god, so she took him back home and raised him up. Since she had wanted to discard him, she named him Qi (lit. discard).'
(SJI, adapted from Zuo 2004: 44)

c. 乃遣晃及史渙邀擊破走之，燒其輜重。 [OC, 3c.]
nǎi qiǎn huàng jí shǐ huàn [$_{V_1}$ *yāo*] [$_{V_2}$ *jī*]
so dispatch Huang and Shi Huan intercept attack

[$_{V_3}$ *pò*] [$_{V_4}$ *zǒu*] [$_O$ *zhī,*] *shāo qí zīzhòng*
break make flee them burn their supplies
'So (the king) sent Huang and Shi Huan to intercept and attack them (i.e., the enemies), break their army, and force them to flee. They burned their supplies.'
(SGZ, adapted from Shi 2002: 69)

The competition is fierce because the verb coordination principle has stably existed for centuries. And when the resultative compound construction finally wins out, it pushes verb coordination out of compositional syntax (Shi 2002: 70), with existing coordinative V-V collocations being either lexicalized as words (e.g., *shōu-yǎng* 'take home and nurture' > 'adopt') or reanalyzed as resultative compounds (e.g., *jī-pò* 'attack and break' > 'attack-broken'). According to Liu (2000: 255) most resultative compounds in late MidC are formed in this way. I illustrate the establishment of the resultative compound construction in Figure 5.7.

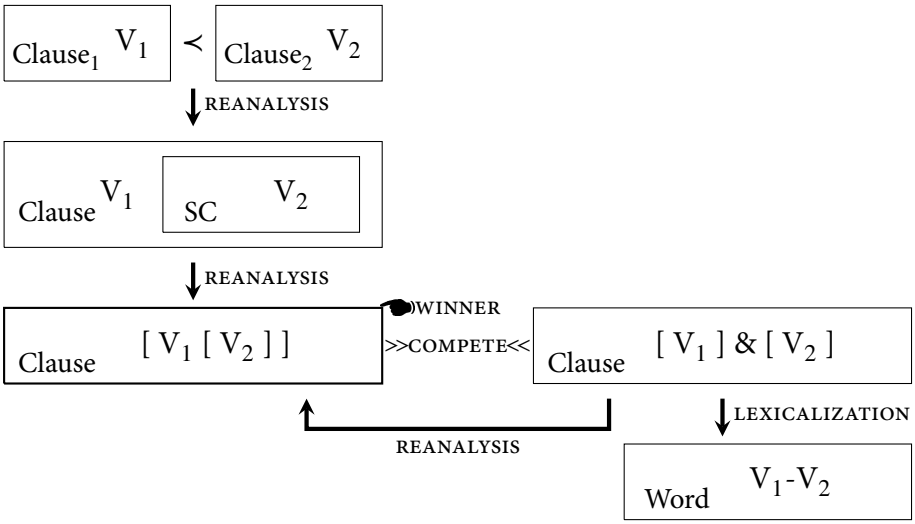


Figure 5.7 Establishment of the resultative compound construction

5.3.4 Distinctness and stative verbalizer

Let us pay closer attention to the competition between V-V coordination and resultative subordination. I display the two structures in tree format in Figure 5.8.

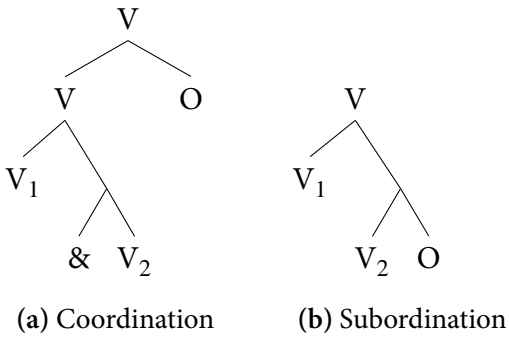


Figure 5.8 V-V coordination and V-V subordination

In the spirit of neo-emergentism, let us assume that all Vs in Figure 5.8 are categorially identical; namely, that they are built from the same mergeme-based category V. This is reasonable because all Vs in Figure 5.8 are freely selected lexical verbs. Under this simplest assumption, however, Figure 5.8 poses a problem: the two V categories are too similar and too close. It is well observed that natural languages discourage adjacent distribution of identical elements within the same local domain. In phonology this is known as the obligatory contour principle (OCP; Leben 1973, Goldsmith 1979, et seq.). See (58) for a general formulation (Ito & Mester 2003: 14).

(58) Obligatory contour principle
Adjacent identical elements are prohibited on an autosegmental tier.

OCP-like effects go far beyond phonology; see Van Oostendorp (2005), Neeleman & Van der Koot (2005), Biberauer (2008) inter alia for examples in morphosyntax. The widespread mani-

festation suggests that OCP is a highly general principle functioning in human language, like a third factor. In fact it has been incorporated into neo-emergentism, as in the following quote:

> [T]he child is specifically looking for … systematic departures from Saussurean arbitrariness [including] recursion. … Significantly, recursion also guarantees distinctness in the sense of Richards (2010), i.e. the requirement that … formally identical elements … cannot surface adjacent to each other within the same domain. (Biberauer 2017a: 43–45)

The distinctness condition Biberauer mentions is formulated as follows :

(59) Distinctness (Richards 2010: 5)
If a linearization statement $\langle \alpha, \alpha \rangle$ is generated, the derivation crashes.

Richards aims to develop a general theory for "bans on multiple objects of the same type that are too close together" (ibid., p. 3). His theory works as follows:

> [The distinctness] condition rejects trees in which two nodes that are both of type α are to be linearized in the same Spell-Out domain [i.e., the same strong phase], and are in an asymmetric c-command relation (so that a linearization statement relating them is generated [à la Kayne 1994]). (Richards 2010: 5)

Thus, distinctness rules out utterances like the following:

(60) a. * Every man admired every woman, except [$_A$ John] [$_B$ Mary].

b. * the singing [$_A$ of songs] [$_B$ of the children]

c. * *Jean a fait manger* [$_A$ *la tarte*] [$_B$ *Paul*]. [French]
Jean has made eat the pie Paul
Intended: 'Jean made Paul eat the pie.' (adapted from Richards 2010: 8–10)

In all three examples the constituents labeled A and B are in an asymmetric c-command relation and have identical syntactic types, so they are unlinearizable.[29] I have two comments on Richards' theory. First, Richards (2010: 6) distinguishes languages where distinctness solely references labels (e.g., English) and languages where it also references features like case and animacy (e.g., German, Japanese). I do not think this distinction is necessary given recent theoretical advancements. According to Chomsky (2013, 2015), labeling is just minimal search, and when the label provider is a head, the label of an SO is just its head per se. This means that all syntactically visible features on labeling heads are usable to render two nodes distinct. Therefore, the really important notion Richards' theory relies on is formal feature bundle rather than label (in

[29] Though according to Ian Roberts (p.c.) (60a) is okay.

its earlier sense). As he points out: "linearization statements make reference to feature bundles …, [which] can vary in their richness from language to language" (Richards 2010: 141).

Second, Richards (2010: 7) follows DM in assuming a fundamental difference between lexical and functional categories, asserting that "distinctness effects will arise just in interactions between functional heads." In particular, he assumes the early insertion of roots and holds that root information can be referenced in the linearization of lexical heads, thus "distinguishing the heads from each other." While I agree that distinctness is a condition on functional heads, I disagree that it may reference root information. Leaving aside the disputed issue of early/late root insertion, if distinctness relies on the comparison of syntactic nodes and a syntactic node is identified by its label (in the current sense of the term), then it follows that distinctness only references syntactic elements that qualify as labels; namely, categories with formal features. As such, radically formally underspecified objects like roots are invisible to the distinctness condition, much as they are invisible to labeling (Chomsky 2013: 47).

With the above considerations, I modify Richards' theory as follows:

(61) The modified distinctness condition
If a linearization statement $\langle \alpha, \alpha \rangle$ is generated, where α is a label in the sense of Chomsky (2013, 2015), the derivation crashes.

The modification is actually more neo-emergentist in spirit, as supported by the quote below:

> Recursion [drives] acquirers to postulate distinguishing [F]s [square brackets in the original] in cases where elements that appear to belong to "the same" category can be systematically stacked in a particular sequence …. Distinctness [is] required … where distribution is governed by categorial "sameness", i.e. *formally* identical elements compete for the same positions, meaning that co-occurring elements must be *formally* distinct in some way. (Biberauer 2016: 7; my emphases)

Returning to Chinese resultative compounds, now the problem in Figure 5.8b is clear: the two formally identical Vs are in a distinctness-violating configuration, as V_1 asymmetrically c-commands V_2 and the two Vs are in the same spell-out domain. Moreover, the distinctness-preserving strategies proposed in Richards (2010), such as insertion of extra elements and movement of nondistinct nodes, do not work here, because it is an empirical fact that the two lexical verbs in a resultative compounds are immediately adjacent. The problem with Richards' strategies is that they can only explain why certain constructions (such as those in (60)) are not attested but cannot explain why seemingly distinctness-violating constructions are licit and learnable. Biberauer's proposal in the above quote remedies this defect. So, the only solution for the problem in Figure 5.8b is to postulate a "distinguishing [F]." What feature to postulate then? Recall that in MidC the resultative compound structure is in competition with the multiple-verb-common-object structure. Given a surface string of the form V_1-V_2-O, the most salient

distinction between the two analyses lies in the nature of V_2: it is transitive in the multiple-verb-common-object analysis but intransitive and more exactly stative in the resultative compound analysis. So, a readily available semantic feature to be formalized is this stative eventuality type; call it [EVENT: STATIVE] or BE for short. This gives rise to a flavored verbal category; call it the *stative verbalizer* or v_{BE}. Since we only need a single feature to distinguish the two Vs, no additional flavor is needed. And importantly, the postulation of BE is not for the sake of semantic interpretation but for that of formal distinction—the eventuality type feature is just chosen because it is the most salient distinguishing [F] in the given scenario. Recall from §5.2.1 that eventuality types need not be encoded in v flavors but may be configurationally derived or simply exist as part of encyclopedic content. Thus, the nonflavored verbalizer may still introduce stative eventualities after v_{BE} is postulated. And in certain cases semantic states must be introduced by the nonflavored verbalizer; for instance, in state-caused results like *zuì-dǎo* 'drunk-fallen' and *là-kū* 'spicy-cry', the causing events are stative but cannot assume v_{BE}, for that would lead to a new distinctness-violating configuration [v_{BE} [v_{BE}]]. The crux here is that the formalized BE is not a semantically-purposed feature but merely "piggybacks" on one (Biberauer's term).

Finally, let us consider an ontological question. Being a flavored category, v_{BE} should be incomparable with v under the logic of this chapter, which means that BE must be a nonmergeme feature. But what would happen if BE enters the feature set of v as a mergeme feature? Since v already has a mergeme feature, there are two possibilities. The first possibility is that BE "squeezes in" the mergeme part of v. This creates an antidecompositional category, which I have argued in §5.2.3.2 to be illegal. The second possibility is that BE replaces the original mergeme specification of v in the newly formed category. In this case, v and v_{BE}, having distinct mergeme values and being in the same c-sequence, become comparable categories; hence, they must fall in a combinatorial-scope-induced ordering. This effectively creates a new functional category in the cartographic V-domain. However, considering functional categories are merged above lexical bases, and assuming the unflavored v is still the default category for run-of-the-mill lexical verbs, the only legitimate ordering is $v \leq v_{\text{BE}}$, which means that v_{BE} has a first-Merge position above v. But this contradicts the whole observation that the result slot is merged below the cause slot and the more general hypothesis that process is ontologically higher than result in event decomposition (Ramchand 2008, Baker 2018). In sum, neither of the two possibilities above is tenable, and so BE can only enter the feature set of v as a nonmergeme feature.

To conclude this section, I list the grammatical conditions or "preparations" in the diachronic emergence of the stative verbalizer in Chinese in (62).

(62) a. Radical pro-drop (without which the reanalyses in Table 5.5 cannot be kick-started).

b. Lack of inflection (without which verb m-roots cannot enter immediate adjacency).

c. A multiple-verb-common-object construction (without which there would be no competition for the resultative compound construction and hence no motivation for BE

to become the most salient formalizable feature).

d. The possibility of coordinator-less coordination (without which the multiple-verb-common-object construction would not exist).

e. The winning of the resultative compound construction in the competition (without which resultative compounds would just be a transient phenomenon in history).

The emergence of v_{BE} in Chinese is a result of the interaction of all the above conditions, the absence of any of which could have led to a completely different scenario than what we see today. The "evolution" of a flavored verbalizer is so difficult to motivate partly because a single nonflavored verbalizer already suffices for ordinary syntactic purposes as argued by opponents of flavored *v* (see §5.2.1). Accordingly, the usage of flavored *v*s should also be kept at a much more parsimonious level than that in the current DM literature, if flavors are taken as I-language primitives rather than just an expository device.

5.3.5 Ramifications of a new categorizer

Although v_{BE} in Chinese has emerged through a series of timely coincidences, once established it acquires the status of a major category by virtue of being a verbalizer, so we expect it to inherit all the properties a verbalizer normally has (e.g., licensing roots, shouldering extended projections). This sharply contrasts v_{BE} with other newly postulated categories such as the split-Infl/C categories. These latter categories merely enrich the "projecting field' of an existing major part of speech (MPOS), whereas v_{BE} essentially sets up a new MPOS. As such, its emergence should bring about more ramifications to the language than just the resultative compound construction. This is exactly the situation observed by Chinese linguists. According to Shi (2002), three out of the five big events in the history of Chinese listed in Wang (1989) are consequences of the establishment of the resultative compound construction, including the establishment of the resultative compound construction itself, that of the aspectual "suffixes," and that of the *bǎ* disposal construction. Shi (2002: 232–234) further identifies the following consequences of the resultative compound construction: the shift of Chinese from a verb-framing to a satellite-framing language (in the sense of Talmy 1972 et seq.), the appearance of the modal "infixes" *-de-* 'can' and *-bu-* '(can)not', the employment of verb reduplication to encode the diminutive aspect, and the expansion of the semantic boundedness requirement on predicates.

Under the current assumption that the establishment of the resultative compound construction is really that of v_{BE}, all the above consequences can be identified as ramifications of the new categorizer. Due to space limitations, here I only discuss two of them—the aspectual suffixes and the modal infixes, as illustrated in (63)–(64)—because these reflect the same diachronic process: upward grammaticalization (in the sense of Roberts & Roussou 2003). Moreover, they suggest a new way grammaticalization can work in human language.

(63) a. *tā chī le sān gè píngguǒ* [Mandarin] (aspectual "suffixes")
he eat PERF three CL apple
'He ate three apples."

b. *tā qù guò běijīng*
he go EXP Beijing
'He has been to Beijing.'

c. *tā zhèng tīng zhe yīnyuè*
he right listen PROG music
'He's listening to music right now.'

(64) *tā xiě de / bù wán zuòyè* (modal "infixes")
he write can cannot finished homework
'He can/can't finish the homework.'

In (63) the three underlined items respectively indicate perfective, experiential, and progressive aspects. In (63) the two items that can intervene between the two components of the resultative compound *xiě-wán* 'write-finished' respectively express the positive and negative potentiality of the accomplishment (i.e., finishing the homework). Note that *bù* in (64) has a different nature from the negative particles optionally preceding V_2 in MidC, such as *wèi* 'not have' in (65).

(65) 吾久欲注，尚未了。 (=53)
[$_{Clause_1}$ *wú jiǔ yù* [$_{V_1}$ *zhù*] [$_O$ pro,]
I long want annotate

[$_{Clause_2}$ [$_S$ pro] *shàng wèi* [$_{V_2}$ *liǎo.*]]
still not have complete
'I have long wanted to annotate (it), but (the annotation) has not yet been completed.'

In (65) what is denied is the independent truth of the situation in $Clause_2$, whereas in (64) what is denied is the likelihood of the complex event; namely, that of the realization of V_2 as a result of V_1. There are still other postverbal grammatical elements in Mandarin. For example, while the three underlined items in (63) all indicate grammatical/viewpoint aspects, the underlined items in (66) convey lexical/situation aspect (in the sense of Smith 1991).

(66) a. *tā xiě-wán/hǎo zuòyè le* (lexical aspect items)
he write-finished/good homework SFP
'He finished the homework.'

b. *tā dòng/è-zhe le*
he cold/hungry-touch SFP
'He got ill due to the cold/hunger.'

c. *tā zhōngyú chī-shàng fàn le*
he finally eat-up meal SFP
'He's finally got some food to eat.'

Wán and *hǎo* in (66a) both express completion of the writing event, *zhe* in (66b) indicates that the cold/hungry state has some negative effect, and *shàng* in (66c) expresses the contrastive inchoation of the eating event, with a positive implication. There are many more postverbal items in Chinese with such subtle aspectual meanings, which are not typically found in grammaticalized aspect markers. Besides, in Chinese the same form may be associated with several aspectual meanings, such as the progressive *zhe* in (63) and the negative-result *zhe* in (66). This is an instantiation of the kind of multifunctionality discussed in §4.3. Sometimes the multiple grammatical meanings of a postverbal item even form a complete grammaticalization path. A typical example is *le*, which originates from *liǎo* 'end, ended' and has developed three functions in Standard Mandarin, as in (67).

(67) a. *bié wàng le xiě zuòyè* [Standard Mandarin]
don't forget TELIC write homework
'Don't forget to do the homework!'

b. *tā chī le sān gè píngguǒ*
he eat PERF three CL apple
'He ate three apples.'

c. *tā qù běijīng le*
he go Beijing CRS
'He went to Beijing.'

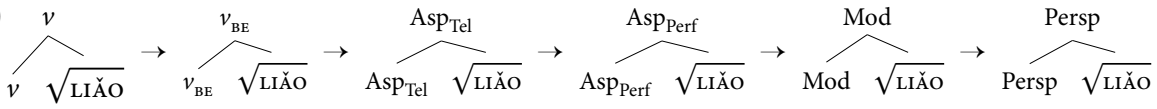
In many Northern Mandarin varieties *le* has also developed a modal function. The following examples are from Dongying Mandarin (where *le* is realized as *liu*, Song 2016c).

(68) a. *ɲi hɒ liu liang pingr ziou liu o* [Dongying Mandarin]
you drink ASP two bottle liquor MOD SFP
'Are you able to down two bottles?'

b. *ɲi zheng ɲimu dianr chi-hang faer liu o*
you earn so little eat-up meal MOD SFP
'Can you make a living (lit. have something to eat) out of so little income?'

c. *te sie-wae zuoye liu liu ae*
he write-finished homework ASP MOD SFP
'He can finish the homework. (confirmative tone)'

The above examples reveal a four-step grammaticalization path for the RSI *liǎo*: resultative → telic → perfective → potential → CRS. Including its initial acquisition of the RSI status, $\sqrt{\text{LIǍO}}$ has undergone five categorial changes from OC to ModC. I display these changes in the root-support format in (69).

(69) 

Recall from Chapter 4 that similar grammaticalization paths can be specified for other Chinese TAM elements. This reveals a striking scenario: Chinese functional elements in the verbal domain are mostly grammaticalized from RSIs; namely, their grammaticalization is based on v_{BE} instead of v.[30] This scenario has a natural explanation in the current theory. Since the newly established v_{BE} is a lexical category in the traditional sense, the items it categorizes (i.e., the RSIs) can have a separate grammaticalization route alongside that of the generic-v items. Additionally, this explains why postverbal grammatical elements had not developed in the OC period despite its impressive length:[31] OC did not have v_{BE} or the new grammaticalization route it makes available. The resultative compound construction matured by the twelfth century, and the eight centuries thereafter may well have witnessed the development of an entire "postverbal functional zone" as illustrated in Figure 5.9.

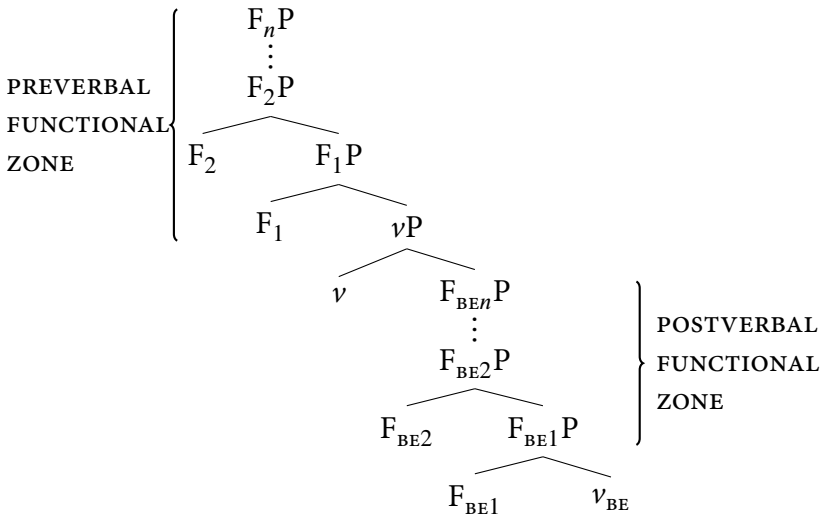
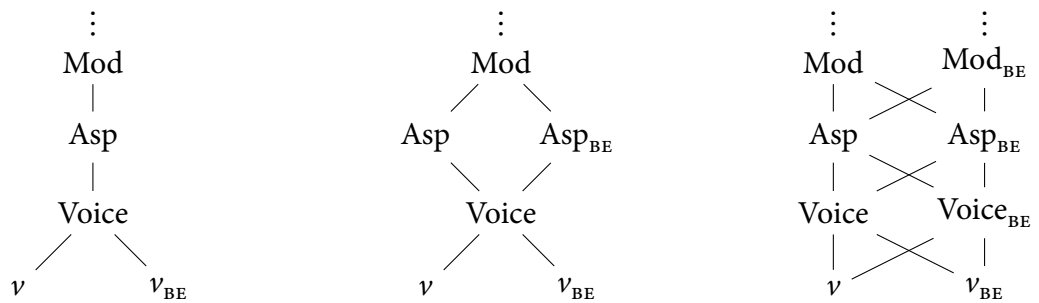


Figure 5.9 Two functional zones in Early Modern Chinese

[30] This does not mean non-RSI verbs cannot grammaticalize; but when they do, they tend to become light verbs or auxiliaries (see Chapter 4) and seldom (if ever) gain an "affixal" or particle status.

[31] The earliest written documents of OC are from the thirteenth century BCE (Baxter & Sagart 2014: 1), and the language used there is already grammatically mature (Zhang 2001).

I have three further comments on the dual-grammaticalization-route conception in Figure 5.9. First, since v_{BE} is essentially a subtype category of v, functional categories based on the two are not really that different but all categories in the "verbal" extended projection. That is, much as v is the basis for Asp, Mod, and so on, so v_{BE} can be the basis for these categories. There are no unpredictable functional categories in the new grammaticalization route. Second, since v and v_{BE} are incomparable categories in the same c-sequence, one might expect the functional categories based on them to be also incomparable. However, this is not necessarily the case. Under neo-emergentism, the simplest assumption is in fact that the functional categories above v and those above v_{BE} are formally identical unless there is specific motivation for further category flavoring. In other words, incomparability at lower positions of a c-sequence does not entail incomparability at its higher positions. Thus, the most likely c-sequence shape for Mid- to ModC is Figure 5.10a, the least likely one is Figure 5.10c, and something like Figure 5.10b is plausible but requires extra empirical motivation at each point of flavoring, just like the kind of motivation I have sought for the $v{\sim}v_{\text{BE}}$ flavoring. The three posets have different mathematical properties,[32] which can be a point of further research; in Chapter 6 I will prove based on the Categorical properties of c-sequences that core functional categories cannot be flavored.



(a) Most likely (b) Plausible if motivated (c) Least likely

Figure 5.10 Potential shapes for a bicategorizer c-sequence

Third, all the functional categories in Figure 5.10, by virtue of being comparable with both v and v_{BE}, can in theory be merged in both the preverbal and the postverbal functional zones. In other words, whatever category is postulated in whichever functional zone, a "replica" of it is automatically mirrored in the other functional zone. This follows from the poset nature of c-sequence (i.e., that flavored categories have shared combinatorial-scope-induced properties) and empirically means that once a functional item is successfully established in one functional zone, it can be readily used in the other. This explains why the postverbal aspectual elements, albeit all grammaticalized from RSIs, can be used with ordinary (i.e., nonstative) verbs as well, as exemplified in (63), (66), and (67).

[32] e.g., none of them is a lattice, though Figure 5.10a and Figure 5.10b may be semilattices.

5.4 Chapter summary

The aim of this chapter is to examine categorial flexibility at the abstraction layer of single c-sequences. The UG template at this abstraction layer is the combinatorial-scope-induced partial ordering, and the flexibility is that this ordering need not be a chain; when it is not a chain, we obtain incomparable categories. This type of flexibility is instantiated by the so-called flavored categories, though their current usage (especially in DM) is a bit excessive. Neo-emergentism promotes the parsimonious postulation of formal features. Following this idea, I have revisited Chinese resultative compounds, which have previously been analyzed in terms of flavored *v*s. After showing that such analyses suffer from a dummy verbalizer pitfall and arguing that the result slot is verbal instead of adjectival. I looked into the historical development of the resultative compound construction and discovered that while a flavored verbalizer (v_{BE}) is indeed involved in this process, its emergence is a result of the interaction of multiple language-specific conditions, none of which is readily replicable. This means that flavored categories, especially flavored categorizers, are not a norm in syntactic category formation but a deviation. Accordingly, we should use them parsimoniously in syntactic analyses. Finally, I discussed the ramifications of v_{BE} in Chinese beyond resultative compounds. It has created a new grammaticalization route for Chinese and hence had global impact on its grammar.

Chapter 6

Global interconnection: A Category-theoretic sketch

6.1 Ascending to the global level

In this dissertation, I have assumed a ladder of abstraction for the SCS that consists of five layers, each associated with a template and a type of flexibility, as in Table 6.1.

Table 6.1 Templates and flexibility types in the syntactic category system (= Table 2.1)

Level of concern	Abstraction layer	Template	Flexibility
Individual	single category	mergeme-based feature set	incomplete mergemes
	single c-sequence	scope-based partial ordering	incomparable categories
Global	single granularity level	universal spine	varied c-sequences
	all granularity levels	category subtyping	varied granularity levels
Supraglobal	entire SCS	syntactic rules	parametric variation

I have taken the templates as universal characteristics of human language[1] and accordingly assumed that their corresponding flexibility types emerge in places UG does not care about.[2] The five abstraction layers are divided into three broad levels of concern: individual, global, and

[1] A further question is whether *universal* equals *UG-given*? Among the five templates in Table 6.1, the first and the last are probably specific to language, but the middle ones all seem to have some domain-general nature; i.e., they may be universal cognitive strategies (or third factors) with linguistic instantiations. Considering this possibility, when I say something is universal or UG-given, I only mean that it is a mental capacity necessary for human language and remain agnostic about its precise location in the cognitive faculties. See Roberts (2017: 15) for a similar discussion in terms of "big UG" vs. "small UG."

[2] This is not to say the flexibility is unconstrained. What I claim is merely that if the templates are well defined, then whatever constrains the flexibility must not come from UG but should be second or third factors instead. There is little doubt that such constraints exist, as revealed in my investigation of flavored categories in Chapter 5, but a proper investigation along this line requires a separate project.

supraglobal. In Chapters 3–5 my attention has been limited to the individual level. In this chapter let us ascend to the global level and explore how syntactic categories and their constellations are interconnected in the SCS universe. My exploration will center on two interrelated phenomena:

(1) a. **Cross-MPOS parallelism:** There is a functional parallelism between syntactic categories across major parts of speech (MPOSs), such as that between C and D.

b. **Inter-GRAN relations:** There are systematic relations between components of different granularity levels (GRANs), such as the inheritance relation between C_{CFC} and split-C categories like Top, Foc, and Fin.

Both phenomena have been discussed in previous studies. In particular, Wiltschko's (2014 et seq.) universal-spine-based perspective on the cross-MPOS parallelism and Biberauer's (2016, 2017a) categorial-space-division-based perspective on the inter-GRAN inheritance have been inspiring for my study. The results in this chapter build on and extend these previous insights. In the remainder of §6.1 I lay out the main questions posed by the two phenomena. As a methodological note, since this chapter mostly deals with purely ontological issues, I will generally avoid derivational terms like *extended projection* and *head* (and use *c-sequence* and *mergeme-bearing category* instead). I do so because meanings of terms like *projection* and *head* depend on a (Chomskyan) derivational context, which we do not have in a purely ontological discourse.

6.1.1 Cross-major-part-of-speech parallelism

The cross-MPOS parallelism can be conceived as an inter-c-sequence connection. So, when we make statements like "T for verbs is like Num for nouns" and "C and D play similar roles,"[3] we are making an ontological connection between two c-sequences X_V and X_N, which is formally a binary relation $\boldsymbol{R} \subseteq X_V \times X_N$ containing pairs like ⟨T, Num⟩ and ⟨C, D⟩. Three points about this relation are worth our attention. First, even though the often-mentioned parallel pairs are just a few (e.g., ⟨*v*, *n*⟩, ⟨C, D⟩), no theoretical limit has been imposed on the extent of the parallelism. This in turn implies the possibility of a total connection, whereby for each verbal category there is a parallel nominal category. Second, insofar as conceptual functions[4] are concerned, and if we identify flavored categories as one (since they have identical mergeme values), each category is the component of a unique parallel pair. Thus, if ⟨C, D⟩ ∈ $\boldsymbol{R}$, then we do not expect to also have, for example, ⟨C, Num⟩ ∈ $\boldsymbol{R}$, as that would make Num and D functionally identical and lead us to ask why they should coexist in the same c-sequence in the first place. Third, the relation $\boldsymbol{R}$ is invertible under the same definition (is-parallel-with), because whenever a statement like "T

[3] I am only suggesting these statements as examples (any similar statements would do), leaving aside their validity.

[4] By *conceptual functions* I mean C-I-oriented functions like Wiltschko's classification, etc. Here I do not consider derivational functions like licensing agreement, hosting subject, etc. Admittedly, any properties—be they conceptual, derivational, or even phonological—may define parallel categories, and if we consider such heterogeneous properties, a single category may be parallel with multiple categories. But since my major concern is the ontological nature of syntactic categories, I restrict my cross-MPOS parallelism to a conceptually based one.

for verbs is like Num for nouns" is true, so is its inverse statement "Num for nouns is like T for verbs" ($\langle \mathrm{T}, \mathrm{Num}\rangle \in \boldsymbol{R} \Leftrightarrow \langle \mathrm{Num}, \mathrm{T}\rangle \in \boldsymbol{R}^{-1}$). That is, is-parallel-with defines a symmetric relation $\boldsymbol{R} \cup \boldsymbol{R}^{-1}$ on the union set $\mathrm{X}_{\mathbf{V}} \cup \mathrm{X}_{\mathbf{N}}$.[5]

The first two points above, total connection and unique pairing, effectively make $\boldsymbol{R}$ into a standard mathematical function, which by definition is total and single-valued. The third point, invertible pairing, further requires that this function have an inverse. The question then is, How can we find such an inter-c-sequence connection? The ideal scenario is a one-to-one correspondence, or bijection, where for each verbal category there is a matching nominal category, and vice versa. However, this is only feasible when the two c-sequences in question are special enough; for example, when they have equal cardinalities and only contain selected categories such as the core functional categories (2a) or phase categories (2b).

(2) a.
$$\begin{array}{ccccccc} \{\mathrm{V} & \leq & v & \leq & \mathrm{T} & \leq & \mathrm{C}\} \\ \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow \\ \{\mathrm{N} & \leq & n & \leq & \mathrm{Num} & \leq & \mathrm{D}\} \end{array}$$

b.
$$\begin{array}{ccc} \{v & \leq & \mathrm{C}\} \\ \updownarrow & & \updownarrow \\ \{n & \leq & \mathrm{D}\} \end{array}$$

Once we move on to the more commonplace c-sequences, such as the moderately elaborate ones in (3) (suppose they are of the same granularity level),[6] a bijection is no longer available.

(3)
$$\begin{array}{c} \{\mathrm{V} \leq \mathrm{Appl} \leq v \leq \mathrm{Asp} \leq \mathrm{T} \leq \mathrm{Mod} \leq \mathrm{Fin} \leq \mathrm{Foc} \leq \mathrm{Top}\} \\ \overset{?}{\Updownarrow} \\ \{\mathrm{N} \leq \mathrm{Gen} \leq n \leq \mathrm{Cl} \leq \mathrm{Num} \leq \mathrm{Q} \leq \mathrm{D}\} \end{array}$$

Of course, we may still pick out special pairs like $\langle \mathrm{T}, \mathrm{Num}\rangle$, but what are the nonspecial categories (e.g., Appl, Gen) parallel with? It seems when we move from (2) to (3), what becomes unavailable is not only the bijection, but in fact any meaningful function from either c-sequence to the other.[7] This is a serious problem, because between (2) and (3), the latter looks more like

[5] Since the parallelism is *across* c-sequences, $\boldsymbol{R} \cup \boldsymbol{R}^{-1}$ is not furthermore an equivalence relation (for it is not reflexive), though it may be identified as a *partial equivalence relation* (an equivalence relation minus reflexivity).

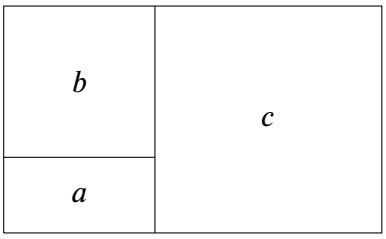
[6] The c-sequences in this chapter are merely expository. I have no commitment to the specific labels or orderings.

[7] One might want to drop the totality condition and view the parallelism as a partial function instead, but two facts render this concession of little help. First, if the lack of a limit on the cross-MPOS pairing implies that the parallelism is possibly total (p. 190), then by contraposition a necessarily nontotal parallelism implies the existence of a limit on what categories can be paired. Yet it is unclear what this limit could be. (2) may suggest that it is the core functional categories (plus V/N), but further pairs (such as $\langle \mathrm{Asp}, \mathrm{Cl}\rangle$ à la Borer 2005a,b) may easily be added to the bijection. Since such addition is legitimate as long as the paired categories can be justified to perform the same conceptual function, there is unlikely to be a fixed limit on the number of pairable categories. Second, since some special categories (e.g., C) dissolve in more elaborate c-sequences, the concessive connection is not only partial but also unstable, because already established pairs like $\langle \mathrm{C}, \mathrm{D}\rangle$ get lost when we move, for example, from (2)

the kind of c-sequence occurring in a working syntactician's day-to-day research, and if we cannot specify what a connection between such ordinary c-sequences is, then we cannot claim to have properly modeled the cross-MPOS parallelism. The intuition may be straightforward, but making it precise is not easy.

6.1.2 Inter-granularity-level relations

Considering c-sequences can be presented with different amounts of details, in Chapter 2 I introduced a granularity level component in the definition of syntactic categories. The c-sequences in (2a), (2b), and (3) are of three different granularity levels: CFC, Ph, and A.[8] Some categorial labels, such as v, are conventionally used at multiple granularity levels, but strictly speaking the categories that I label as v_{CFC}, v_{Ph}, and v_{A} have different extensions (qua sets of SOs), such that what is a member of v at the CFC level may become a member of Appl at the A level. From an acquisitional perspective (à la Biberauer 2017a), this reflects a scenario where the categorial space for v is successively divided, as in Figure 6.1.[9]



$$v_{\mathrm{Ph}} := a + b + c$$
$$v_{\mathrm{CFC}} := a + b$$
$$v_{\mathrm{A}} := a$$

Figure 6.1 Division of categorial space for v

In Figure 6.1 we can view the outer rectangle $a + b + c$ as v_{Ph}, the composite cell $a + b$ as v_{CFC}, and the monadic cell a as v_{A}. If we conceive these labels as terms of a formal language and the comparted rectangle (more exactly the set-theoretic structure it represents) as its model, then semantically we have $[\![v_{\mathrm{A}}]\!] \subseteq [\![v_{\mathrm{CFC}}]\!] \subseteq [\![v_{\mathrm{Ph}}]\!]$, which can be straightforwardly translated into the featural metalanguage in Chapter 2 as the sort-type inheritance relation $\mathtt{v_A} \sqsubseteq \mathtt{v_{CFC}} \sqsubseteq \mathtt{v_{Ph}}$, where $\mathtt{v_A}$ is a subtype of $\mathtt{v_{CFC}}$ and $\mathtt{v_{CFC}}$ a subtype of $\mathtt{v_{Ph}}$. This correspondence is not accidental but reflects the blurred boundary between first-order-predicate-denoting features and categories (see §2.2.2). The inheritance relation can be extended to entire c-sequences ($\mathrm{A_v} \sqsubseteq \mathrm{CFC_v} \sqsubseteq \mathrm{Ph_v}$) and granularity levels ($\mathrm{A} \sqsubseteq \mathrm{CFC} \sqsubseteq \mathrm{Ph}$), which makes it a viable tool to organize the *granularity level space* (GLS); namely, the collection of all theoretically possible granularity levels.

to (3), though the ontological parallelism is intuitively still there. Considering such undesirable effects, I will not compromise on the total function nature of the parallelism but aim to construct the desired total connection via more advanced mathematical tools instead.

[8] Recall from §2.3.2 that I use A, B, etc. to denote arbitrary granularity levels. A is between CFC and Cart.

[9] I use the different sizes of a, b, and c to show that division of the categorial space need not be even.

Recall from §2.3.2 that not all granularity levels are comparable. For example, neither of the two arbitrary levels D_1 and D_2 in (4) inherits the other.[10]

(4) (Adapted from (23) in Chapter 2)

a. $D_1 = \{\{V \leq v \leq \underbrace{\text{Asp} \leq \text{Tns} \leq \text{Mod}}_{T} \leq C\}, \{N \leq n \leq \text{Num} \leq D\}, \ldots\}$

b. $D_2 = \{\{\underbrace{\text{Res} \leq \text{Proc} \leq \text{Init}}_{V} \leq v \leq T \leq C\}, \{N \leq n \leq \text{Num} \leq D\}, \ldots\}$

The two granularity levels differ in the way CFC_v (the first row in (2a)) is subtyped: D_1 subtypes T and D_2 subtypes V.[11] Since division of the categorial space does not have an undo-and-redo option, neither D_1 nor D_2 can be obtained by manipulating the categorial space configuration of the other; they are only related by a common supertype level CFC instead, which is an upper bound of the subset $\{D_1, D_2\}$ in order-theoretic terms.

There is still another form of incomparability among granularity levels. Typologically, we cannot guarantee that what is labeled T in one language variety is the same category as what is labeled T in another variety. This is because division of the categorial space can be guided by any SCDPs (§2.3.1.2),[12] and just because successive division may produce the same geometrical shape (e.g., the tripartite rectangle in Figure 6.1) in two varieties does not mean that each cell in the shape also has the same featural content. So, we expect to have not only incomparable granularity levels produced by heterogeneous subtyping paths as in (4) but also those produced by different instantiations of the same c-sequence position as in (5) (the T/Loc variation is based on Wiltschko 2014) and those resulted from parametric settings of the same c-sequence as in (6).

(5) a. $CFC_1 = \{\{V \leq v \leq T \leq C\}, \ldots\}$

b. $CFC_2 = \{\{V \leq v \leq \text{Loc} \leq C\}, \ldots\}$

(6) a. $CFC_3 = \{\{V \leq v_{\text{head-initial}} \leq T_{\varphi\text{-complete}} \leq C\}, \ldots\}$

b. $CFC_4 = \{\{V \leq v_{\text{head-final}} \leq T_{\varphi\text{-incomplete}} \leq C\}, \ldots\}$

Given the various types of incomparability, we expect the GLS to be a much larger poset than that initially introduced in §2.3.2.

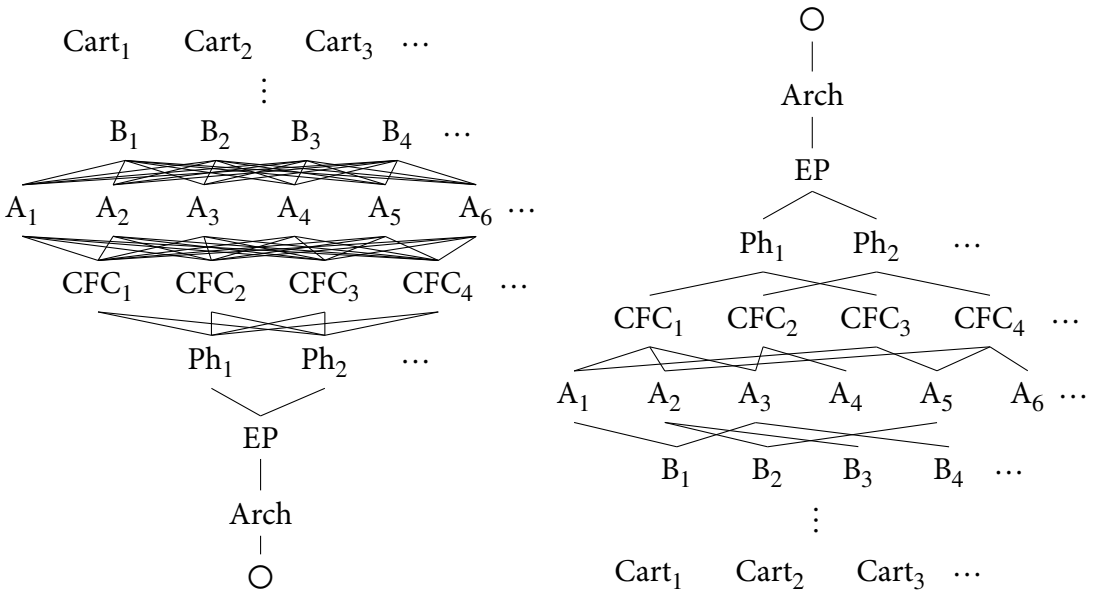
Apart from inheritance, there are still other relations that may be used to organize gran-

[10] I use numerical subscripts to distinguish incomparable granularity levels of the same general level.

[11] The two levels D_1 and D_2 can be understood in both analytical and acquisitional terms. Analytically, a linguist studying T-related phenomena may not need an elaborate decomposition of V. Acquisitionally, a child may have acquired the split-T categories but not yet encountered sufficient evidence motivating the further subtyping of V.

[12] As an aside, though a c-sequence and a divided categorial space both have individual categories as components, they highlight different facets of the categories. To obtain a well-formed c-sequence we only need the mergemes, but in division of the categorial space it is mostly the nonmergeme features that are at work.

ularity levels. For example, the inheritance-based ordering A $\sqsubseteq$ CFC $\sqsubseteq$ Ph is opposite to the fine-grainedness-based ordering Ph $\leq$ CFC $\leq$ A (where $\leq$ denotes is-coarser-than-or-equal-to). We may also redefine $\leq$ as is-finer-than-or-equal-to (see note 67, p. 31) and get another fine-grainedness-based ordering A $\leq$ CFC $\leq$ Ph.[13] All these—and possibly other—perspectives on granularity level ordering are legitimate, and they all reflect the global interconnection in the GLS. However, a disadvantage with such order-theoretic organizations is that they make the interconnection look rather convoluted, as in Figure 6.2.



(a) GLS ordered by is-coarser-than-or-equal-to **(b)** GLS ordered by inherits/is-a-subtype-of

Figure 6.2 Order-theoretic organizations of the GLS

The two diagrams in Figure 6.2 both reveal interesting information about the GLS. Figure 6.2a is a meet-semilattice, where every two granularity levels have a common coarsest predecessor (i.e., a meet), while Figure 6.2b is a join-semilattice where every two granularity levels have a common smallest supertype (i.e., a join).[14] In addition, Figure 6.2b is not simply Figure 6.2a reversed, as the inheritance relation is not the exact inverse of the coarser-than-or-equal-to relation (which is finer-than-or-equal-to).[15] Despite their informativeness, however, in both diagrams (especially in Figure 6.2a) the edges in the middle zone get so dense when we add in more incomparable granularity levels that they quickly become illegible. Informally, an order-theoretic organization of the GLS is a "low-level" description (analogous to an assembly computer lan-

[13] The opposite orderings here are reminiscent of a similar scenario in the HPSG feature structure/type hierarchy (Pollard & Sag 1987: 42, note 10), where subtypes are subsumed by supertypes (subtype $\preceq$ supertype) and supertypes are less informative than subtypes (supertype $\sqsubseteq$ subtype). In fact, our inheritance relation is just Pollard & Sag's subsumption relation rephrased in the active voice (inherits $\equiv$ is-subsumed-by), and their subsumption hierarchies are indeed also called "inheritance hierarchies" (Pollard & Sag 1987, Carpenter 1992).

[14] See Appendix C for definitions of the order-theoretic terms used here.

[15] For example, that A_1 is finer than CFC_2 does not entail that A_1 is derivable from CFC_2 via division of the categorial space; they may follow independent subtyping paths instead.

guage); but to make more sense out of the global interconnection, we need a "higher-level" description, one that encodes some more abstract inter-GRAN relation(s).

Last but not least, if we compare the cross-MPOS parallelism and the inter-GRAN relations side by side; it becomes clear that they share the same basic components: c-sequences. The status of c-sequences at the global level is similar to that of syntactic categories at the individual level: both are ground vocabulary out of which more complex relations are built and more abstract properties formulated. This similarity leads us to a new question: Can the cross-MPOS parallelism and the inter-GRAN relations be somehow unified?

In sum, three questions are interesting regarding the global interconnection in the SCS:

(7) a. The cross-MPOS parallelism is intuitively an invertible function, but there is often no obvious bijection between two given c-sequences. How can a linguistically meaningful function be constructed that can adequately and consistently formalize the parallelism?

b. The GLS can be ordered by different relations, but order-theoretic organizations tend to be cluttered with low-level details. How can a higher-level organization be attained which is sufficiently abstract but also reveals no less information about the GLS?

c. The cross-MPOS parallelism and the inter-GRAN relations are phenomena at different abstraction layers, but they belong to the same (global) level of concern and also share the same ground vocabulary (c-sequences). Is it possible to have a unified theory for the global interconnection of the SCS?

I aim to answer all three questions in this chapter by borrowing some well-established results from a flourishing subfield of mathematics: Category theory.

6.2 Why Category theory?

As mentioned in §2.4.2, I take feature theory and Category theory as two halves of a complete metalanguage for the SCS. Feature theory has proven advantages in describing the internal content of syntactic categories and in explaining how individual categories are composed into interpretable utterances.[16] But when we shift our attention from derivations/representations of concrete utterances to abstract ontological properties of the SCS, we also move beyond the design capacity of feature theory. This point is clearly reflected in the nature of the two global-level phenomena laid out in §6.1. It is hard to see how we can employ feature theory to construct a bijection between c-sequences of different sizes or to organize intertwining granularity levels. Features by definition are inhabitants of specific categories; yet the global-level inquiries not only go beyond specific categories and utterances but also transcend individual speakers, language varieties, and

[16] Not that Category theory cannot tackle these (see Lambek 1988 for an introduction to Categorical grammar), but the descriptive and explanatory adequacy feature theory has achieved on these issues is unparalleled.

even possible worlds (recall that the GLS contains all theoretically possible granularity levels realizable in human language). Since the feature-theoretic mechanisms (Match, Copy, Unify, etc.) are not designed for such large-scale purposes, we need an alternative tool to complement the foundational role of features in linguistic theorization.

Category theory is a mathematical theory put forward by Samuel Eilenberg and Saunders Mac Lane in the 1940s (the seminal paper is Eilenberg & Mac Lane 1945). Its original purpose was to bridge two subfields in mathematics, algebra and topology, so one can fairly say that Category theory is a boundary breaker by birth. Spivak (2014: 3) remarks that the "language and tool set" provided by Category theory are "useful throughout science." Fong & Spivak (2018: iii, 5) further characterize Category theory as being "all about organizing and layering structures" and "unmatched in its ability to … find commonalities." In a different context (about autonomous evolutionary systems), Ehresmann & Vanbremeersch (2007: 2) give the following summary, which appears highly relevant for the study of human language as well:

> [C]ategory theory could provide tools to study concepts germane to complexity[:]
>
> (i) *The binding problem*: how do simple objects bind together to form a 'whole that is greater than the sum of its parts'?
>
> (ii) *The emergence problem*: how do the properties of a complex object relate to the properties of the more elementary objects that it binds?
>
> (iii) *The hierarchy problem*: how may we explain the formation of increasingly complex objects, beginning with elementary particles … up through increasingly complicated systems …?

The role of Category theory as a "universal language of science" is witnessed by its fast and far-reaching interdisciplinary application. Till now, it has been successfully applied to a wide range of subjects including mathematical foundation (e.g., Lawvere 1964 et seq.), physics (e.g., Lawvere & Schanuel 1986, Baez 2001), logic (e.g., Lambek & Scott 1986, Pitts 2001, Goldblatt 2006), computer science (e.g., Barr & Wells 1990, Rydeheard & Burstall 1988, Pierce 1991), linguistics (e.g., Lambek 1988 et seq., Dougherty 1993, Moortgat 2007, 2009, Gillibert & Retoré 2014, Preller 2014, Piedeleu et al. 2015), cognitive science (e.g., Magnan & Reyes 1994, Phillips & Wilson 2014, 2016), philosophy (Ellerman 2007, Weatherall 2016, Landry 2017), music (e.g., Mazzola & Mannone 2016), and so forth. Recently there has even appeared a burgeoning subfield called applied Category theory (see Bradley 2018 for an introduction).

However, so far the application of Category theory in linguistics has been limited to two subfields: (i) categorial grammar[17] and its descendants (based on Lambek calculus), (ii) computational linguistics and natural language processing.[18] Research in these subfields has been

[17] Categorial grammar ≠ Categorical grammar. The latter is a particular descendant of the former.

[18] Notable exceptions are Pollard's (2004 et seq.) higher-order grammar (an extension of HPSG) and the categorial

fruitful, but we can surely ask if there may still be other possible applications of Category theory in linguistics. The study in this chapter is an innovative attempt, as it shows (i) that the linguistic application of Category theory need not be limited to certain (NLP-oriented) subfields, and (ii) that natural languages provide rich ground to experiment various Category-theoretic ideas.

6.3 A Category theory primer

Now I introduce the basic framework of Category theory. Unlike the rest of the dissertation, this section is presented in the mathematical style. I will generally omit the proofs for well-established theorems, which can be found in any textbook. The major sources consulted in writing this introduction are Lambek & Scott (1986), Crole (1993), Mac Lane (1998), Goldblatt (2006), Lawvere & Schanuel (2009), Awodey (2010), Simmons (2011), Spivak (2014), Leinster (2016), Fong & Spivak (2018), Pitts (2018), and Smith (2018). For space reasons, I will only introduce notions and results that will be directly applied to my study of the SCS global interconnection.[19] I follow the conventional path in introductory texts on Category theory and define Category, Functor, natural transformation, and Adjunction in turn;[20] there will also be a brief introduction of Yoneda embedding. See Appendices C–E for notions that are not directly used in this chapter but needed to define notions that *are* used.

6.3.1 Category

Definition 6.3.1.1 (Category). A Category $\mathscr{C}$ comprises

1. a collection of *objects* $A, B, C \ldots$;
2. a collection of *arrows* (or *morphisms*) $f, g, h \ldots$;
3. two operations **src** and **tar** assigning to each arrow f a *source* (or *domain*) object $\mathbf{src}\, f$ and a *target* (or *codomain*) object $\mathbf{tar}\, f$. If $A = \mathbf{src}\, f$ and $B = \mathbf{tar}\, f$ we display this as

$$f : A \to B \qquad \text{or} \qquad A \xrightarrow{f} B\,;$$

4. an operation assigning to each pair of arrows $\langle g, f \rangle$ with $\mathbf{src}\, g = \mathbf{tar}\, f$ a *composite* arrow $g \circ f$ which satisfies *assosciativity*; that is, given the configuration

$$A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{h} D\,,$$

minimalist grammar developed by a group of French scholars (see, e.g., Lecomte 2005, 2008, Amblard et al. 2010, Amblard 2011). However, since both approaches are based on Lambek's theory, their applications of Category theory are still within the above-mentioned limitation.

[19] So notions like universal construction and limit, albeit of vital importance in Category theory, will be left out.

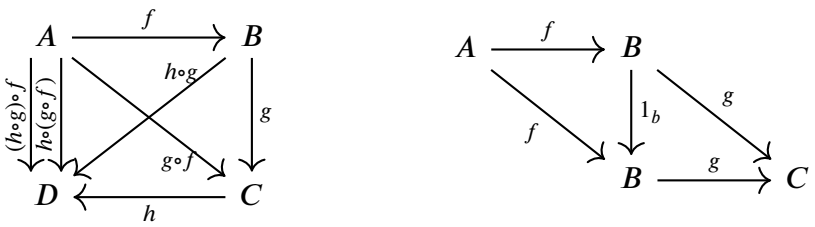
[20] To avoid confusion, I capitalize the initial letters of Category-theoretic terms that have linguistic homonyms.

we have $h \circ (g \circ f) = (h \circ g) \circ f$;

5. an assignment to each object B of an *identity arrow* $1_B : B \to B$ which satisfies the *unit law*; that is, for any arrows $f : A \to B$ and $g : B \to C$ we have

$$1_B \circ f = f \qquad \text{and} \qquad g \circ 1_B = g\,.$$

Note 6.3.1.1 (commutative diagram). Equations like those in Definition 6.3.1.1 are illustrated by *commutative diagrams*, as in Figure 6.3. Commutativity is the equivalence of parallel paths between a pair of objects. In Figure 6.3a, for instance, $A \to B \to D = A \to C \to D$.



(a) Associativity (b) Unit law

Figure 6.3 Commutative diagrams for associativity and unit law

Remark 6.3.1.1. Some terms in Definition 6.3.1.1 (e.g., [co]domain) suggest links to set theory, but in general the objects/arrows in a Category may have nothing to do with sets/functions.

Example 6.3.1.1 (Category of sets). All sets and functions between them form a Category **Set**, where arrow composition is just function composition. In the configuration of sets and functions

$$A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{h} D\,,$$

we have $h \circ (g \circ f)(a) = (h \circ g) \circ f(a)$ for all $a \in A$ (associativity). For any functions $f : A \to B$ and $g : B \to C$, we have

$$1_b \circ f(a) = f(a) \qquad \text{and} \qquad g \circ 1_b(b) = g(b)$$

for all $a \in A$ and $b \in B$ (unit law).

Example 6.3.1.2 (Category of posets/preorders). All posets and monotone functions between them form a Category **Pos**. Recall that a poset is a set equipped with a partial order. Remove antisymmetry from the definition of poset and we are left with a *preordered set*. All preordered sets and monotone functions between them form another Category **Pre**.

Example 6.3.1.3 (poset/preorder Category). A single poset $\langle P, \leq \rangle$ can be viewed as a Category $\mathscr{P}_{\leq}$ (called a *poset Category*), whose objects are the poset elements and arrows are instances of

the order relation; that is, for $p, q \in P$ there is an arrow $p \to q$ when $p \leq q$. Composition and identity are respectively defined by order transitivity ($p \to q \to r \equiv p \leq q \leq r$) and reflexivity ($1_p \equiv p \leq p$). The identity axiom is the reason why I have adopted $\leq$ instead of $<$ in §2.3.2 to define c-sequences. Similarly, a single preordered set can be viewed as a Category, called a *preorder Category*. Note that in an arbitrary poset or preorder Category, the objects are not sets and arrows not functions. Some useful order-theoretic notions are given in Appendix C.

Example 6.3.1.4 (Magnan & Reyes 1994). Consider Julie's family relations in Figure 6.4.

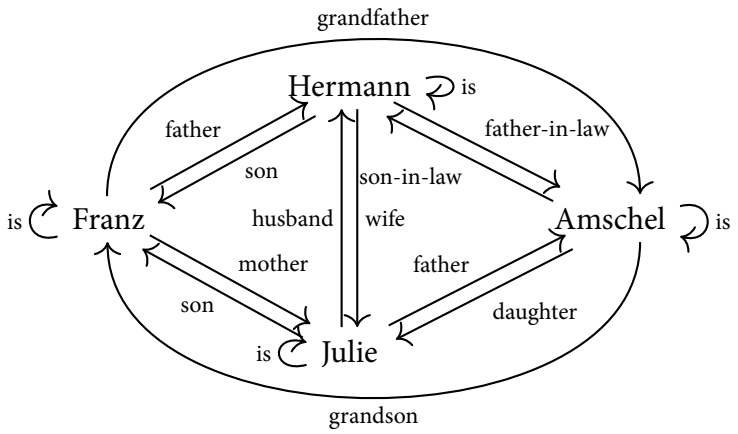


Figure 6.4 A real-life Category

Figure 6.4 is a preorder Category, where each arrow f defines a specific instance of the general relation (**tar** f -)is-a-family-member-of(-**src** f). Composition is defined by family relation composition (e.g., the father-in-law of the father of Franz is the grandfather of Franz) and identity by personal identity (e.g., Franz is Franz).

Note 6.3.1.2. Poset and preorder Categories are special in that from any object to any object there is at most one arrow (i.e., only when the objects in question are comparable). More generally, any Category with this property is called a *posetal* or *thin Category*.

Below we define some important special Categories (see Appendix E for more).

Definition 6.3.1.2 (discrete Category)**.** A Category whose only arrows are identities is *discrete*.

Example 6.3.1.5. Any set can be viewed as a discrete Category.

Definition 6.3.1.3 (Subcategory)**.** A *Subcategory* $\mathscr{J}$ of a Category $\mathscr{C}$ consists of a selection of $\mathscr{C}$'s objects and arrows (NB $\mathscr{J}$ must still meet the definition of a Category). $\mathscr{J}$ is a *full Subcategory* if it keeps all the arrows between its selection of $\mathscr{C}$-objects.

Example 6.3.1.6. The Category $\mathbf{Set}_f$ of all finite sets is a full Subcategory of **Set**.

Definition 6.3.1.4 (Category of Categories)**.** All (small) Categories and Functors between them form a Category **Cat** (which is large). Composition and identities are defined by Functor composition and identity Functors. See Appendix D for definitions of *small* and *large*.

Definition 6.3.1.5 (skeletal Category)**.** A Category $\mathscr{J}$ is a *skeleton* of another Category $\mathscr{C}$ if $\mathscr{J}$ is a full Subcategory of $\mathscr{C}$ and contains exactly one object from each class of isomorphic objects in $\mathscr{C}$. A Category is said to be *skeletal* if it is a skeleton of some Category.

Example 6.3.1.7. A poset Category is a skeletal preorder Category.

Theorem 6.3.1.1. *If $\mathscr{J}$ is a skeleton of $\mathscr{C}$, then the two Categories are equivalent (Definition* 6.3.3.3*).*

A general rule in Category theory is that we are not allowed to peek inside objects, but this restriction may sometimes be bypassed. For instance, sometimes a mathematical construct can be viewed as either an objects or a Category (a poset is such a construct; see Examples 6.3.1.2 and 6.3.1.3). There are also special arrows allowing us to deduce information about objects. Some important special objects/arrows are defined in the remainder of §6.3.1.

Definition 6.3.1.6 (initial/terminal/zero object)**.** In a Category $\mathscr{C}$, an object I is *initial* iff for every X in $\mathscr{C}$ there is a unique arrow $I \to X$, an object T is *terminal* iff for every X in $\mathscr{C}$ there is a unique arrow $X \to T$, and an object O is *zero* (or *null*) iff it is both initial and terminal.

Example 6.3.1.8. In **Set** the empty set is initial, and any singleton set is terminal, for there is a unique (empty) function from the empty set to any set and a unique function from any nonempty set to a singleton set. In **Pos** the empty poset is initial, and any singleton poset is terminal.

Example 6.3.1.9. In the Category **Grp** of groups and group homomorphisms,[21] there is a unique arrow from the trivial group Gr_0 comprising only an identity element e to any group, which maps e to its identity element, so Gr_0 is initial. There is also a unique arrow from any group to Gr_0 mapping all its elements to e, so Gr_0 is terminal. Hence, Gr_0 is a zero object in **Grp**.

[21] A *group* is an algebra with an associative binary operation, an identity element, and an inverse element for each element. For example, the set of integers with addition forms a group, where the identity element is 0 and the inverse for any element a is $-a$.

Definition 6.3.1.7 (mono/epi/iso)**.** In any Category $\mathscr{C}$, an arrow $f : A \to B$ is called a *monomorphism* (or *mono*) if for all $g, h : C \to A$

$$f \circ g = f \circ h \Rightarrow g = h .$$

Dually, $f : A \to B$ is called an *epimorphism* (or *epi*) if for all $i, j : B \to D$

$$i \circ f = j \circ f \Rightarrow i = j .$$

An arrow $f : A \to B$ is called an *isomorphism* (or *iso*) if there is an arrow $g : B \to A$ such that

$$g \circ f = 1_A \text{ and } f \circ g = 1_B . \tag{6.3.1.1}$$

In this case, we say A is isomorphic to B and write $A \cong B$.

Example 6.3.1.10. In **Set**, monomorphisms are injective functions, epimorphisms are surjective functions, and isomorphisms are bijective functions.

Definition 6.3.1.8 (split mono/epi)**.** Given an arrow $s : A \to B$ in a Category $\mathscr{C}$, if there exists another arrow $r : B \to A$ such that $r \circ s = 1_A$, pictorially

$$\begin{tikzcd} & B \arrow[dr, "r"] & \\ A \arrow[ur, "s"] \arrow[rr, "1_A"'] & & A , \end{tikzcd} \tag{6.3.1.2}$$

then s is called a *split monomorphism* (or *right-inverse/section*) and r is called a *split epimorphism* (or *left-inverse/retraction*). Accordingly, A is called a *retract* of B.

Note 6.3.1.3. Comparing (6.3.1.1) and (6.3.1.2), we see that the former is a two-sided or full inverse, whereas the latter is only a one-sided or half inverse. In the half-inverse situation, though $s \circ r \neq 1_B$, the composite $s \circ r$ does have another nice property called *idempotence* (Definition 6.3.1.9). Every section-retraction pair $\langle s, r \rangle$ is the splitting of an idempotent map $s \circ r$.

Definition 6.3.1.9 (idempotence)**.** An *endomorphism* e is an arrow from an object to itself. It is *idempotent* if $e \circ e = e$.

Example 6.3.1.11 (Spivak 2014)**.** In the following diagram, M is the set of mothers and C the set of children, both viewed as objects in **Set**.

$$\begin{tikzcd} M \arrow[r, shift left, "f = \text{has as first-born}"] & C \arrow[l, shift left, "g = \text{has}"] \end{tikzcd}$$

This is a half-inverse situation. It says that while every mother is the mother of her first-born child ($g \circ f = 1_M$), not every child is the first-born child of its mother ($f \circ g \neq 1_C$). And the

idempotence $(f \circ g) \circ (f \circ g) = f \circ g$ says that applying the find-first-born-child-for-some-child's-mother procedure twice returns the same child as applying it once.

Example 6.3.1.12 (Lawvere & Schanuel 2009). In the following diagram, C is the set of US citizens and D the set of congressional districts, both viewed as objects in **Set**.

$$C \underset{s=\text{ representative}}{\overset{f=\text{ residence district}}{\rightleftarrows}} D$$

This is another half-inverse situation. It says that while every representative is a citizen of the district they represent ($f \circ s = 1_B$), not every citizen is a representative of their district ($s \circ f \neq 1_A$). And the idempotence $(s \circ f) \circ (s \circ f) = s \circ f$ says that applying the find-representative-for-some-citizen's-district procedure twice returns the same person as applying it once.

Remark 6.3.1.2. A section-retraction pair typically exists for objects different in size, with the smaller object being a retract of the larger object. When the objects are (possibly structured) sets, the smaller set can also be viewed as a *sorting object* by which elements in the larger set are sorted. If there is additionally no empty sort, then the sorting becomes a *partition*. Note that in Example 6.3.1.12, we did not specify how the representatives are chosen. Since a district has many citizens, there could be many possible choices. For instance, one district may choose its richest citizen as the representative, while another may choose its most popular singer. We will find ourselves in exactly the same situation in formalizing the cross-MPOS parallelism, and Category theory will help us find the "best choice."

6.3.2 Functor and natural transformation

Definition 6.3.2.1 (Functor). A Functor F from $\mathscr{C}$ to $\mathscr{D}$ is a mapping of objects to objects and arrows to arrows such that for A, B and f, g in $\mathscr{C}$

1. $F(f : A \to B) = F(f) : F(A) \to F(B)$;
2. $F(1_A) = 1_{F(A)}$; and
3. $F(g \circ f) = F(g) \circ F(f)$, whenever $g \circ f$ is defined in $\mathscr{C}$.

That is, F preserves sources and targets, identity arrows, and compositions; pictorially

$$\begin{tikzcd}
 & \mathscr{C} \arrow[rr, "F"] & & \mathscr{D} \\
1_A \circlearrowright A \arrow[r, "f"] \arrow[dr, "h"'] & B \circlearrowleft 1_B \arrow[d, "g"] & 1_{F(A)} \circlearrowright F(A) \arrow[r, "F(f)"] \arrow[dr, "F(h)"'] & F(B) \circlearrowleft 1_{F(B)} \arrow[d, "F(g)"] \\
 & C \circlearrowleft 1_C & & F(C) \circlearrowleft 1_{F(C)}
\end{tikzcd}$$

Remark 6.3.2.1. A Functor produces an image of the source Category in the target Category. Strictly speaking, a Functor consists of two suitably related mappings, one for objects and the other for arrows, but in mathematics both mappings get the same name (e.g., F above).[22]

Definition 6.3.2.2 (full/faithful Functor). A Functor $F : \mathscr{C} \to \mathscr{D}$ is *faithful* if given parallel $\mathscr{C}$-arrows $A \underset{g}{\overset{f}{\rightrightarrows}} B$, $F(f) = F(g)$ implies $f = g$. Dually, F is *full* if given $\mathscr{C}$-objects A, B, each $\mathscr{D}$-arrow $g : F(A) \to F(B)$ implies a $\mathscr{C}$-arrow $f : A \to B$ for which $F(f) = g$. A Functor is *fully faithful* if it is both full and faithful.

Note 6.3.2.1. For locally small Categories, fullness and faithfulness yield special functions on *hom-sets* (see Appendix D.2). A full Functor is a surjective function, a faithful Functor is an injective function, and a fully faithful Functor is a bijective function. An *inclusion Functor* (see Appendix E.3) $F : \mathscr{S} \to \mathscr{C}$ is automatically faithful; when it is also full, $\mathscr{S}$ is a full Subcategory of $\mathscr{C}$ (Definition 6.3.1.3). A fully faithful Functor injective on objects[23] is an *embedding Functor*.

Example 6.3.2.1 (forgetful Functor). The Functor $U : \mathbf{Pos} \to \mathbf{Set}$ sending posets to their underlying sets is *forgetful* (in that it "forgets" the order relations). There are many forgetful Functors, representatively those from Categories of structured sets (preordered sets, groups, etc.) to **Set**.

Example 6.3.2.2 (Lambek 1988). Given a Category **Syn** of syntax and a Category **Sem** of semantics for any natural or formal language, there is an *interpretation Functor* $\Phi : \mathbf{Syn} \to \mathbf{Sem}$ sending each syntactic expression to its meaning.

Example 6.3.2.3 (monotone function). Given two poset Categories $\mathscr{P}_{\leq}$ and $\mathscr{Q}_{\sqsubseteq}$, a Functor between them, $F : \mathscr{P}_{\leq} \to \mathscr{Q}_{\sqsubseteq}$, is simply a monotone function.

Note 6.3.2.2. A monotone function f qua Functor is automatically faithful as there is at most one arrow between any two objects in a poset Category. If f is also full and injective on objects, it becomes an *order embedding*—an embedding Functor between poset Categories. The section arrow in Definition 6.3.1.8 qualifies as an order embedding if the two sets involved are posets.

Definition 6.3.2.3 (natural transformation). Given Categories and Functors in the configuration $\mathscr{A} \underset{G}{\overset{F}{\rightrightarrows}} \mathscr{B}$, a natural transformation $\alpha : F \Rightarrow G$ is a family of arrows in $\mathscr{B}$, one for each object in $\mathscr{A}$, such that for every $\mathscr{A}$-arrow $f : A \to A'$ we have the commutative square

$$\begin{tikzcd} F(A) \arrow[r, "F(f)"] \arrow[d, "\alpha_A"'] & F(A') \arrow[d, "\alpha_{A'}"] \\ G(A) \arrow[r, "G(f)"] & G(A'), \end{tikzcd}$$

which is called a *naturality square*. The maps $\alpha_{A \in \mathscr{A}}$ are α's *components*. If each α-component is an isomorphism, the natural transformation is a *natural isomorphism*, which I notate by $\Leftrightarrow$.

[22] They get separate names in functional programming (e.g., in Haskell), where the term *Functor* is usually reserved for arrow mappings, while object mappings are called *type constructors* (see, e.g., Milewski 2018).

[23] A Functor $F : \mathscr{A} \to \mathscr{B}$ is injective on objects if for any objects A, A' in $\mathscr{A}$, $F(A) = F(A')$ implies $A = A'$.

Remark 6.3.2.2. A natural transformation is intuitively a transformation in the target Category between two images of the source Category created by two Functors. There are two ways to locate natural transformations in the Categorical big picture. On the one hand, we may view all Functors from $\mathscr{A}$ to $\mathscr{B}$ as objects in a *Functor Category*, notated $\mathscr{B}^{\mathscr{A}}$, whose arrows are natural transformations. On the other hand, we may also view $\mathscr{A}$ and $\mathscr{B}$ as objects in **Cat** (Definition 6.3.1.4), which is a *2-Category* that directly has arrows between arrows (called *2-cells*).

6.3.3 Adjunction and Category comparison

The significance of Adjunction in mathematics is beyond doubt. As Goldblatt (2006: 438) points out, it is "perhaps the most profound contribution" of Category theory to "the history of general mathematical ideas." Awodey (2010: 207) also remarks that it is "a concept of fundamental logical and mathematical importance that is not captured elsewhere in mathematics." Insofar as we are concerned, Adjunction will find its place in our answers to all the three questions raised at the end of §6.1. It is *the* notion that has inspired the application of Category theory in this dissertation.

Definition 6.3.3.1. Let $\mathscr{A} \underset{G}{\overset{F}{\rightleftarrows}} \mathscr{B}$ be Categories and Functors. We say that F is *left adjoint* to G and G is *right adjoint* to F, and write $F \dashv G$, if there is a hom-set isomorphism

$$\mathscr{B}(F(A), B) \cong \mathscr{A}(A, G(B)) \tag{6.3.3.1}$$

that is natural in $A \in \mathscr{A}$ and $B \in \mathscr{B}$. This situation is called an *adjoint situation* or *Adjunction*.

Note 6.3.3.1. The *natural* in Definition 6.3.3.1 refers to a natural isomorphism between two bifunctors (Appendix E.5) extended from the two hom-sets in (6.3.3.1), pictorially

$$\begin{tikzcd} \mathscr{A}^{\mathrm{op}} \times \mathscr{B} \arrow[r, bend left, "\mathscr{B}(F(-),-)"] \arrow[r, bend right, "\mathscr{A}(-,G(-))"'] \arrow[r, phantom, "\Updownarrow\theta"] & \mathbf{Set} \end{tikzcd} . \tag{6.3.3.2}$$

An Adjunction can thus be defined as a triple $\langle F, G, \theta \rangle$. Note that both bifunctors in Definition 6.3.3.1 are *contravariant* (Appendix E.7) in the first component. Abstracting away from technical details, we can understand (6.3.3.1) as a set-theoretic bijection, whereby any arrow from $F(A)$ to B in $\mathscr{B}$ is matched uniquely by an arrow from A to $G(B)$ in $\mathscr{A}$, pictorially

$$\begin{tikzcd} \mathscr{A} \arrow[r, bend left, "F"] \arrow[r, phantom, "\perp"] & \mathscr{B} \arrow[l, bend left, "G"] \\ A \arrow[d, "f"'] \arrow[r, phantom, "\rotatebox{-90}{$\cong$}"] & F(A) \arrow[d, "g"] \\ G(B) & B \end{tikzcd} .$$

In sum, the basic data for an Adjunction are two Categories, two Functors between them (one in each direction), and an interchange of their objects and arrows via these Functors.

Example 6.3.3.1 (free/forgetful Functors). The forgetful Functor U in Example 6.3.2.1 has a left adjoint $F : \mathbf{Set} \to \mathbf{Pos}$, which sends sets to *free posets* based on them, pictorially

$$\begin{tikzcd}
\mathbf{Set} \arrow[rr, bend left, "F"] \arrow[rr, phantom, "\perp"] & & \mathbf{Pos} \arrow[ll, bend left, "U"] \\
X \arrow[rr, "F(X)"] \arrow[d, "f"'] \arrow[rrd, phantom, "\wr\|"] & & \langle X, = \rangle \arrow[d, "g"] \\
P & & \langle P, \leq \rangle \arrow[ll, "U(\langle P, \leq \rangle)"]
\end{tikzcd} .$$

The free poset $\langle X, = \rangle$ over a set X is X with an equality relation. It is *free* in the sense that it has just enough structure to qualify as a poset. Functors mapping sets (or other basic mathematical constructs) to freely generated structured sets (resp. other free objects), called *free Functors*, are generally left-adjoint to the corresponding forgetful Functors.

Example 6.3.3.2 (Spivak 2014). Let $\mathscr{A}$ be the Category of an adult's words (i.e., sound-meaning pairs) and $\mathscr{B}$ be the Category of a baby's noises (i.e., sounds without meanings). The "translation" between baby language and adult language can be viewed as an Adjunction, with the left adjoint $F : \mathscr{B} \to \mathscr{A}$ mapping the baby's noises to free words that pair a sound with a meaning placeholder μ and the right adjoint $U : \mathscr{A} \to \mathscr{B}$ mapping an adult word to its sound part by forgetting its meaning. This yields a pair of free and forgetful Functors, pictorially

$$\begin{tikzcd}
\mathscr{B} \arrow[rr, bend left, "F=\langle -, \mu \rangle"] \arrow[rr, phantom, "\perp"] & & \mathscr{A} \arrow[ll, bend left, "U"] \\
B \arrow[rr, "F(B)"] \arrow[d, "f"'] \arrow[rrd, phantom, "\wr\|"] & & \langle B, \mu \rangle \arrow[d, "g"] \\
A & & \langle A, \mu_A \rangle \arrow[ll, "U(\langle A, \mu_A \rangle)"]
\end{tikzcd} .$$

The hom-set bijection can be understood as the adult's deciphering of the baby's noises by matching the free words against the correct words (by comparing sounds) and the baby's acquisition of pronunciation by emulating the sounds of adult words.

Remark 6.3.3.1. The adult and the baby in Spivak's example cannot understand each other, for there is no deciphering/acquisition of meaning alongside that of sound (so the baby is like a parrot). That said, Spivak provides a preliminary Categorical modeling of language acquisition that may be further developed.

Definition 6.3.3.2 (unit/co-unit)**.** Take an Adjunction $\mathscr{A} \underset{G}{\overset{F}{\underset{\perp}{\rightleftarrows}}} \mathscr{B}$. Let $A = G(B)$ in (6.3.3.1) and obtain for any $B \in \mathscr{B}$ a family of components for the natural isomorphism θ in (6.3.3.2)

$$\theta_{G(B),B} : \mathscr{B}(F(G(B)), B) \cong \mathscr{A}(G(B), G(B))$$

This sends the identity arrow $1_{G(B)}$ in $\mathscr{A}$ to a special arrow $\varepsilon_B : F(G(B)) \to B$ in $\mathscr{B}$. Since B can freely vary in $\mathscr{B}$, we have a natural transformation

$$\varepsilon : F \circ G \Rightarrow \mathrm{Id}_{\mathscr{B}} \tag{6.3.3.3}$$

called the *co-unit* of the Adjunction. Dually, we can let $B = F(A)$ in (6.3.3.1) and obtain for any $A \in \mathscr{A}$ a family of θ-components

$$\theta_{A,F(A)} : \mathscr{B}(F(A), F(A)) \cong \mathscr{A}(A, G(F(A))),$$

which sends the identity arrow $1_{F(A)}$ in $\mathscr{B}$ to another special arrow $\eta_A : A \to G(F(A))$ in $\mathscr{A}$. Since A can freely vary in $\mathscr{A}$, we have another natural transformation

$$\eta : \mathrm{Id}_{\mathscr{A}} \Rightarrow G \circ F \tag{6.3.3.4}$$

called the *unit* of the Adjunction. An Adjunction can be defined as a quadruple $\langle F, G, \eta, \varepsilon \rangle$.

Note 6.3.3.2. The two definitions of an Adjunction, $\langle F, G, \theta \rangle$ and $\langle F, G, \eta, \varepsilon \rangle$, are equivalent.

Unit and co-unit are important notions in that they provide a way to formally compare Categories. Depending on the nature of (6.3.3.3) and (6.3.3.4), there are three graded levels of similarity between two Categories.

Definition 6.3.3.3 (Category comparison)**.** Given Categories $\mathscr{A}$, $\mathscr{B}$ and Functors $F : \mathscr{A} \to \mathscr{B}$, $G : \mathscr{B} \to \mathscr{A}$:

1. If $F \circ G = \mathrm{Id}_{\mathscr{B}}$ and $\mathrm{Id}_{\mathscr{A}} = G \circ F$, then $\mathscr{A}$ and $\mathscr{B}$ are *isomorphic* ($\mathscr{A} \cong \mathscr{B}$).
2. If $F \circ G \Leftrightarrow \mathrm{Id}_{\mathscr{B}}$ and $\mathrm{Id}_{\mathscr{A}} \Leftrightarrow G \circ F$, then $\mathscr{A}$ and $\mathscr{B}$ are *equivalent* ($\mathscr{A} \simeq \mathscr{B}$).
3. If $F \circ G \Rightarrow \mathrm{Id}_{\mathscr{B}}$ and $\mathrm{Id}_{\mathscr{A}} \Rightarrow G \circ F$, then $\mathscr{A}$ and $\mathscr{B}$ are in an Adjunction ($F \dashv G$).

These scenarios can be illustrated by the picture

$$\begin{tikzcd}
\mathscr{A} \arrow[r, bend left, "F"] \arrow[r, phantom, "\perp"] & \mathscr{B} \arrow[l, bend left, "G"]
\end{tikzcd}$$

$$\begin{tikzcd}
 & A \arrow[r, "F"] \arrow[dl, "\eta_A"'] \arrow[d, dashed, "f"] & F(A) \arrow[loop right, "1_{F(A)}"] \arrow[d, dashed, "g"] \arrow[dr, "F(f)"] & \\
G(F(A)) \arrow[dr, "G(g)"] & G(B) \arrow[loop left, "1_{G(B)}"] & B \arrow[l, "G"] & F(G(B)) \arrow[dl, "\varepsilon_B"] \\
 & & &
\end{tikzcd} , \tag{6.3.3.5}$$

where the two arrows η_A and ε_B may be equalities (whence $\mathscr{A} \cong \mathscr{B}$), isomorphisms (whence $\mathscr{A} \simeq \mathscr{B}$), or merely one-sided arrows (whence $F \dashv G$).[24]

Note 6.3.3.3. If $\mathscr{A}$ and $\mathscr{B}$ are poset Categories, then $\mathscr{A} \cong \mathscr{B}$ iff $\mathscr{A} \simeq \mathscr{B}$, because in a poset, due to antisymmetry, any two elements that are isomorphic must actually be the same.

Remark 6.3.3.2. Isomorphism is the strongest similarity between Categories, equivalence of Categories is the second best, and Adjunction is the weakest (though still better than having nothing). As we will see, the cross-MPOS parallelism involves a special Adjunction that instantiates an enhanced weak similarity. However, to formulate it we need a few more definitions.

Definition 6.3.3.4. Let $\mathscr{D}$ be a full Subcategory (Note 6.3.2.1) of $\mathscr{C}$ and write $\iota : \mathscr{D} \to \mathscr{C}$ for the inclusion Functor (Appendix E.3). We call $\mathscr{D}$ a *reflective Subcategory* of $\mathscr{C}$ if ι has a left adjoint $L : \mathscr{C} \to \mathscr{D}$ (called a *reflector*); pictorially

$$\begin{tikzcd}
\mathscr{C} \arrow[r, bend left, "L"] \arrow[r, phantom, "\perp"] & \mathscr{D} \arrow[l, bend left, "\iota"]
\end{tikzcd}$$

$$\begin{tikzcd}
 & C \arrow[dl, "\eta_C"'] \arrow[d, dashed, "f"] & L(C) \arrow[loop right, "1_{L(C)}"] \arrow[d, dashed, "g"'] \arrow[dr, "L(f)"] & \\
\iota(L(C)) \arrow[dr, "\iota(g)"] & \iota(D) \arrow[loop left, "1_{\iota(D)}"] & D & L(\iota(D)) \arrow[dl, "\varepsilon_D"] \\
 & & &
\end{tikzcd} ,$$

Note 6.3.3.4. In the above situation, the unit $\eta : \mathrm{Id}_{\mathscr{C}} \Rightarrow \iota \circ L$ is a normal natural transformation, but the co-unit $\varepsilon : L \circ \iota \Leftrightarrow \mathrm{Id}_{\mathscr{D}}$ is a natural isomorphism, for $D \cong L(\iota(D)) = L(D)$.

Remark 6.3.3.3. Definition 6.3.3.4 can be further generalized. We can replace the inclusion Functor ι by any fully faithful Functor (Definition 6.3.2.2) and accordingly replace the reflective Subcategory by any Category isomorphic to it without losing the reflective property.

[24] The dashed arrows f and g are in one-to-one correspondence; i.e., for each f there is a unique g and vice versa.

Definition 6.3.3.5 (Galois connection). Given two poset Categories $\mathscr{P}_{\leq}$ and $\mathscr{Q}_{\sqsubseteq}$ and two monotone functions qua Functors $f : \mathscr{P}_{\leq} \to \mathscr{Q}_{\sqsubseteq}$ and $g : \mathscr{Q}_{\sqsubseteq} \to \mathscr{P}_{\leq}$. We have an Adjunction $f \dashv g$, commonly called a *Galois connection*, between $\mathscr{P}_{\leq}$ and $\mathscr{Q}_{\sqsubseteq}$ if

$$f(p) \sqsubseteq q \Leftrightarrow p \leq g(q) \tag{6.3.3.6}$$

for any $p \in \mathscr{P}_{\leq}$ and $q \in \mathscr{Q}_{\sqsubseteq}$. Since a hom-set in a poset Category contains at most one arrow, the two hom-sets $\mathscr{Q}_{\sqsubseteq}(f(p), q)$ and $\mathscr{P}_{\leq}(p, g(q))$ must be either both singleton or both empty.

Note 6.3.3.5. The unit and co-unit components in poset Categories are simply order relation instances. In the above Galois connection, pictorially

$$\begin{tikzcd}
\mathscr{P}_{\leq} \arrow[r, bend left, "f"] \arrow[r, phantom, "\perp"] & \mathscr{Q}_{\sqsubseteq} \arrow[l, bend left, "g"]
\end{tikzcd}$$

$$\begin{tikzcd}
 & p \arrow[dl, "\eta_p"'] \arrow[d, dashed] & & f(p) \arrow[loop right, "1_{L(C)}"] \arrow[dr] \arrow[d, dashed] & \\
g(f(p)) \arrow[dr] & & & & f(g(q)) \arrow[dl, "\varepsilon_q"] \\
 & g(q) \arrow[loop left, "1_{g(q)}"] & & q &
\end{tikzcd}$$

the co-unit and unit are respectively

$$\begin{aligned}
&\varepsilon : f \circ g \Rightarrow \mathrm{Id}_{\mathscr{Q}_{\sqsubseteq}} \quad \text{and} \\
&\eta : \mathrm{Id}_{\mathscr{P}_{\leq}} \Rightarrow g \circ f \,,
\end{aligned}$$

whose components at p, q are

$$\begin{aligned}
&\varepsilon_q : f(g(q)) \sqsubseteq q \quad \text{and} \\
&\eta_p : p \leq g(f(p)) \,.
\end{aligned}$$

Informally, this says that after a round trip via the Galois connection, q becomes no larger than before while p becomes no smaller than before.

Remark 6.3.3.4. Intuitively, a Galois connection pairs up elements from two posets that may be very different. And for posets with all joins/meets (Appendix C.3), we can actually compute either adjoint Functor from the other via a specialized version of the Adjoint Functor theorem (first proposed in Freyd 1964).

Theorem 6.3.3.1 (Adjoint Functor theorem for posets). *Suppose $\mathscr{P}_{\leq}$ and $\mathscr{Q}_{\sqsubseteq}$ are in a Galois connection, and suppose $\mathscr{P}_{\leq}$ has all joins. Then given a monotone function qua Functor $f : \mathscr{P}_{\leq} \to \mathscr{Q}_{\sqsubseteq}$, its right adjoint $g : \mathscr{Q}_{\sqsubseteq} \to \mathscr{P}_{\leq}$ is uniquely determined by the formula below: for any $q \in \mathscr{Q}_{\sqsubseteq}$*

$$g(q) = max\{p \in \mathscr{P}_{\leq} \mid f(p) \sqsubseteq q\} \,; \tag{6.3.3.7}$$

that is, $g(q)$ is the maximum among all such $\mathscr{P}_{\leq}$ elements that make $f(p) \sqsubseteq q$ hold.

Dually, if $\mathscr{Q}_{\sqsubseteq}$ has all meets, then given a monotone function qua Functor $g : \mathscr{Q}_{\sqsubseteq} \to \mathscr{P}_{\leq}$, its left adjoint $f : \mathscr{P}_{\leq} \to \mathscr{Q}_{\sqsubseteq}$ is uniquely determined by the formula below: for any $p \in \mathscr{P}_{\leq}$

$$f(p) = min\{q \in \mathscr{Q}_{\sqsubseteq} \mid p \leq g(q)\}\,;$$

that is, $f(p)$ is the minimum among all such $\mathscr{Q}_{\sqsubseteq}$ elements that make $p \leq g(q)$ hold.

The reflective situation (Definition 6.3.3.4) also has a poset version, where a reflective Subcategory is simply a special subposet (or a poset isomorphic to such a subposet considering Remark 6.3.3.3). This special subposet has some nice properties that enable us to describe the adjoint situation more accurately. The following notions are from Erné (2004).

Definition 6.3.3.6 (epi-Adjunction)**.** Given a Galois connection $f \dashv g$ between poset Categories $\mathscr{P}_{\leq}$ and $\mathscr{Q}_{\sqsubseteq}$, pictorially

$$\begin{tikzcd} \mathscr{P}_{\leq} \arrow[r, bend left, "f"] \arrow[r, phantom, "\perp"] & \mathscr{Q}_{\sqsubseteq} \arrow[l, bend left, "g"] \end{tikzcd},$$

if $\mathscr{Q}_{\sqsubseteq}$ is isomorphic to a subposet of $\mathscr{P}_{\leq}$, then the left adjoint f is surjective, and the situation is called a *right perfect Galois connection* or *epi-Adjunction*. The composite Functor $g \circ f$ is called a *closure operation*, for it maps $\mathscr{P}_{\leq}$ to its *fixpoint set*; that is, applying $g \circ f$ twice yields the same result as applying it once. All these notions can be dualized, to a *left perfect Galois connection* or *mono-Adjunction*, a *kernel* or *interior operation*, and another *fixpoint set*.

Note 6.3.3.6. The unit and co-unit in an epi- (resp. mono-) Adjunction behave in the same way as those in a reflective situation, with the unit η being a normal natural transformation and the co-unit ε being a natural isomorphism, pictorially

$$\begin{tikzcd} \mathscr{P}_{\leq} \arrow[r, bend left, "f"] \arrow[r, phantom, "\perp"] & \mathscr{Q}_{\sqsubseteq} \arrow[l, bend left, "g"] \end{tikzcd}$$

$$\begin{tikzcd} & p \arrow[dl, "\eta_p"'] \arrow[d, dashed] & & f(p) \arrow[loop right, "1_{L(C)}"] \arrow[dr] \arrow[d, dashed] & \\ g(f(p)) \arrow[dr] & & & & f(g(q)) \\ & g(q) \arrow[loop left, "1_{g(q)}"] & & q \arrow[ur, "\varepsilon_q"'] & \end{tikzcd},$$

where the two components η_p and ε_q effectively say that after the Galois round trip, p may grow larger, while q goes back to an object isomorphic to itself, which in a poset is simply itself.

Remark 6.3.3.5. A reflective situation, or epi-Adjunction, can be viewed as a section-retraction situation (Definition 6.3.1.8) at the Functorial level, where the fully faithful Functor (right adjoint) is like a section, the surjective "reflector" (left adjoint) is like a retraction, the smaller Cat-

egory isomorphic to a reflective Subcategory is like a retract, and the closure operation is like an idempotent mapping.

6.3.4 Yoneda embedding

The last Categorical notion I will directly use is the Yoneda embedding, which is the basis for the Yoneda lemma (named after the Japanese mathematician Nobuo Yoneda). The Yoneda lemma may be "the single most used result" of Category theory according to Awodey (2010: 185), but since we will not need it, here I only introduce the more basic notion of Yoneda embedding.

Observe that the hom-functor $\mathscr{C}(-,-) : \mathscr{C}^{\mathrm{op}} \times \mathscr{C} \to \mathbf{Sets}$ (Appendix E.8) can be curried in two ways, either as

$$k : \mathscr{C}^{\mathrm{op}} \to \mathbf{Set}^{\mathscr{C}}$$

or as

$$y : \mathscr{C} \to \mathbf{Set}^{\mathscr{C}^{\mathrm{op}}} . \tag{6.3.4.1}$$

It is this second Functor y, which sends $\mathscr{C}$-objects to contravariant set-valued Functors (called *presheaves* on $\mathscr{C}$), that is used to define the Yoneda embedding.

Note 6.3.4.1. $\mathbf{Set}^{\mathscr{C}^{\mathrm{op}}}$ in (6.3.4.1) is a Functor Category (Remark 6.3.2.2) with all Functors from $\mathscr{C}^{\mathrm{op}}$ to **Set** as objects and natural transformations between them as arrows.

Definition 6.3.4.1. A Yoneda embedding $y : \mathscr{C} \to \mathbf{Set}^{\mathscr{C}^{\mathrm{op}}}$ sends each $\mathscr{C}$-object C to the presheaf

$$y(C) = \mathscr{C}(-, C) : \mathscr{C}^{\mathrm{op}} \to \mathbf{Set}$$

and each $\mathscr{C}$-arrow $f : C \to D$ to the natural transformation

$$y(f) = \mathscr{C}(-, f) : \mathscr{C}(-, C) \Rightarrow \mathscr{C}(-, D) .$$

Note 6.3.4.2. Recall from Note 6.3.2.1 that a Functor is an embedding if it is fully faithful and injective on objects. The Yoneda embedding is just such a Functor. This means that $\mathscr{C}$ is isomorphic to a full Subcategory of $\mathbf{Set}^{\mathscr{C}^{\mathrm{op}}}$; that is, (up to isomorphism) each $\mathscr{C}$-object is determined by a Functor $\mathscr{C}(-, C)$, which is the partly applied hom-functor sending any $\mathscr{C}$-object to the hom-set from that object to C.

Remark 6.3.4.1. Mathematicians speak of the "Yoneda philosophy," which is often conveyed in the Dutch proverb "tell me who your friends are, and I will tell you who you are" (*zeg me wie uw vrienden zijn, dan zeg ik wie u bent*; see, e.g., Boisseau & Gibbons 2018). Such "externalist" methodology is prevalent in Category theory.

Example 6.3.4.1 (Yoneda embedding for posets)**.** Given a poset Category $\mathscr{P}_{\leq}$ and an object X in it, since for any W in $\mathscr{P}_{\leq}$, the hom-set $\mathscr{P}(W, X)$ is either singleton or empty, the hom-functor

$\mathscr{P}(-, X)$ effectively divides $\mathscr{P}_{\leq}$ into two parts—one part "below" X (inclusively) and the other "not below" it (exclusively).[25] The set-valued Functor $\mathscr{P}(-, X)$ picks out X's below-set, called its *principal lower set* and denoted by $\downarrow X$, and the Yoneda Functor y embeds $\mathscr{P}_{\leq}$ into the set of all its lower sets, called its *lower set completion*. Thus, any object in $\mathscr{P}_{\leq}$ is fully determined by its principal lower set; namely, the set of all $\mathscr{P}_{\leq}$-objects lower than it—or in more linguistically friendly terms, the set of all $\mathscr{P}_{\leq}$-objects in its scope.[26] As we will see, this provides a Category-theoretic formulation of the mergeme template.

This is the end of the introduction to Category theory.

6.4 Cross-major-part-of-speech parallelism

With the Categorical notions laid out, next I will apply them to the global-level questions about the SCS raised in §6.1, beginning with the cross-MPOS parallelism.

6.4.1 No direct connection

Recall that a categorial sequence is defined as a poset. Now we can further view it as a poset Category. Take the expository c-sequences in (8) for example.

(8) a. $\mathrm{A_V} := \mathrm{V} \leq \mathrm{Appl} \leq v \leq \mathrm{Asp} \leq \mathrm{Tns} \leq \mathrm{Mod} \leq \mathrm{Fin} \leq \mathrm{Foc} \leq \mathrm{Top} \leq \mathrm{SA}$

b. $\mathrm{A_N} := \mathrm{N} \leq \mathrm{Gen} \leq n \leq \mathrm{Cl} \leq \mathrm{Num} \leq \mathrm{Q} \leq \mathrm{Det} \leq \mathrm{K}$

These give rise to two poset Categories; call them $\mathscr{A}_{\mathbf{V}}$ and $\mathscr{A}_{\mathbf{N}}$. I approach the parallelism between them by asking how similar the two Categories are; namely, whether they are isomorphic, equivalent, or connected by Adjunction. Isomorphism can be ruled out right away, for $\mathrm{A_V}$ and $\mathrm{A_N}$ as posets have different cardinalities, which means it is impossible to construct a bijection between them.[27] And since two poset Categories are isomorphic iff they are equivalent (Note 6.3.3.3), equivalence of Categories is ruled out as well. Thus, if a general inter-c-sequence connection can be constructed at all, it must be an Adjunction (i.e., a Galois connection).[28]

To construct a Galois connection between $\mathscr{A}_{\mathbf{V}}$ and $\mathscr{A}_{\mathbf{N}}$, we first need to specify a pair of monotone functions between them. Mathematically, there are a multitude of choices. For ex-

[25] More precisely, the Yoneda Functor "factors through the embedding of $\mathbf{2}^{\mathscr{C}^{\mathrm{op}}}$ in $\mathbf{Set}^{\mathscr{C}^{\mathrm{op}}}$" (Andrew Pitts, p.c.), where $\mathbf{2}$ is a full Subcategory of **Set** comprising the empty set (call it 0) and a singleton set (call it 1). For any $X \in \mathscr{P}_{\leq}$, there is a $\mathbf{2}$-valued Functor (the "factor" in Pitts' words) that sends all $\mathscr{P}_{\leq}$-objects below or equal to X to 1 and all other $\mathscr{P}_{\leq}$-objects to 0. Thus, $\mathbf{2}$ effectively defines a subset of the poset $\langle P, \leq \rangle$ via the property below-or-equal-to-X.

[26] See Partee et al. (1993: 285–286) for an order-theoretic formulation of the same idea, in *ideal representation* terms.

[27] Recall from §6.1.1 that direct bijections exist between special c-sequences; however, since the cross-MPOS parallelism intuitively exists for any c-sequences, a connection that only holds for special ones is undesirable.

[28] I use the terms *Adjunction* and *Galois connection*, and accordingly *Functor* and *monotone function*, interchangeably in the context of poset Categories.

ample, we can send all $\mathscr{A}_{\mathbf{V}}$-objects to a single $\mathscr{A}_{\mathbf{N}}$-object (e.g., N) and send all $\mathscr{A}_{\mathbf{V}}$-arrows to the identity arrow on 1_{N}, as in (6.4.1.1).[29] This is a constant Functor (Appendix E.4).

$$\begin{tikzcd}
\mathscr{A}_{\mathbf{V}} \arrow[d] & \mathrm{V} \arrow[r] \arrow[d, mapsto] & \mathrm{Appl} \arrow[r] \arrow[dl, mapsto] & v \arrow[r] \arrow[dll, mapsto] & \mathrm{Asp} \arrow[r] \arrow[dlll, mapsto] & \mathrm{T} \arrow[r] \arrow[dllll, mapsto] & \mathrm{Mod} \arrow[r] \arrow[dlllll, mapsto] & \mathrm{Fin} \arrow[r] \arrow[dllllll, mapsto] & \mathrm{Foc} \arrow[r] \arrow[dlllllll, mapsto] & \mathrm{Top} \arrow[r] \arrow[dllllllll, mapsto] & \mathrm{SA} \arrow[dlllllllll, mapsto] \\
\mathscr{A}_{\mathbf{N}} & \mathrm{N} \arrow[r] & \mathrm{Gen} \arrow[r] & n \arrow[r] & \mathrm{Cl} \arrow[r] & \mathrm{Num} \arrow[r] & \mathrm{Q} \arrow[r] & \mathrm{Det} \arrow[r] & \mathrm{K} & &
\end{tikzcd} \tag{6.4.1.1}$$

The same can be done in the other direction, for example, by sending all $\mathscr{A}_{\mathbf{N}}$-objects to SA and all $\mathscr{A}_{\mathbf{N}}$-arrows to 1_{SA}. The two constant Functors are clearly well defined. However, neither is linguistically interesting, and so are most other Functors between $\mathscr{A}_{\mathbf{V}}$ and $\mathscr{A}_{\mathbf{N}}$. Remember that our task is not to find just any Functor pair, but to find a linguistically interesting pair. Since Functors map all objects in the source Category to certain objects in the target Category, this requires us to select the "certain objects" in a linguistically interesting way; namely, they need to have some special status in the target c-sequence. An obvious solution, as suggested in §6.1, is mapping to core categories, such as phase categories or core functional categories. We can prove that maps between $\mathscr{A}_{\mathbf{V}}$ and $\mathscr{A}_{\mathbf{N}}$ sending objects to core categories are indeed Functorial.

Proposition 6.4.1.1. *There are Functors $f : \mathscr{A}_{\mathbf{V}} \to \mathscr{A}_{\mathbf{N}}$ and $g : \mathscr{A}_{\mathbf{N}} \to \mathscr{A}_{\mathbf{V}}$ between c-sequence Categories $\mathscr{A}_{\mathbf{V}}$ and $\mathscr{A}_{\mathbf{N}}$ that send all categories in one c-sequence to core categories in the other.*

Proof. Suppose the core categories in $\mathscr{A}_{\mathbf{V}}$ are v and Fin, and those in $\mathscr{A}_{\mathbf{N}}$ are n and Det (these choices may vary). We can define f and g for objects $x \in \mathscr{A}_{\mathbf{V}}$ and $y \in \mathscr{A}_{\mathbf{N}}$ as follows (arrows are mapped according to the order relations):

$$\begin{aligned} f(x) &= \begin{cases} n & \text{for } x \in \{\mathrm{V}, \mathrm{Appl}, v\} \\ \mathrm{Det} & \text{for } x \in \{\mathrm{Asp}, \mathrm{Tns}, \mathrm{Mod}, \mathrm{Fin}, \mathrm{Foc}, \mathrm{Top}, \mathrm{SA}\} \end{cases}, \\ g(y) &= \begin{cases} v & \text{for } y \in \{\mathrm{N}, \mathrm{Gen}, n\} \\ \mathrm{Fin} & \text{for } y \in \{\mathrm{Cl}, \mathrm{Num}, \mathrm{Q}, \mathrm{Det}, \mathrm{K}\} \end{cases}, \end{aligned} \tag{6.4.1.2}$$

pictorially[30]

$$\begin{tikzcd}
\mathscr{A}_{\mathbf{V}} \arrow[d, bend right, "f"] & \mathrm{V} \arrow[r] \arrow[dr, mapsto] & \mathrm{Appl} \arrow[r] \arrow[dr, mapsto] & \boxed{v} \arrow[r] \arrow[d, dashed, mapsto] & \mathrm{Asp} \arrow[r] \arrow[dr, mapsto] & \mathrm{T} \arrow[r] \arrow[dr, mapsto] & \mathrm{Mod} \arrow[r] \arrow[dr, mapsto] & \boxed{\mathrm{Fin}} \arrow[r] \arrow[d, dashed, mapsto] & \mathrm{Foc} \arrow[r] \arrow[dl, mapsto] & \mathrm{Top} \arrow[r] \arrow[dll, mapsto] & \mathrm{SA} \arrow[dll, mapsto] \\
\mathscr{A}_{\mathbf{N}} \arrow[u, bend right, "g"] & \mathrm{N} \arrow[r] & \mathrm{Gen} \arrow[r] & \boxed{n} \arrow[r] & \mathrm{Cl} \arrow[r] & \mathrm{Num} \arrow[r] & \mathrm{Q} \arrow[r] & \boxed{\mathrm{Det}} \arrow[r] & \mathrm{K}. & &
\end{tikzcd} \tag{6.4.1.3}$$

[29] I have omitted the arrow mappings to reduce clutter.

[30] Here and below I use boxes and dashed arrows to highlight the invertibly associated syntactic categories in an inter-c-sequence connection. These are the unequivocally parallel categories recognized by the connection.

It is easy to verify that for any $x, x' \in \mathscr{A}_{\mathbf{V}}$, we have

$$x \leq x' \Rightarrow f(x) \leq f(x')\,.$$

For instance, $v \leq$ Asp $\Rightarrow f(v) \leq f(\text{Asp})$ (i.e., $n \leq$ Det). Therefore, f is a monotone function and so a Functor between poset Categories. Similarly, we can verify that g is also a Functor. □

As (6.4.1.3) shows, the inter-c-sequence connection given by f and g is divided into three zones, which incidentally reflects two intuitions: (i) c-sequences are partitioned into sub-c-sequences; (ii) the core categories themselves are parallel. The Categories and Functors in (6.4.1.3) together form the configuration needed in Category comparison:

$$\mathscr{A}_{\mathbf{V}} \underset{g}{\overset{f}{\rightleftarrows}} \mathscr{A}_{\mathbf{N}}\,. \tag{6.4.1.4}$$

Do f and g form an Adjunction? Recall from Definition 6.3.3.5 that to verify this, we only need to check if the inequalities

$$x \leq g(f(x)) \quad \text{and} \quad f(g(y)) \leq y \tag{6.4.1.5}$$

hold for all $x \in \mathscr{A}_{\mathbf{V}}$ and $y \in \mathscr{A}_{\mathbf{N}}$; namely, if all $\mathscr{A}_{\mathbf{V}}$-objects become no smaller after the closure round trip $g \circ f$ and all $\mathscr{A}_{\mathbf{N}}$-objects become no larger after the interior round trip $f \circ g$ (Definition 6.3.3.6). If both are true, then we have the natural transformations

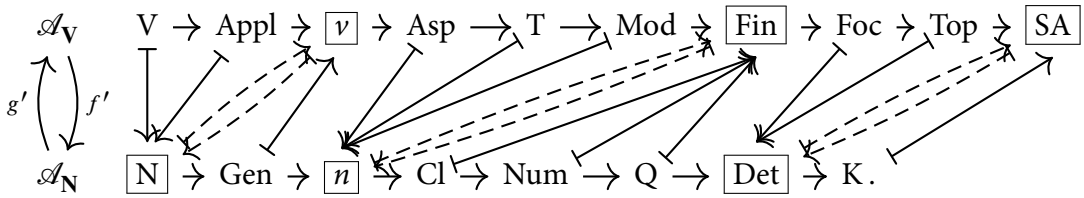
$$\mathrm{Id}_{\mathscr{A}_{\mathbf{V}}} \overset{\eta}{\Rightarrow} g \circ f \quad \text{and} \quad f \circ g \overset{\varepsilon}{\Rightarrow} \mathrm{Id}_{\mathscr{A}_{\mathbf{N}}}\,, \tag{6.4.1.6}$$

which are exactly the unit and co-unit defining an Adjunction.

However, a quick verification shows that neither inequality in (6.4.1.3) holds. For instance, Foc $\geq g(f(\text{Foc})) =$ Fin, and $f(g(\text{Cl})) =$ Det $\geq$ Cl. In fact, we have $x \geq g(f(x))$ for any x above Fin and $f(g(y)) \geq y$ for any y below Det. Hence, the Functors f and g, though linguistically meaningful, fail to form an Adjunction. One may wish to fix things by tweaking the choices of core categories. For instance, we could add two more core categories SA and N and redefine f, g in (6.4.1.2) as f', g' below:

$$\begin{aligned} f'(x) &= \begin{cases} \text{N} & \text{for } x \in \{\text{V, Appl, } v\} \\ n & \text{for } x \in \{\text{Asp, Tns, Mod, Fin}\} \\ \text{Det} & \text{for } x \in \{\text{Foc, Top, SA}\}\,, \end{cases} \\ g'(y) &= \begin{cases} v & \text{for } y \in \{\text{N, Gen}\} \\ \text{Fin} & \text{for } y \in \{n, \text{Cl, Num, Q}\} \\ \text{SA} & \text{for } y \in \{\text{Det, K}\}\,, \end{cases} \end{aligned} \tag{6.4.1.7}$$

pictorially



(6.4.1.8)

We can verify that this configuration does satisfy (6.4.1.5); hence $f' \dashv g'$, pictorially

$$\mathscr{A}_{\mathbf{V}} \underset{g'}{\overset{f'}{\underset{\longleftarrow}{\overset{\longrightarrow}{\perp}}}} \mathscr{A}_{\mathbf{N}}\,. \tag{6.4.1.9}$$

However, this Adjunction is counterintuitive, for it systematically skews the core category associations leftward, which is not what we know about world languages (e.g., it says that "v is parallel with N"). Perhaps we can find better Adjunctions by further tweaking the Functors, but that is not worthwhile, because two general problems render any direct inter-c-sequence Adjunction linguistically uninteresting. First, since the closure round trip must not shrink objects, the greatest $\mathrm{A}_{\mathbf{v}}$-element SA must be a core category (otherwise it would shrink); dually, since the interior round trip must not enlarge objects, the least $\mathrm{A}_{\mathbf{v}}$-element N must also be a core category (otherwise it would grow). But such obligatory choices are unmotivated, for nothing makes SA and N more special than the other (nonobligatory) core categories.[31] Moreover, if they are indeed special—say, by being extremal—then we may ask why the other two extremal categories, V and K, are not equally special. The answer is clear: SA and N are only special for the Adjunction but not for the SCS ontology.

Second, since the Adjunction in (6.4.1.9) is strictly defined with $f'/\mathscr{A}_{\mathbf{V}}$ on the left and $g'/\mathscr{A}_{\mathbf{N}}$ on the right, the inter-c-sequence connection it establishes is direction-sensitive; that is, the category associations would change if we swap the left/right placement of $\mathscr{A}_{\mathbf{V}}$ and $\mathscr{A}_{\mathbf{N}}$. However, recall from §6.1.1 that the cross-MPOS parallelism is an invertible pairing—whichever c-sequence we pick as the source of comparison, we ought to get the same parallel associations; for instance, $\langle \mathrm{V}, \mathrm{N} \rangle$ when $\mathscr{A}_{\mathbf{V}}$ is placed on the left should not become $\langle \mathrm{V}, n \rangle$ when $\mathscr{A}_{\mathbf{V}}$ is placed on the right, because we are comparing the same two c-sequences after all. Yet this is exactly what happens if we swap $\mathscr{A}_{\mathbf{V}}$ and $\mathscr{A}_{\mathbf{N}}$ in (6.4.1.9) (I omit the detailed depiction for space reasons).

[31] By *nonobligatory* I mean that those core categories may be variably picked depending on one's theoretical or empirical assumptions without affecting the overall Adjunction; e.g., one could choose Mod instead of Fin to be the middle core category in (6.4.1.8).

6.4.2 Connection via universal spine

6.4.2.1 Epi-Adjunction

Since the cross-MPOS parallelism cannot be directly characterized by a suitable Adjunction, we must seek an alternative configuration if we want to compare c-sequences as Categories. It turns out that a suitable configuration can be built via certain part-of-speech-neutral conceptual categories such as those in Wiltschko (2014). Wiltschko puts these on a "universal spine," which for us is just a very short poset (more exactly a chain), as in (9).

(9) {classification ≤ point-of-view ≤ anchoring ≤ linking} (universal spine)

I will call this poset A_0 and view it as a Category $\mathscr{A}_0$.[32] Below I replace $\mathscr{A}_{\mathbf{N}}$ in (6.4.1.9) with $\mathscr{A}_0$:

$$\mathscr{A}_{\mathbf{V}} \underset{e}{\overset{h}{\rightleftarrows}} \mathscr{A}_0 \,, \tag{6.4.2.1}$$

where the Functors h and e are defined as follows:

$$\begin{aligned} h(x) &= \begin{cases} \text{classification} & \text{for } x \in \{\text{V, Appl, } v\} \\ \text{point-of-view} & \text{for } x \in \{\text{Asp}\} \\ \text{anchoring} & \text{for } x \in \{\text{Tns, Mod}\} \\ \text{linking} & \text{for } x \in \{\text{Fin, Foc, Top, SA}\} \end{cases} , \\ e &:= \begin{cases} \text{classification} & \mapsto v \\ \text{point-of-view} & \mapsto \text{Asp} \\ \text{anchoring} & \mapsto \text{Mod} \\ \text{linking} & \mapsto \text{SA} \end{cases} , \end{aligned} \tag{6.4.2.2}$$

or pictorially

$$\begin{tikzcd} \mathscr{A}_{\mathbf{V}} \arrow[d, bend right, "h"] & \text{V} \arrow[r] \arrow[dd, mapsto] & \text{Appl} \arrow[r] \arrow[dd, mapsto] & \boxed{v} \arrow[r] \arrow[dd, dashed, bend left] & \boxed{\text{Asp}} \arrow[r] \arrow[dd, dashed, bend left] & \text{Tns} \arrow[r] \arrow[ddr, mapsto] & \boxed{\text{Mod}} \arrow[r] \arrow[dd, dashed, bend left] & \text{Fin} \arrow[r] \arrow[ddr, mapsto] & \text{Foc} \arrow[r] \arrow[ddr, mapsto] & \text{Top} \arrow[r] \arrow[ddr, mapsto] & \boxed{\text{SA}} \arrow[dd, dashed, bend left] \\ \mathscr{A}_0 \arrow[u, bend right, "e"] & \text{classification} \arrow[r] & \text{point-of-view} \arrow[r] & \text{anchoring} \arrow[r] & \text{linking} \,. & & & & & & \end{tikzcd} \tag{6.4.2.3}$$

The Functor e is full, faithful, and injective on objects, so it is an embedding Functor (Note 6.3.2.1) and more exactly an order embedding (Note 6.3.2.2). It is routine to check that h and e form an Adjunction $h \dashv e$, by checking that they yield the desired unit and co-unit natural transforma-

[32] The motivation for the subscript 0 will become clear by the end of this section.

tions. Specifically, we can verify that the two round-trip relations

$$x \leq e(h(x)) \quad \text{and} \quad h(e(y)) \leq y$$

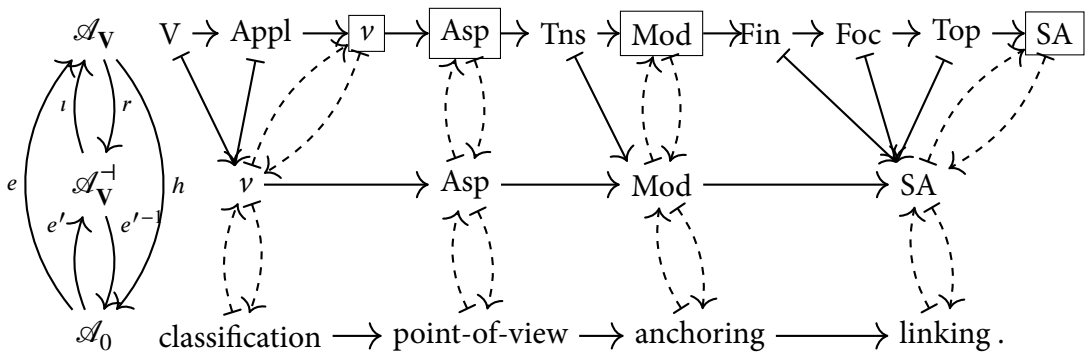
hold for any $x \in \mathscr{A}_{\mathbf{V}}$ and $y \in \mathscr{A}_0$; for instance, Appl $\leq e(h(\text{Appl})) = v$, and $h(e(\text{anchoring})) =$ anchoring. Note that the latter is an equality, which means the co-unit ε is a natural isomorphism (Note 6.3.3.4):

$$h \circ e \overset{\varepsilon}{\Leftrightarrow} \mathrm{Id}_{\mathscr{A}_0} .$$

This makes the Adjunction $h \dashv e$ into an epi-Adjunction (Definition 6.3.3.6), with $\mathscr{A}_0$ being isomorphic to a reflective Subcategory of $\mathscr{A}_{\mathbf{V}}$ (I denote the latter by $\mathscr{A}_{\mathbf{V}}^{\dashv}$ following Smith 2018):

$$\mathscr{A}_0 \cong \mathscr{A}_{\mathbf{V}}^{\dashv} := e(\mathscr{A}_0) = v \leq \text{Asp} \leq \text{Mod} \leq \text{SA} ,$$

pictorially



(6.4.2.4)

Since our $\mathscr{A}_{\mathbf{V}}$ was arbitrarily chosen, this scenario can be generalized into the following lemma:[33]

Lemma 6.4.2.1. *Categorically, the universal spine and any c-sequence of a given granularity level form an epi-Adjunction, with the former being isomorphic to a reflective Subcategory of the latter.*

In Remark 6.3.3.5 we saw that an epi-Adjunction is like a section-retraction situation. Thus, the partition of a c-sequence by a universal spine is analogous to the sorting of US citizens by congressional districts in Example 6.3.1.12 (when there is no empty sort). Specifically, e in (6.4.2.4) (the section) chooses a representative syntactic category for each functional domain, and h (the retraction) sends each category to its domain. Recall from Definitions 6.3.1.8 and 6.3.1.9 that composing a section-retraction pair in one direction yields an identity map and composing them in the other direction yields an idempotent map. In (6.4.2.4), the identity map is $\mathrm{Id}_{\mathscr{A}_0}$ and the idempotent map is the closure operation $e \circ h$. The idempotence is easy to verify: applying the find-core-category-for-some-category's-functional-domain process twice returns the same result as applying it once; for instance, $e \circ h(e \circ h(\text{Appl})) = e \circ h(\text{Appl}) = v$. In a section-retraction situation, there are multiple choices for the section map (Remark 6.3.1.2). I have let e in (6.4.2.1)

[33] The proof for Lemma 6.4.2.1 is just a reinterpretation of $\mathscr{A}_{\mathbf{V}}$ above as an arbitrary c-sequence.

map to the core syntactic categories, which demarcate functional domain boundaries by being the highest category in their respective domain. If it turns out that the functional domains still have other roles—for example, if they correspond to the Chomskyan phases—then those roles will also be reflected in the core categories. In Category-theoretic terms, the core categories are the *reflections* of syntactic categories in general.

In fact, with Lemma 6.4.2.1 in hand, we do not need to define the embedding Functor by ourselves every time we meet a new c-sequence, because in the established Galois connection the left adjoint (i.e., the send-to-functional-domain Functor) is predetermined based on independent factors, notably the C-I function of each syntactic category. Recall from Remark 6.3.3.4 that in a Galois connection, a right adjoint (when it exists) is uniquely determined by its corresponding left adjoint via the Adjoint Functor theorem, more exactly by the poset version given in (6.3.3.7) and repeated below:

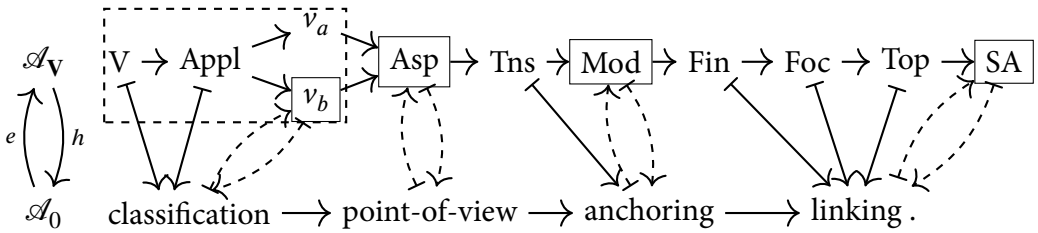
$$g(q) = max\{p \in \mathscr{P}_{\leq} \mid f(p) \sqsubseteq q\}\,.$$

Substituting the parameters in this formula with those from (6.4.2.3), we obtain

$$e(y) = max\{x \in \mathscr{A}_{\mathbf{V}} \mid h(x) \leq y\}\,, \tag{6.4.2.5}$$

which says that $e(y)$ is the maximum (Appendix C.2) of the set of all such $\mathscr{A}_{\mathbf{V}}$-elements that make $h(x) \leq y$ hold. But this is just saying that $e(y)$ is the highest category in its functional domain, which is exactly how we have defined e earlier. As such, the linguistic motivation and the mathematical motivation for our specific definition of e converge, and the Adjoint Functor theorem provides additional mathematical support for our choices of core syntactic categories.

A caveat must be given here regarding flavored categories, which I have set aside in this chapter till now, because all our Category-theoretic discussion requires is that c-sequences be posets, which they all are by definition. The trouble with flavored categories is that category flavoring at certain points in a c-sequence runs afoul of the Adjoint Functor theorem. If a core category, say v, is flavored, then only one of its flavors can be hit by the embedding Functor e, because Functors, like functions, are single-valued. I illustrate this scenario below:



(6.4.2.6)

However, the mapping from classification to v_b cannot be derived from (6.4.2.5), for the classification-domain (or V-domain) subposet (indicated by dashed box) has no maximum; namely, there

is no such element in $\mathscr{A}_{\mathbf{V}}$ that is both an upper bound of the boxed subposet and part of it. On the one hand, since the intuitive cross-MPOS parallelism is not contingent on the (non)existence of category flavoring—much as it is not contingent on the cardinality of c-sequences (§6.1.1)—our inter-c-sequence connection must be generally applicable to c-sequences of all legitimate shapes. On the other hand, since the Adjoint Functor theorem is a well-established mathematical result, that it fails to give us a right adjoint in (6.4.2.6) can hardly be used as evidence to falsify the theorem itself. As such, the most likely conclusion from the above conflict is that the c-sequence in (6.4.2.6) is illegitimate. I state this in the following corollary:

Corollary 6.4.2.1 (of Ajoint Functor Theorem). *The distribution of flavored categories in any c-sequence is strictly limited to noncore syntactic categories.*

Therefore, among others, no phase category can be flavored. This may turn out to have independent linguistic motivation, which I leave to future research.[34] Considering our conclusion in Chapter 5 that flavored categories (as defined in this dissertation) are produced by very special diachronic conditions and so ought to occur scarcely in world languages, this extra restriction, if true, would not be surprising.

6.4.2.2 Flower-shaped parallelism

Lemma 6.4.2.1 captures a general characteristic of the SCS in line with the interaction of template and flexibility at the abstraction layer of single granularity levels: the template is the universal spine, the flexibility space contains the concrete c-sequences that the universal spine can give rise to (in a certain language variety), and the two are connected by epi-Adjunction. For each c-sequence $\mathscr{A}_{\mathbf{X}}$ of part of speech $\mathbf{X}$ at a granularity level A, we have the following configuration:

$$\begin{tikzcd} \mathscr{A}_{\mathbf{X}} \arrow[r, bend left, "h_{\mathbf{X}}"] \arrow[r, phantom, "\perp"] & \mathscr{A}_0 \arrow[l, bend left, "e_{\mathbf{X}}"] \end{tikzcd} . \tag{6.4.2.7}$$

So, for a pair of c-sequences, say $\mathscr{A}_{\mathbf{V}}$ and $\mathscr{A}_{\mathbf{N}}$, we have a joint configuration

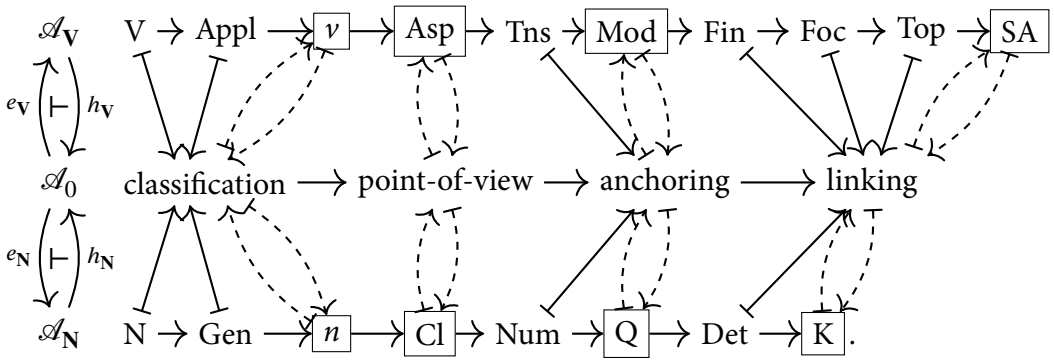
$$\begin{tikzcd} \mathscr{A}_{\mathbf{V}} \arrow[r, bend left, "h_{\mathbf{V}}"] \arrow[r, phantom, "\perp"] & \mathscr{A}_0 \arrow[l, bend left, "e_{\mathbf{V}}"] \arrow[r, bend right, "e_{\mathbf{N}}"'] \arrow[r, phantom, "\perp"] & \mathscr{A}_{\mathbf{N}} \arrow[l, bend right, "h_{\mathbf{N}}"'] \end{tikzcd} \tag{6.4.2.8}$$

consisting of an epi-Adjunction on either side of $\mathscr{A}_0$. This qualifies as an inter-c-sequence connection and can serve as a formal description of the cross-MPOS parallelism.

[34] Theresa Biberauer (p.c.) suggests a possible motivation from an acquisitional perspective: since phase categories are formed early in the successive division of the categorial space, while flavored categories ought to be formed late (as they reference very specific formal features), the successive division algorithm predicts that these two subtypes of syntactic categories do not coincide.

Theorem 6.4.2.1 (cross-MPOS Parallelism). *For any pair of c-sequences at a given granularity level, there is a joint Categorical connection between them via the mediation of a universal spine and two epi-Adjunctions, each comprising an embedding right adjoint and a surjective left adjoint.*

I illustrate this scenario below:



(6.4.2.9)

There are three ways to interpret the above parallelism (let **X** and **Y** be arbitrary MPOSs):

(10) a. (**category-to-category**) Each category in a c-sequence $\mathrm{A}_{\mathbf{X}}$ is parallel with a core category in another c-sequence $\mathrm{A}_{\mathbf{Y}}$ via the composite Functor $e_{\mathbf{Y}} \circ h_{\mathbf{X}}$. For example, Appl in $\mathscr{A}_{\mathbf{V}}$ is parallel with n in $\mathscr{A}_{\mathbf{N}}$ via $e_{\mathbf{N}} \circ h_{\mathbf{V}}$, and Num in $\mathscr{A}_{\mathbf{N}}$ is parallel with Fin in $\mathscr{A}_{\mathbf{V}}$ via $e_{\mathbf{V}} \circ h_{\mathbf{N}}$. This is the most specific description of the cross-MPOS parallelism, which combines insights from (6.4.1.3) and (6.4.1.8) and is rid of their drawbacks.

b. (**category-to-domain**) Each category $c_{\mathbf{X}}$ in a c-sequence $\mathrm{A}_{\mathbf{X}}$ is parallel with an entire categorial domain $C_{\mathbf{Y}}$ in another c-sequence $\mathrm{A}_{\mathbf{Y}}$ via the Functorial configuration $c_{\mathbf{X}} \xrightarrow{h_{\mathbf{X}}} u \xleftarrow{h_{\mathbf{Y}}} C_{\mathbf{Y}}$ (called a *co-span*), where u is a category in the universal spine. For example, Appl is parallel with $\{\mathrm{N} \to \mathrm{Gen} \to n\}$ as $h_{\mathbf{V}}(\mathrm{Appl}) = h_{\mathbf{N}}(\{\mathrm{N} \to \mathrm{Gen} \to n\}) =$ classification, and Num is parallel with $\{\mathrm{Tns} \to \mathrm{Mod}\}$ as $h_{\mathbf{N}}(\mathrm{Num}) = h_{\mathbf{V}}(\{\mathrm{Tns} \to \mathrm{Mod}\}) =$ anchoring. This is a moderate description of the cross-MPOS parallelism, which pairs multiple categories with a single sub-c-sequence (e.g., all of V, Appl, and v are parallel with $\{\mathrm{N}\to\mathrm{Gen}\to n\}$). Such an element-to-subset mapping is impossible in a direct inter-c-sequence connection (qua a first-order function), but it is perfectly feasible in the mediated connection.

c. (**domain-to-domain**) Each categorial domain in a c-sequence $\mathrm{A}_{\mathbf{X}}$ is parallel with a domain in another c-sequence $\mathrm{A}_{\mathbf{Y}}$, also via the co-span in (10b). For example, $\{\mathrm{V} \to \mathrm{Appl} \to v\}$ is parallel with $\{\mathrm{N} \to \mathrm{Gen} \to n\}$ as $h_{\mathbf{V}}(\{\mathrm{V} \to \mathrm{Appl} \to v\}) = h_{\mathbf{N}}(\{\mathrm{N} \to \mathrm{Gen} \to n\}) =$ classification, and $\{\mathrm{Cl}\}$ is parallel with $\{\mathrm{Asp}\}$ as $h_{\mathbf{N}}(\{\mathrm{Cl}\}) = h_{\mathbf{V}}(\{\mathrm{Asp}\}) =$ point-of-view. This is the most general description of the cross-MPOS parallelism and also impossible to formulate in a direct inter-c-sequence connection.

In all three interpretations above, the joint connection is invertible; that is, no matter which c-

sequence we put on which side of the universal spine, we get the same category associations.[35] Also remember that all c-sequences in our foregoing illustration are arbitrarily chosen, so one can freely add, remove, or relabel the specific categories in (6.4.2.9) without losing the Categorical connection. As such, the joint connection is indeed a solution to the question raised in (7a). It is a linguistically meaningful function that can adequately and consistently formalize the cross-MPOS parallelism. The joint connection can be further generalized. We can in principle add more c-sequences to (6.4.2.8) and establish a mediated parallelism across multiple MPOSs (if those exist). The multi-MPOS parallelism is shaped like a flower, as in Figure 6.5.

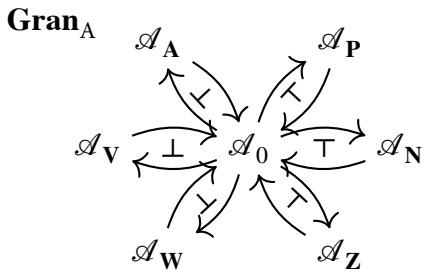


Figure 6.5 Categorical parallelism across multiple MPOSs

In Figure 6.5, the center of the flower is the universal spine, and each petal of it is a c-sequence connected to the center via an epi-Adjunction. If we add in all the c-sequences at a granularity level, we effectively obtain a Category $\mathbf{Gran}_{\mathrm{A}}$ of the granularity level A itself, where each poset Category, be it a c-sequence or a universal spine, is an object and each adjoint Functor is an arrow.

Before ending this section, I highlight three characteristics of $\mathbf{Gran}_{\mathrm{A}}$. First, since there is at most one arrow between each pair of objects in $\mathbf{Gran}_{\mathrm{A}}$, this is a thin Category (Note 6.3.1.2).[36] Second, since there is a unique arrow from every object to $\mathscr{A}_0$ and also a unique arrow from $\mathscr{A}_0$ to every object, $\mathscr{A}_0$ is a zero object. This is why I have notated the universal spine with a subscript 0. Third, since Adjunction resides in the context of Categories and Functors, now that we are viewing Categories and Functors as objects and arrows, we need to update our perspective on the adjoint situation too. This perspective shift is straightforward based on the correspondence between epi-Adjunction and section-retraction situation (Remark 6.3.3.5): each right adjoint in Figure 6.5 (leaving $\mathscr{A}_0$) is a section and each left adjoint (entering $\mathscr{A}_0$) is a retraction. The above Categorification of a single granularity level will be placed in a bigger picture as we move on to the next abstraction layer in §6.5.

[35] Pictorially, this simply turns (6.4.2.9) upside down with all arrows preserved.

[36] Specifically, it can be obtained via a *thinning Functor* (Appendix E.6) from a (tiny) full Subcategory (Note 6.3.2.1) of the Category **Pos** of all posets (Example 6.3.1.2).

6.5 Inter-granularity-level relations

In this section I turn to the next abstraction layer of the SCS and examine the inter-granularity-level relations, which will lead us to a Categorification of the entire granularity level space (GLS).

6.5.1 Isomorphic granularity levels

In §6.1.2 I ordered all granularity levels realizable in (not necessarily mature states of) human language into a huge poset by the subtyping-induced inheritance relation. Now we can further view the GLS poset as a Category; call it **GLS**. Recall that a major drawback of order-theoretic organizations of the GLS is their low-level nature; namely, they involve too much arbitrary detail of a single type (i.e., an ordering) and quickly become illegible in visual presentation. **GLS** is also subject to such problems, as in Figure 6.6 (adapted from Figure 6.2b).

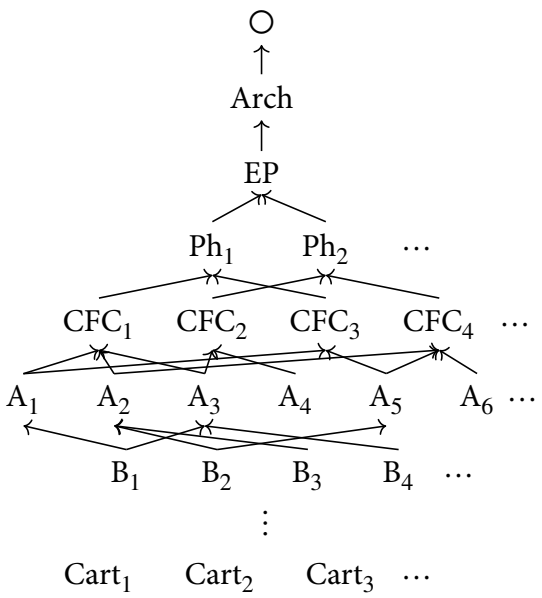


Figure 6.6 The poset Category **GLS** of granularity levels ordered by inheritance

The above said, we may use Figure 6.6 as a starting point and modify it into an adequate higher-level description of the GLS. To begin with, note that *subtyping* and *inheritance* are not exact synonyms in the current context. Inheritance is a relation that statically holds between granularity levels, whereas subtyping is more naturally conceived as a process that dynamically grows c-sequences. Categorically speaking, the two notions require different object-and-arrow configurations. While inheritance can be directly stated in **GLS**, to describe the subtyping process we must access the internal structure of each granularity level, which is not allowed in **GLS**. To get an insider's view we need to zoom in on **GLS** and view its objects as Categories. Accordingly, we view **GLS** as a tiny Subcategory of **Cat** (Definition 6.3.1.4); call it $\mathbf{Cat}_{\mathrm{GLS}}$. This perspective shift gives us more freedom in describing granularity levels. Let us begin with the three top levels ○,

Arch, and EP;[37] call them *protolevels* as no nontrivial c-sequence[38] has been formed yet. Since ○ contains nothing but the undivided categorial space (a most generic supertype), it as a Category has only one object (call it 0) and one (identity) arrow, pictorially

$$0 \circlearrowleft^{\mathrm{id}} . \tag{6.5.1.1}$$

Arch and EP are trickier to characterize due to varied assumptions about the arity of category subtyping and the number of MPOSs, which can be phrased as two questions: (i) How many immediate subtypes can each supertype category be divided into—two (binary division) or more (n-ary division)? (ii) How many MPOSs can there be—two (e.g., N and V) or more (e.g., also A and P)? Depending on one's answers to these questions, there may be four scenarios:

(11) a. Binary division + two MPOSs ⇒ Arch and EP as one level (call it ArchEP)

b. n-ary division + two MPOSs ⇒ one level ArchEP (same as (11a))

c. n-ary division + multiple MPOSs ⇒ one level ArchEP (same as (11a))

d. Binary division + multiple MPOSs ⇒ two or more protolevels Archi–EP–EP′ ⋯

Insofar as the GLS is concerned, the four scenarios in (11) yield two possibilities: Arch and EP are either a single granularity level or separate levels. Scenarios (11a) and (11b) are effectively the same (i.e., $n = 2$), where ArchEP can be conceived as a *discrete poset* (Appendix C.6), which may be further viewed as a discrete Category (Definition 6.3.1.2) with two objects and no nonidentity arrows; call it $\mathbf{ArchEP}_{\mathrm{bin}}$. Pictorially

$$\bullet \circlearrowleft^{\mathrm{id}} \qquad \bullet \circlearrowleft^{\mathrm{id}} . \tag{6.5.1.2}$$

In scenario (11c) there is also just one level ArchEP, with all MPOSs obtained from one n-ary division of 0. This may be viewed as another discrete Category $\mathbf{ArchEP}_{\mathrm{multi}}$ with a number of objects (exemplified by four below) and no nonidentity arrow; pictorially

$$\bullet \circlearrowleft^{\mathrm{id}} \qquad \bullet \circlearrowleft^{\mathrm{id}} \qquad \bullet \circlearrowleft^{\mathrm{id}} \qquad \bullet \circlearrowleft^{\mathrm{id}} . \tag{6.5.1.3}$$

In scenario (11d) Arch and EP cannot be identified. Thus, when there are four MPOSs (e.g., N, V, A, P), we need to apply binary division at least twice to obtain all of them from 0, which makes **Arch** into a discrete Category like (6.5.1.2) and **EP** one like (6.5.1.3). Comparing the four scenarios above, we can see that whichever of them we assume, the protolevels are always discrete Categories.[39] Hence, the varied assumptions on category subtyping are not that conse-

[37] Recall from §2.3.2 that these are respectively the granularity level of the undivided categorial space, that of archi-categories, and that of extended projections (conceived as monadic categories, e.g., V and N).

[38] By a *trivial c-sequence* I mean a c-sequence with only one syntactic category (i.e., a singleton poset).

[39] In other words, they are always sets. I keep to Categorical parlance to facilitate later discussion.

quential for our Categorification of the GLS. For convenience I will work with scenario (11a) in my illustration.

Assuming that all MPOSs become available by the end of the last protolevel, say EP, the post-EP levels can no longer be viewed as discrete Categories as such, because further category division begins to yield nontrivial c-sequences. For example, **Ph** as displayed below has both identity and nonidentity arrows.

$$\overset{\text{id}}{\bullet} \longrightarrow \overset{\text{id}}{\bullet} \qquad \overset{\text{id}}{\bullet} \longrightarrow \overset{\text{id}}{\bullet} \quad . \tag{6.5.1.4}$$

Further down Figure 6.6, while the c-sequences given by category subtyping grow longer and longer, the overall number of c-sequences remains the same; that is, the number of MPOSs remains stable. Thus, the granularity levels CFC, A, B, …, Cart, when viewed as Categories, all have the same basic shape and only differ in the internal content of their objects. But since the objects in a Category are opaque anyway, this lends us a new way to view the post-EP granularity levels as discrete Categories, by treating each c-sequence (instead of each category) as an object. Consequently we obtain the following Categorical description of the GLS.

Theorem 6.5.1.1. *All post-EP granularity levels (including EP) viewed as discrete Categories are isomorphic.*

Proof. As mentioned in note 39, p. 222, the granularity-level discrete Categories are just sets (equipped with an identity arrow for each element).[40] Thus, to prove Theorem 6.5.1.1 we only need to find a bijection between each pair of post-EP granularity levels qua sets of c-sequences. There is an obvious one that sends each c-sequence $G_{\mathbf{x}}$ of granularity level G and MPOS $\mathbf{x}$ to a c-sequence $G'_{\mathbf{x}}$ of the same MPOS at another granularity level G'. Take Ph/CFC with $\mathbf{N}/\mathbf{V}$ for example. We have the following bijection:

$$\begin{array}{ccccc} \mathbf{Ph} & & \text{id} \circlearrowright \mathrm{Ph}_{\mathbf{N}} & & \text{id} \circlearrowright \mathrm{Ph}_{\mathbf{V}} \\ f \Big\downarrow\Big\uparrow f^{-1} & & \Big\downarrow\Big\uparrow & & \Big\downarrow\Big\uparrow \\ \mathbf{CFC} & & \text{id} \circlearrowright \mathrm{CFC}_{\mathbf{N}} & & \text{id} \circlearrowright \mathrm{CFC}_{\mathbf{V}} \end{array} \tag{6.5.1.5}$$

Bijections between other granularity levels can be proved similarly. Since we are abstracting away from the c-sequence-internal content, the two trivial EP-level c-sequences, whose only members are the undivided MPOS categories, can be treated in the same way (i.e., as objects $\mathrm{EP}_{\mathbf{V}}$ and $\mathrm{EP}_{\mathbf{V}}$ of a discrete Category **EP**). Hence, we have $\mathbf{EP} \cong \mathbf{Ph} \cong \mathbf{CFC} \cong \mathbf{A} \cong \mathbf{B} \cong \cdots \cong \mathbf{Cart}$. □

The above discrete-Category-based perspective on post-EP granularity levels has an additional merit: it unifies the interpretation of *objects* in granularity-level Categories, with each

[40] This does not mean all discrete Categories are sets—ours just happen to be small enough.

object being identified as a (possibly trivial) c-sequence. This is consistent with the ladder-of-abstraction conception, where the abstraction layer of single granularity levels has c-sequences instead of single categories as basic components. If all post-EP granularity levels are isomorphic, what about the protolevels? Since we are assuming two MPOSs, there is only one protolevel: ○. And since ○ as a Category has only one object and one arrow, it can be viewed as the terminal object (Definition 6.3.1.6) in $\mathbf{Cat}_{\mathrm{GLS}}$, whose objects are granularity-level Categories and arrows are functions qua Functors between granularity levels, including the above-defined isomorphisms and the unique Functors from each granularity level to ○. See Figure 6.7 for an illustration.

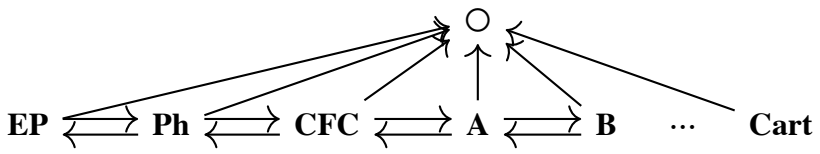


Figure 6.7 The Category $\mathbf{Cat}_{\mathrm{GLS}}$ of granularity-level Categories

Next let us add in the incomparable granularity levels, which were the main cause of the entanglement in Figure 6.6. Since all incomparable levels of the same general level have the same MPOSs, they too are all isomorphic as discrete Categories of c-sequences. Thus the diagram in Figure 6.7 can be extended to the "three-dimensional" version in Figure 6.8.[41]

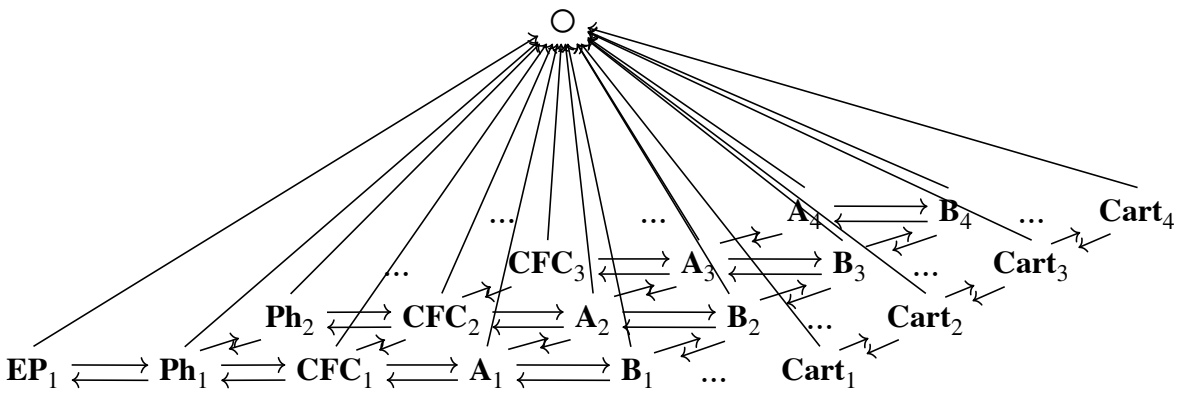


Figure 6.8 $\mathbf{Cat}_{\mathrm{GLS}}$ extended with incomparable granularity levels

In Figure 6.8, all granularity levels within the same plane are isomorphic, as they all emanate from the same EP, where the MPOSs for the entire plane are fixed. By comparison, variation at the EP-level would generate incomparable planes (not shown in Figure 6.8), which are nevertheless all connected to ○. Interestingly, since isomorphic objects in a Category can be treated as the same, the huge Category in Figure 6.8 is equivalent (in the Categorical sense in Definition 6.3.3.3) to the simple Category in Figure 6.9, which is a skeleton of $\mathbf{Cat}_{\mathrm{GLS}}$ (Definition 6.3.1.5).

The "minimalist" Category in Figure 6.9 abstracts away from all details of the GLS but two most crucial pieces of information. First, all GLS-layer variation are given rise to by the single

[41] In Figure 6.8 all post-EP granularity levels are isomorphic. I have omitted the diagonal bijective arrows (e.g., those between $\mathbf{Ph}_1$ and $\mathbf{CFC}_2$), since they are guaranteed by arrow composition.

$$\begin{tikzcd}
 & \bigcirc & \\
\mathbf{EP}_1 \arrow[ur] & \mathbf{EP}_2 \arrow[u] & \mathbf{EP}_3 \arrow[ul] & \cdots \arrow[ull]
\end{tikzcd}$$

Figure 6.9 A Categorical skeleton of $\mathbf{Cat}_{\mathrm{GLS}}$

mechanism of category subtyping—plus the essential bootstrapping step of category formation from nothing to something (i.e., the MPOSs). In other words, Figure 6.9 is the template at the abstraction layer of GLS, and Figure 6.8 is (a corner of) its flexibility space. Second, Figure 6.9 also retains the information that $\bigcirc$ is the part of GLS reachable from any granularity level. This is reminiscent of the defective category Cat ($\neq$**Cat**) studied in Chapter 3, which accompanies all granularity levels and underlies the omnipresent grammatical phenomenon of adjunction ($\neq$Adjunction). By identifying Cat with $\bigcirc$, we can now reinterpret *defective* as describing a default or unactivated state of the GLS.

At this point one may wonder how our starting point, the inter-GRAN inheritance, can be encoded in the new Category. This is an easy task. Since we are now in a nonposetal Category (but a Category of discrete Categories qua objects and Functors qua arrows), it is perfectly feasible to add in more arrows. One option is to reinterpret each instance of the inheritance relation as a function (and Functor) that actually fleshes out the category subtyping process. For example, given the two comparable granularity-level Categories below:

$$\begin{array}{lcll}
\mathbf{Ph}_1 & := & \mathrm{id} \circlearrowright \mathrm{Ph}_{\mathbf{V}} & \mathrm{id} \circlearrowright \mathrm{Ph}_{\mathbf{N}} \\
\mathbf{CFC}_1 & := & \mathrm{id} \circlearrowright \mathrm{CFC}_{\mathbf{V}} & \mathrm{id} \circlearrowright \mathrm{CFC}_{\mathbf{N}}\,,
\end{array}$$

whose c-sequences are defined as follows:

$$\begin{array}{lcl}
\mathrm{Ph}_{\mathbf{V}} & := & v_{\mathrm{Ph}} \rightarrow \mathrm{C}_{\mathrm{Ph}} \\
\mathrm{Ph}_{\mathbf{N}} & := & n_{\mathrm{Ph}} \rightarrow \mathrm{D}_{\mathrm{Ph}} \\
\mathrm{CFC}_{\mathbf{V}} & := & \mathrm{V} \rightarrow v_{\mathrm{CFC}} \longrightarrow \mathrm{T} \longrightarrow \mathrm{C}_{\mathrm{CFC}} \\
\mathrm{CFC}_{\mathbf{N}} & := & \mathrm{N} \rightarrow n_{\mathrm{CFC}} \rightarrow \mathrm{Num} \rightarrow \mathrm{D}_{\mathrm{CFC}}\,,
\end{array}$$

we can reinterpret the ordering $\mathbf{CFC}_1 \leq \mathbf{Ph}_1$ as a Functor $H : \mathbf{CFC}_1 \rightarrow \mathbf{Ph}_1$ defined as follows:

$$\begin{array}{lcll}
\mathbf{Ph}_1 & := & \mathrm{id} \circlearrowright \mathrm{Ph}_{\mathbf{V}} & \mathrm{id} \circlearrowright \mathrm{Ph}_{\mathbf{N}} \\
{\scriptstyle H=\langle j,k\rangle}\uparrow & & {\scriptstyle j}\uparrow & {\scriptstyle k}\uparrow \\
\mathbf{CFC}_1 & := & \mathrm{id} \circlearrowright \mathrm{CFC}_{\mathbf{V}} & \mathrm{id} \circlearrowright \mathrm{CFC}_{\mathbf{N}}\,,
\end{array} \tag{6.5.1.6}$$

where j and k are respectively defined as follows:

$$\begin{tikzcd}
\mathrm{Ph}_{\mathbf{V}} \arrow[r, phantom, ":="] & v_{\mathrm{Ph}} \arrow[r] & \mathrm{C}_{\mathrm{Ph}} & & \\
\mathrm{CFC}_{\mathbf{V}} \arrow[u, "j"] \arrow[r, phantom, ":="] & \mathrm{V} \arrow[u] \arrow[r] & v_{\mathrm{CFC}} \arrow[ul, mapsto] \arrow[r] & \mathrm{T} \arrow[ul, mapsto] \arrow[r] & \mathrm{C}_{\mathrm{CFC}} \arrow[ull, mapsto]
\end{tikzcd} \tag{6.5.1.7}$$

$$\begin{tikzcd}
\mathrm{Ph}_{\mathbf{N}} \arrow[r, phantom, ":="] & n_{\mathrm{Ph}} \arrow[r] & \mathrm{D}_{\mathrm{Ph}} & & \\
\mathrm{CFC}_{\mathbf{N}} \arrow[u, "k"] \arrow[r, phantom, ":="] & \mathrm{N} \arrow[u] \arrow[r] & n_{\mathrm{CFC}} \arrow[ul, mapsto] \arrow[r] & \mathrm{Num} \arrow[ul, mapsto] \arrow[r] & \mathrm{D}_{\mathrm{CFC}} \arrow[ull, mapsto]
\end{tikzcd}.$$

By reinterpreting each instance of the inter-GRAN inheritance relation in this way, we effectively record all possible versions of Biberauer & Roberts' (2015: 299) category division hierarchy given in Figure 6.10.[42]

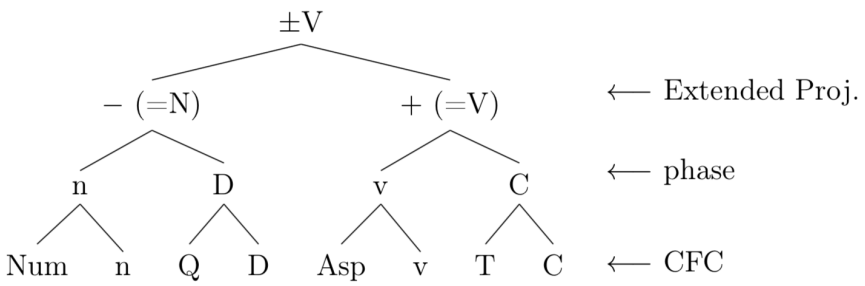


Figure 6.10 Biberauer & Roberts' (2015) syntactic category division hierarchy

Note that none of the arrows in (6.5.1.6) and (6.5.1.7) need to be present in $\mathbf{Cat}_{\mathrm{GLS}}$; they are encoded in the definition of the single $\mathbf{Cat}_{\mathrm{GLS}}$-arrow H instead, as in Figure 6.11.

$$\begin{tikzcd}
\mathbf{Ph}_1 \arrow[loop left, "\mathrm{id}"] \arrow[r, shift left, "f^{-1}"] & \mathbf{CFC}_1 \arrow[l, shift left, "f"] \arrow[loop right, "\mathrm{id}"] \arrow[ll, bend left, "H"]
\end{tikzcd}$$

Figure 6.11 A corner in $\mathbf{Cat}_{\mathrm{GLS}}$ (cf. Figure 6.8)

In addition, H is half of an epi-Adjunction, or more exactly a conjunction of two smaller epi-Adjunctions, one between $\mathrm{CFC}_{\mathbf{V}}$ and $\mathrm{Ph}_{\mathbf{V}}$ and the other between $\mathrm{CFC}_{\mathbf{N}}$ and $\mathrm{Ph}_{\mathbf{N}}$. In (6.5.1.8)

$$\begin{tikzcd}
\mathbf{CFC_V} \arrow[r, bend left, "j"] \arrow[r, phantom, "\perp"] & \mathbf{Ph_V} \arrow[l, bend left, "e"]
\end{tikzcd} \tag{6.5.1.8}$$

[42] Each H corresponds to an inter-row connection (i.e., a set of edges) in Biberauer & Roberts' diagram.

the Functor j is defined as in (6.5.1.7), and the Functor e is an embedding that sends each Ph-level category to a CFC-level core category. It is routine to verify that $j \dashv e$ is an epi-Adjunction; for example, we have $e \circ j(\mathrm{T}) = \mathrm{C}_{\mathrm{CFC}} \geq \mathrm{T}$ (a unit component), and $j \circ e(\mathrm{C}_{\mathrm{Ph}}) = \mathrm{C}_{\mathrm{Ph}}$ (a co-unit component), as displayed below:

$$\begin{tikzcd}
\mathbf{CFC_V} \arrow[d, bend right, "j"'] \arrow[d, phantom, "\vdash"] \\
\mathbf{Ph_V} \arrow[u, bend right, "e"]
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathrm{V} \arrow[r] \arrow[d, mapsto] & \boxed{\nu_{\mathrm{CFC}}} \arrow[r] & \mathrm{T} \arrow[r] \arrow[d, mapsto] & \boxed{\mathrm{C}_{\mathrm{CFC}}} \\
\nu_{\mathrm{Ph}} \arrow[rr] \arrow[ur, dashed, bend left] & & \mathrm{C}_{\mathrm{Ph}} \arrow[ur, dashed, bend left] &
\end{tikzcd}$$

Similarly, k in (6.5.1.6) also has an embedding right adjoint d (i.e., $k \dashv d$). Hence, $H \coloneqq \langle j, k \rangle$ as a whole has an embedding right adjoint $E \coloneqq \langle e, d \rangle$ (i.e., $H \dashv E$), as displayed below:

$$\begin{tikzcd}
\mathbf{Ph}_1 \arrow[d, bend right, "E=\langle e,d\rangle"] \arrow[d, phantom, "\dashv"] & \coloneqq & \mathrm{Ph}_{\mathbf{V}} \arrow[loop above, "\mathrm{id}"] \arrow[d, bend right, "e"] \arrow[d, phantom, "\dashv"] & \mathrm{Ph}_{\mathbf{N}} \arrow[loop above, "\mathrm{id}"] \arrow[d, bend right, "d"] \arrow[d, phantom, "\dashv"] \\
\mathbf{CFC}_1 \arrow[u, bend right, "H=\langle j,k\rangle"] & \coloneqq & \mathrm{CFC}_{\mathbf{V}} \arrow[loop below, "\mathrm{id}"] & \mathrm{CFC}_{\mathbf{N}} \arrow[loop below, "\mathrm{id}"] .
\end{tikzcd} \tag{6.5.1.9}$$

Based on (6.5.1.9), we can now complete Figure 6.11 as in Figure 6.12.

$$\begin{tikzcd}
\mathbf{Ph}_1 \arrow[loop left, "\mathrm{id}"] \arrow[r, shift left, "f^{-1}"] \arrow[rr, bend left, "E"] & \mathbf{CFC}_1 \arrow[l, shift left, "f"] \arrow[ll, bend left, "H"] \arrow[loop right, "\mathrm{id}"]
\end{tikzcd}$$

Figure 6.12 A corner in $\mathbf{Cat}_{\mathrm{GLS}}$ (complete)

And this multi-Functor connection can be extended to the entire $\mathbf{Cat}_{\mathrm{GLS}}$. What is more, the epi-Adjunctions thus established are composable. For instance, in the following configuration:

$$\begin{tikzcd}
\mathbf{Ph}_1 \arrow[r, shift left, "f^{-1}"] \arrow[rr, bend left, "E"] & \mathbf{CFC}_1 \arrow[l, shift left, "f"] \arrow[ll, bend left, "H"] \arrow[r, shift left, "g^{-1}"] \arrow[rr, bend left, "N"] & \mathbf{A}_1 \arrow[l, shift left, "g"] \arrow[ll, bend left, "M"]
\end{tikzcd}, \tag{6.5.1.10}$$

there are three epi-Adjunctions: (i) $H \dashv E$, (ii) $M \dashv N$, and (iii) $H \circ M \dashv N \circ E$.[43]

Back to our starting point, how is $\mathbf{Cat}_{\mathrm{GLS}}$ (a Category of Categories) a better organization of the GLS than **GLS** (a poset Category)? I answer this question by the following summarizing comparison. First, $\mathbf{Cat}_{\mathrm{GLS}}$ allows us to simultaneously describe multiple inter-GRAN relations, such as the inter-GRAN inheritance and the post-EP isomorphism, while this is impossible in **GLS** for two reasons: (i) between each pair of objects in a poset Category, there can be at most

[43] There may still be other interesting inter-granularity-level Functors, but the above suffice for my purposes.

one arrow (Note 6.3.1.2); (ii) under antisymmetry, isomorphic elements in a poset must be identical. Second, although $\mathbf{Cat}_{\mathrm{GLS}}$ is not organized by any order relation, it in fact encodes richer information about the inter-GRAN inheritance than **GLS** does, for arrows like H not only state which two granularity levels are related by inheritance, but also specify how they are related down to the single category layer, which is a mission impossible in **GLS** where arrows are not functions. Third, in $\mathbf{Cat}_{\mathrm{GLS}}$ arrow entanglement is no longer a problem, even though there are actually more arrows than in **GLS**. This is because (i) there is an additional dimension in Figure 6.8 to accommodate the extra arrows, and (ii) we can optionally neglect one type of arrow (e.g., the isomorphism) when focusing on another type (e.g., the inheritance) in $\mathbf{Cat}_{\mathrm{GLS}}$. We cannot do this in **GLS** since there only exists one type of arrow (i.e., that instantiating the order relation). In conclusion, for the purpose of global-level exploration $\mathbf{Cat}_{\mathrm{GLS}}$ is a better tool.

6.5.2 Augmented granularity level isomorphism

The granularity-level Categories in the last section ($\mathbf{Ph}_1$, $\mathbf{CFC}_1$, etc.) differ from the flower-shaped $\mathbf{Gran}_{\mathrm{A}}$ proposed at the end of §6.4 in that the former lack the universal spine component. This difference reflects a theoretical conflict concerning universal spines. On the one hand, they are needed to formulate the cross-MPOS parallelism, but on the other hand, they cannot be derived by successively dividing the syntactic categorial space, as they do not comprise *syntactic* categories after all.[44] If we can integrate the universal spines into our new GLS Category (i.e., $\mathbf{Cat}_{\mathrm{GLS}}$), however, we would be able to unify the two global-level phenomena studied in this chapter (i.e., cross-MPOS parallelism and inter-GRAN relations) and thereby have an answer to the third question at the end of §6.1.

Intuitively, categories in the universal spine are from a separate cognitive domain (perhaps that of thought), which may have its own ontological structure and merely feed the language domain via an interface (see Ramchand & Svenonius 2014 for a similar idea). If so, then the flower-shaped $\mathbf{Gran}_{\mathrm{A}}$ in fact combines information from two categorial systems: one is the SCS, and the other is a *conceptual category system* (CCS). Recall from §6.4.2.1 that a universal spine is formally a poset just like a c-sequence. Since the objects in a granularity-level Category like $\mathbf{Ph}_1$ or $\mathbf{CFC}_1$ are also all posets, we can view the object collection $Ob(\mathbf{Gran}_{\mathrm{A}})$ as the union $Ob(\mathbf{A}) \cup \{\sigma\}$, where $\mathbf{A}$ is a universal-spine-less granularity-level Category (i.e., an object in $\mathbf{Cat}_{\mathrm{GLS}}$) and σ a universal spine. We can let arrows in $\mathbf{Gran}_{\mathrm{A}}$ be the same as in Figure 6.5.

The perspective shift above allows us to define additional Functors between flower-shaped Categories. Suppose we have another flower-shaped Category $\mathbf{Gran}_{\mathrm{B}}$ in the same EP-plane as $\mathbf{Gran}_{\mathrm{A}}$, whose object collection is $Ob(\mathbf{B}) \cup \{\sigma'\}$ (σ' is another universal spine not necessarily

[44] In particular, all syntactic categories are mergeable but classification, anchoring, etc. are not.

distinct from σ). We can define a Functor $K : \mathbf{Gran}_{\mathrm{A}} \to \mathbf{Gran}_{\mathrm{B}}$ that maps objects as follows:

$$\begin{tikzcd} Ob(\mathbf{Gran}_{\mathrm{A}}) \arrow[d, "K_{Ob}=f\cup g"'] & = & Ob(\mathbf{A}) \arrow[d, "f"'] & \cup & \{\sigma\} \arrow[d, "g"'] \\ Ob(\mathbf{Gran}_{\mathrm{B}}) & & Ob(\mathbf{B}) & \cup & \{\sigma'\}\,. \end{tikzcd} \tag{6.5.2.1}$$

In (6.5.2.1) f maps each c-sequence in $\mathbf{A}$ to its same-MPOS counterpart in $\mathbf{B}$ (as in (6.5.1.5)), and g maps σ to σ'. As a Functor, K must also map arrows. Since we have let arrows in $\mathbf{Gran}_{\mathrm{A}}$ (resp. $\mathbf{Gran}_{\mathrm{B}}$) remain as originally conceived (i.e., as adjoint Functors between c-sequences and the universal spine), we can map the adjoint Functors in $\mathbf{Gran}_{\mathrm{A}}$ to those in $\mathbf{Gran}_{\mathrm{B}}$ and identities to identities. Functors like K give us an augmented version of the inter-GRAN isomorphism.

Theorem 6.5.2.1. *If we add universal spines to the (post-EP) granularity-level Categories in Theorem* 6.5.1.1*, the augmented Categories thus obtained are still all isomorphic.*

Proof. Since f in (6.5.2.1) is defined like that in (6.5.1.5), it is bijective; and since g is trivially also bijective, K_{Ob} as a whole is bijective. On the other hand, since $\mathbf{Gran}_{\mathrm{A}}$ and $\mathbf{Gran}_{\mathrm{B}}$ as granularity levels within the same EP-plane have the same MPOSs, they also have the same number of adjoint Functors; hence, K is bijective on arrows too. Since K is bijective on both objects and arrows, it has an inverse Functor $K^{-1} : \mathbf{Gran}_{\mathrm{B}} \to \mathbf{Gran}_{\mathrm{A}}$, whence $\mathbf{Gran}_{\mathrm{A}} \cong \mathbf{Gran}_{\mathrm{B}}$. Similarly, we can prove that other pairs of flower-shaped Categories within the same EP-plane are also isomorphic. Hence, $\mathbf{Gran}_{\mathrm{EP}} \cong \mathbf{Gran}_{\mathrm{Ph}} \cong \mathbf{Gran}_{\mathrm{CFC}} \cong \mathbf{Gran}_{\mathrm{A}} \cong \mathbf{Gran}_{\mathrm{B}} \cong \cdots \cong \mathbf{Gran}_{\mathrm{Cart}}$. □

With Theorem 6.5.2.1 we can now replace all granularity-level Categories in $\mathbf{Cat}_{\mathrm{GLS}}$ (Figure 6.8) with their augmented versions; call the updated GLS Category $\mathbf{Cat}_{\mathrm{GLS}^+}$. The inheritance function between each pair of comparable granularity levels and its embedding right adjoint are preserved in $\mathbf{Cat}_{\mathrm{GLS}^+}$. For example, $H \dashv E$ in (6.5.1.9) is augmented to $H^+ \dashv E^+$ below:

$$\begin{tikzcd} \mathbf{Gran}_{\mathrm{Ph}_1} := & \mathrm{Ph}_{\mathbf{v}} \arrow[loop above, "\mathrm{id}"] \arrow[r, bend left, "\perp"] \arrow[d, bend left, "e"] & \mathrm{Ph}_0 \arrow[loop above, "\mathrm{id}"] \arrow[l, bend left] \arrow[r, bend left, "\top"] \arrow[d, bend left, "v"] & \mathrm{Ph}_{\mathbf{N}} \arrow[loop above, "\mathrm{id}"] \arrow[l, bend left] \arrow[d, bend left, "d"] \\ H^+=\langle j,u,k\rangle \arrow[u, bend left, "\dashv"] & \mathbf{Gran}_{\mathrm{CFC}_1} := & \mathrm{CFC}_{\mathbf{v}} \arrow[loop below, "\mathrm{id}"] \arrow[r, bend left, "\perp"] \arrow[u, bend left, "j"] & \mathrm{CFC}_0 \arrow[loop below, "\mathrm{id}"] \arrow[l, bend left] \arrow[r, bend left, "\top"] \arrow[u, bend left, "u"] & \mathrm{CFC}_{\mathbf{N}}\,. \arrow[loop below, "\mathrm{id}"] \arrow[l, bend left] \arrow[u, bend left, "k"] \arrow[u, bend left, "\dashv"] \end{tikzcd} \tag{6.5.2.2}$$

As we let $\mathbf{Gran}_{\mathrm{CFC}_1}$ and $\mathbf{Gran}_{\mathrm{Ph}_1}$ vary in $\mathbf{Cat}_{\mathrm{GLS}^+}$, there may be some slight variation in the definition of the "triplet Adjunction" above: one or two of the three vertical (i.e., inter-GRAN) connections on the right-hand side of $:=$ in (6.5.2.2) may actually be isomorphisms; for example, when the two granularity levels in question have a common universal spine or when they

differ only in one c-sequence. However, as long as the two granularity levels differ in at least one component, be it a c-sequence or the universal spine, they are necessarily connected by an epi-Adjunction (instead of an isomorphism). The augmented epi-Adjunctions are all composable too; for example, in the following configuration:

$$\begin{tikzcd} \mathbf{Gran}_{\mathrm{Ph}_1} \arrow[r, bend left, "J^{-1}"] \arrow[rr, bend left=60, "E^+"] & \mathbf{Gran}_{\mathrm{CFC}_1} \arrow[l, bend left, "J"] \arrow[r, bend left, "L^{-1}"] \arrow[rr, bend left=60, "N^+"] \arrow[ll, bend left=60, "H^+"] & \mathbf{Gran}_{\mathrm{A}_1} \arrow[l, bend left, "L"] \arrow[ll, bend left=60, "M^+"] \end{tikzcd} \tag{6.5.2.3}$$

there are three epi-Adjunctions: (i) $H^+ \dashv E^+$, (ii) $M^+ \dashv N^+$, and (iii) $H^+ \circ M^+ \dashv N^+ \circ E^+$. Recall from Remark 6.3.3.5 that an epi-Adjunction is a section-retraction situation at the Functorial level, which can in turn be viewed as a sorting or partitioning situation (Remark 6.3.1.2). We can view the successive epi-Adjunctions in $\mathbf{Cat}_{\mathrm{GLS}^+}$ as successive inter-GRAN partitions, with the supertype granularity levels partitioned into their immediate subtype levels. In Categorical terms, $\mathbf{Gran}_{\mathrm{Ph}_1}$ is a retract of $\mathbf{Gran}_{\mathrm{CFC}_1}$, $\mathbf{Gran}_{\mathrm{CFC}_1}$ is a retract of $\mathbf{Gran}_{\mathrm{A}_1}$, and so forth.

We now have a unified theory for cross-MPOS parallelism and inter-GRAN relations based on epi-Adjunction. Specifically, by the Functorial interconnection across granularity levels, we can reach any c-sequence or universal spine at any granularity level from any c-sequence or universal spine at any other granularity level within the same EP-plain, with all the connected c-sequences or universal spines being in epi-Adjunctions, which makes them all similar in a strict Categorical sense (Remark 6.3.3.2). Epi-Adjunction as a basic pattern is so fundamental in this chapter that it may be considered an overarching template at the two c-sequence-based abstraction layers of the SCS; namely, the entire global level of concern.

6.5.3 Multiple universal spines

A consequence of the foregoing discussion is that universal spines may also vary in granularity. I have assumed that each non-○ granularity level is equipped with a universal spine, which serves as the template for the concrete c-sequences; I have also shown that a c-sequence and its universal spine are connected by an epi-Adjunction. However, since in such an epi-Adjunction the universal spine must not be larger in size than the c-sequence (for the Functor from the latter to the former must be surjective; see Definition 6.3.3.6), this means that, for instance, the universal spine for the post-CFC granularity levels cannot also be that for the Ph (or EP) level. Therefore, if we want to maintain the generality of Theorem 6.4.2.1, we must hypothesize that universal spines come in different sizes as c-sequences do. I state this result in the following corollary:

Corollary 6.5.3.1 (of Theorem 6.4.2.1). *There are multiple universal spines of different granularities for the SCS, which are part of an inheritance hierarchy for conceptual categories.*

In the previous section I took a universal spine as a "c-sequence" in a separate cognitive domain CCS, which may have its own ontological structure. Here I further conjecture that the organization of the CCS is similar to that of the SCS: with conceptual categories being grouped into c-sequences, conceptual c-sequences grouped into granularity levels, and so forth. This suggests that the Wiltschkovian universal spine is part of an inheritance hierarchy just like any syntactic c-sequence is. I tentatively propose the hierarchy in Figure 6.13.

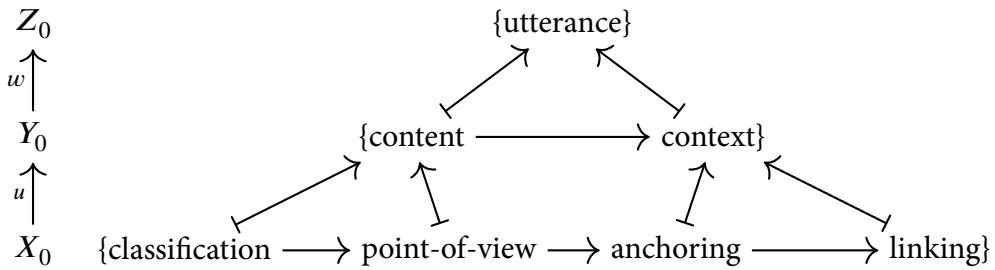


Figure 6.13 An inheritance hierarchy for universal spines

In Figure 6.13 there are three universal spines X_0, Y_0, and Z_0. We can also define two inheritance functions u and w mapping each conceptual category to its immediate supertype. I call the supertype of classification and point-of-view *content* since these primarily concern the semantic core of an utterance, and I call the supertype of anchoring and linking *context* since these primarily concern circumstance- or discourse-related information. However, I do not take the most general conceptual category (which I call *utterance*) as an undivided conceptual categorial space, for the tiny hierarchy in Figure 6.13 is unlikely to represent the entire cognitive domain it belongs to. It is more likely just the corner specialized for language.

There is a close resemblance between Figure 6.13 and Biberauer & Roberts' (2015) category division hierarchy (Figure 6.10), an adapted version of which is presented in Figure 6.14. Here $\mathbf{CFC}_2$, $\mathbf{Ph}_1$, $\mathbf{EP}_1$, and $\bigcirc$ are universal-spine-less granularity-level Categories (i.e., objects in $\mathbf{Cat}_{\mathrm{GLS}}$), and H', G, ! are inheritance Functors (i.e., arrows in $\mathbf{Cat}_{\mathrm{GLS}}$).[45]

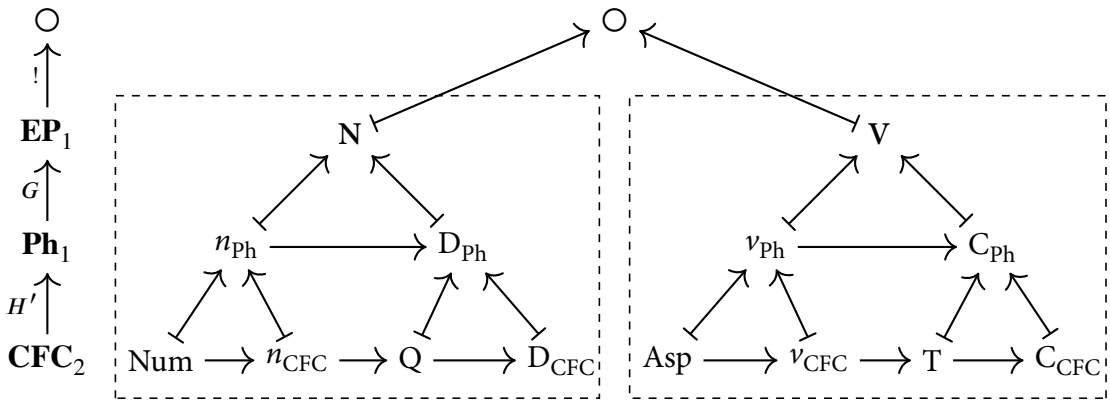


Figure 6.14 Category division hierarchy overlaid with inheritance Functors

[45]It is a Category-theoretic convention to denote the unique arrow to a terminal object by !.

Each boxed substructure in Figure 6.14 is isomorphic to the structure in Figure 6.13. So, X_0 can be conceived as the universal spine at the (post-)CFC levels, Y_0 as that at the Ph levels (so u in Figure 6.13 is just u in (6.5.2.2)), and Z_0 as that at the EP levels. The ○ level, representing the undivided categorial space, has no corresponding universal spine. A series of questions may be asked about Figure 6.13; for instance, Are there still longer universal spines than X_0? Might there be some "higher-order template" guiding the successive subtyping of utterance? And so on. Below I present some general ideas on these issues.

First, since there is no fixed limit on the size of the inheritance hierarchy in Figure 6.13, in theory there could be longer universal spines if those can be backed up by empirical evidence (see Wiltschko 2016 et seq. for a possibility). Second, a possible candidate for the higher-order universal spine is the two-membered chain {core→periphery} (à la Biberauer 2018), which captures a common pattern in the subtyping of utterance and that of content/context; thus, context is peripheral to content, and point-of-view is peripheral to classification. Importantly, however, such a higher-order template is not a template in the SCS but one in the CCS; that is, insofar as the SCS is concerned, there is still only one template per abstraction layer.

6.6 Revisiting categorial sequence indexing

So far I have Categorified three abstraction layers of the SCS: the layer of single c-sequences, that of single granularity levels, and that of the entire GLS. Since the final layer (i.e., the entire SCS) is not central to this dissertation, I will only discuss it briefly in §6.7. For now let us return to the first layer (i.e., that of single syntactic categories) and complete a half-done formalization initiated in Chapter 2. To facilitate our discussion I repeat the featural template for syntactic categories in (12).

(12) syntactic category$_{\text{LANG, GRAN}}$:= [[MERGEME:[CAT: `cat`, POS: `nat`]], NM]

In Chapter 2 I mentioned two ways to locate a syntactic category in its ambient c-sequence: (i) by a numerical index; (ii) by a relational predicate. For example, given a c-sequence CFC_v := {V→v→T→C}, we can pick out T either by indexing CFC_v with the natural numbers 1–4 and assigning 3 as the value of the [POS] feature in T or by assigning [POS] a more elaborate value $\lambda c : \text{CFC}_\text{v} . v \leq c \wedge c \leq \text{C}$, which returns `true` iff $c = \text{T}$. The purpose of encoding such "coordinates" in a category's feature set is to ensure that when the category is taken out of the lexicon (where the SCS ontology lives) and put to derivational use, it can correctly find its way in (i.e., its first-Merge position).[46] In this sense, the mergeme is a channel for the rich SCS structure explored in this dissertation to influence or guide the generation of concrete utterances. In other words, it bridges the ontological and derivational aspects of syntactic categories.

[46] The coordinates reduce the selection of mergeme-bearing items from the LA to the workspace to featural terms without resorting to a [SELECT] feature.

As mentioned in Chapter 2, I merely adopt the numerical indices for their readability but do not assume that they are also the underlying representations of [POS] values in the mental knowledge. The question then is, What could the underlying representations be? This was not an issue in Chapter 2 because there we only had one option, the relational predicates. Those are not bad, as they are nonnumerical and also successfully ban multiple flavors of a category from co-occurring in the same derivational layer (see Chapter 5 for the motivation of this ban). Yet they are not that good either due to three design problems. First, there is no obvious way to convert a relational predicate into a numerical index,[47] so the former is more like a separate valuation for [POS] than an underlying representation for the latter. Second, the relational predicates are "resource-demanding" because to locate any category we must first locate two[48] other categories (e.g., v and C for T). Third, c-sequences are built on combinatorial-functional scope, but no scope information is reflected in the relational predicates, which makes them less informative than desired. Thus, if we could find some alternative method to represent the [POS] values that is free from the above problems, then that would be a better option than the relational predicates.

Roberts (2019, p.c.) suggests such an alternative based on c-command, whereby a (minimal) category in an extended projection is indexed in terms of the number of heads in the same extended projection it asymmetrically c-commands. Thus, the indexing for the above-mentioned $\mathrm{CFC_v}$ is (in pair notation) $\langle \mathrm{V}, 0\rangle$, $\langle v, 1\rangle$, $\langle \mathrm{T}, 2\rangle$, and $\langle \mathrm{C}, 3\rangle$. Roberts' method solves the problems of relational predicates: the set of c-commanded heads defines a scope and its cardinality defines an index, and in this process no extra information is required for each category. Also, the c-command scope is extensionally the same as the combinatorial scope underlying c-sequences, which makes it potentially compatible with my framework. However, a problem with c-command is that its definition relies on a broadly derivational[49] context, where a parse tree or some equivalent/abstraction thereof can be used; so c-command cannot be meaningfully referenced in a discourse where no such context is readily available. In short, Roberts' method points out a promising direction, but we need something other than c-command as the scope-defining device, something equivalent in power but does not require a broadly derivational context.

Category theory has such a device: the Yoneda embedding. Recall from Note 6.3.4.2 that this is an embedding Functor $y : \mathscr{C} \to \mathbf{Set}^{\mathscr{C}^{\mathrm{op}}}$ that sends a Category $\mathscr{C}$ into a Category of set-valued Functors (i.e., presheaves on $\mathscr{C}$). Intuitively, this means that every object in $\mathscr{C}$ is determined by how other objects in the same Category map to it. In the case of posets, the Yoneda Functor

[47] Although the numerical indices are merely expository, for theoretical coherence I still need to specify how to get them from the underlying representations I propose.

[48] Or one, if the category to be located occupies an extremal position.

[49] By *derivational* I do not only mean the kind of derivation in GB/minimalism (i.e., procedural structure building). Rather, any rule-based construction of linguistic objects from smaller parts, be it procedural or declarative, can be considered derivational in a proof-theoretic sense, so GPSG/HPSG and categorial grammars are also broadly derivational in that their techniques can be given proof-theoretic interpretations. It is the access to specific syntactic objects or some abstraction thereof, including structure-building-oriented notions like sisterhood and domination/containment, that we do not have in a purely ontological study of syntactic categories, where the focus is on *categories* rather than *syntax*. This is somewhat reminiscent of Ellis' (1993) assertion cited in Chapter 1.

effectively embeds a poset into its lower set completion (Example 6.3.4.1), and any element in the poset is fully determined by its principal lower set (i.e., the subposet containing all elements less than or equal to it). The Yoneda embedding is relevant for us because if we understand *fully determine* as *uniquely pick out*, then the principle lower set of a syntactic category becomes a tool to represent its [POS] value or its coordinates. In $\mathrm{CFC_v}$, for example, V is uniquely picked out by its principle lower set {V}, v by {V→v}, T by {V→v→T}, and C by {V→v→T→C}. In fact, subtracting a category c from its own principal lower set $\downarrow c$ gives us a new poset $\langle K(c), \leq\rangle$ (call it the K-set of c) whose underlying set $K(c)$ is exactly the c-command scope set $com(c)$ in Roberts' method (call it the com-set of c); that is, for any category c in a given c-sequence, we have

$$com(c) = K(c)\,. \tag{6.6.0.1}$$

For example, $com(\mathrm{T}) = K(\mathrm{T}) = \{\mathrm{V}, v\}$. Under (6.6.0.1) the K-set can be considered a non-derivational counterpart of the com-set. Accordingly, we could grant the Yoneda embedding a similar status in the SCS ontology as the status of c-command in syntactic derivation.

Furthermore, since K-sets are posets, they systematically encode more information about the categories they pick out than com-sets do, with the latter being plain, unordered sets. For this reason, we may also call K-sets *ordered c-command scope sets*. The additional order information makes K-sets more efficient than com-sets in certain circumstances. For example, it provides a very simple algorithm for the minimalist derivational system to deduce what category to merge next at a given stage, by comparing the least-embedded head in the active workspace with the maximum of the K-set of each category in the LA: the category with a matching K-set maximum is the one to select and merge next. For example, if the least-embedded head in the workspace of a CFC-level derivational stage is v, then the next-merged head is T, since $v = max(\langle K(\mathrm{T}), \leq\rangle)$.

A similar algorithm in com-set terms is hard to formulate. Since com-sets are unordered, no special elements can be singled out in algorithm design. The only conceivable element-wise comparison in this situation is between entire com-sets of categories in the LA and the set of heads in the workspace at a given stage, whereby the category with a matching com-set is selected and merged next.[50] Thus, if the workspace of a CFC-level derivational stage contains V and v, then the next-merged head is T, as $\{\mathrm{V}, v\} = com(\mathrm{T})$. However, this algorithm quickly loses power as we take phases into consideration, because a com-set containing categories across multiple phases might never be identified as a match (and so never gets merged) due to cyclic spell-out. Take Top for example, and suppose it has the com-set {V, Appl, v, Asp, Tns, Mod, Fin, Foc}. Since by the stage Foc is merged, at least V and Appl have been transferred,[51] the set of heads in the workspace does not match com(Top) or any com-set in the LA, and the derivation erroneously

[50] I abstract away from satellites, assuming that they do not hinder the c-command relation between spine heads.

[51] That is, if we assume $\mathrm{PIC_2}$ (Chomsky 2001). If $\mathrm{PIC_1}$ (Chomsky 2000) is adopted, then even more heads would have been transferred by the stage Foc is merged.

halts.[52] No such trouble occurs in the K-set-based algorithm: Top is merged next when Foc is the least-embedded head, which is not affected by cyclic spell-out (at least under PIC_2).

Of course, both *com*-set-based and K-set-based representations of POS values retain the merits of relational predicates: they are both nonnumerical and can both rule out co-occurring flavors of a single category (by assigning them identical numerical indices). But the additional advantages of the K-set-based representation make it a more desirable option, especially in an ontological study of syntactic categories. As such, Category theory yields nice results not only at the global level of the SCS, but also at the individual level and for certain derivational issues.

6.7 Chapter summary

In this chapter I have applied mathematical Category theory to studying the global interconnection in the SCS. Category theory provides a coherent description of the SCS ontology beyond the layers of single categories and c-sequences. It also helps us formulate the global-level template-flexibility pairs in rigorous, succinct terms, which is a purpose lying beyond the design capacity of feature theory. I summarize the most important results of this chapter in (13), in response to the three questions raised in §6.1 (cf. (7)):

(13) a. *How can a linguistically meaningful inter-c-sequence connection be constructed that can adequately and consistently formalize the cross-MPOS parallelism?*
There is no way to directly Categorify the cross-MPOS parallelism, but a mediated connection can be built between any c-sequences at a given granularity level through the universal spine via a special adjoint situation, an epi-Adjunction.

b. *What is an adequate description of the GLS organization that is both comparatively abstract and sufficiently informative?*
Compared to a poset Category, a Category of Categories is a more suitable Categorification of the GLS, with the universal-spined (or "vertebrate") granularity levels being its objects and various inter-GRAN relations its arrows, such as the MPOS-to-MPOS mappings and the inheritance functions. The former identify all granularity levels descending from the same EP level as an isomorphism class, while the latter are surjective Functors with embedding right adjoints, whence we have epi-Adjunctions again.

c. *Is it possible to have a unified theory for the global interconnection of the SCS?*
Following (13b), the intra- and inter-GRAN epi-Adjunctions are a unifying pattern of fundamental significance at the global level of the SCS.

[52] Ian Roberts (p.c.) suggests a potential solution to this problem: one could (and maybe should) define *com*-sets in relation to phases, so Top would not have V or Appl (or maybe *v* either) in its *com*-set to begin with. I leave a comparison of this phase-based *com*-set and the K-set to future research.

There is still a final, most abstract layer in the SCS: the entire SCS, including both mergeme-bearing and acategorial categories. A Categorification of this abstraction layer would abstract away from all but the most fundamental characteristics shared by all and only syntactic categories (e.g., combinability). This layer is in fact the focus in mainstream Category-theoretic studies of natural languages, where the entire syntactic system is viewed as a Category **Syn** (Example 6.3.2.2) with some special combinatorially oriented properties (e.g., monoidal) and connected by an interpretation Functor to a semantics Category **Sem** with some other special properties (e.g., closed).[53] In this extremely abstract discourse, none of the GLS-internal organization studied in this chapter is directly visible.

However, such a chasm need not exist in the application of Category theory to linguistics, because both the ontological issues I have explored here and the derivational issues in others' studies are indispensable aspects of human language. In fact, the two areas of interest may well be joined. Among others, the following questions are worth exploring. First, how can we ascend to the SCS layer without losing insights accumulated at previous abstraction layers? Since categories at the SCS layer are still defined by features, a possibility is to let the mergeme be a channel to feed syntactic derivations with ontological information, as already touched on in §6.6. Second, how can we combine results from the minimalist program and those from Categorical syntactic theories? As an example, the minimalist combinatorial operation Merge does not distinguish left and right, while the left/right directionality is firmly built into the combinatorial machinery of Categorical grammars (see, e.g., Moortgat 2010, Lambek 2011, Gillibert & Retoré 2014, Piedeleu et al. 2015, and references therein). The directionality component facilitates NLP-oriented research, yet from a mental-biological perspective it is hardly maintainable.[54] If Categorists want to study natural languages really as natural objects, they will benefit from learning about minimalist ideas; conversely, if minimalists want to make their theories more precise, then Categorical grammars set a good example. I leave these and other issues to future research.

[53] Accordingly, natural transformations between interpretation Functors are shifts across different models.

[54] e.g., Why should the language faculty evolve two combinatorial operations (say, left- and right-Merge) for English *yellow banana* and Spanish *plátano amarillo* 'banana yellow' just because they differ in word order?

Chapter 7

Conclusion

The concept of category is of paramount importance to the study of natural languages in general and to that of syntax in particular. This dissertation is about the formal flexibility of syntactic categories: *flexibility* under the assumption that UG only provides templatic guidance to syntactic category formation but does not impose strict requirements on how the templates should be implemented, and *formal* in the sense that the flexibility types in question are all logically deduced from the abstract templates and hence are theorems of the UG formal system. As such, each type of flexibility that has been discussed in this dissertation is by assumption universal.

My study of categorial flexibility began in Chapter 2 with a (re)examination of the linguistic feature systems and a (re)definition of syntactic categories. My purpose was not to reinvent the wheel but to demonstrate that insights from multiple fields and traditions can be coherently integrated to yield more objective perspectives on thorny issues. Specifically, my examination of feature systems took into account theories from phonology, morphosyntax (both GB/minimalism and unification-based frameworks), computer science, and mathematical logic. As it turns out, some traditions are more concerned with feature ontology (i.e., What features are there? How are they classified or organized?), some are more concerned with "feature engineering" (i.e., What can be done with features? How to let features function like cogs and gears?), and still some are more concerned with feature logic (i.e., How should features be interpreted as terms in a formal system? What are the possible mathematical models for feature systems?). Building on all these traditions, I suggested a featural metalanguage consisting of four key components: specification (monadic or bipartite), valuation (coefficient or attributive), typing (logico-mathematical or taxonomic), and granularity (syntagmatic or paradigmatic). Using this metalanguage, I proposed the featural definition for syntactic categories in (1), which integrates the combinatorial and taxonomic perspectives on syntactic categories in the `mergeme`-typed feature structure.

(1) syntactic $\text{category}_{\text{LANG, GRAN}}$ ≔ [[MERGEME: ⟨CAT: cat, POS: nat⟩], NM]

The concept mergeme encodes the scope-based ordering among functional categories on the one hand and represents the minimal mergeable and composable unit relativized to a granularity level on the other. The definition in (1) is the UG template at the first layer of abstraction in my proposed categorial universe (i.e., the layer of single categories).

From this first-layer template I deduced two types of formal categorial flexibility: the defective category (2a) and the acategorial category (2b). The latter further subsumes two subtypes: the logical operator (2c) and the Root (2d). These flexibility types are all yielded by incomplete feature specification, which in turn builds on the notion of nonequipollence (i.e., the conception that underspecified features have special theoretical status).

(2) a. defective $\text{category}_{\text{LANG}}$ ≔ [[MERGEME: ⟨CAT:_, POS:_⟩], NM]

b. acategorial $\text{category}_{\text{LANG}}$ ≔ [(NM)]

c. logical $\text{operator}_{\text{LANG}}$ ≔ [NM]

d. $\text{Root}_{\text{LANG}}$ ≔ [ ]

In Chapters 3–4 I respectively studied the defective category (Cat) and the Root category (leaving the logical operator category to future research). The defective category, with its chameleon-like character (i.e., that it gets assimilated, via Agree, into whatever nondefective category it merges with), provides a feature-based account of (certain types of) adjunction, because it can introduce an additional "segment" to a "category" without affecting the latter's projecting status. I illustrated the potential explanatory power of the defective category with two case studies: one on modifier-head compounds across languages and the other on Chinese-style sentence-final particles. As it turns out, the two apparently disparate phenomena can be analyzed as instantiations of the same structure respectively at the word-formation level and at the interactive discourse level.

As for the Root category, I objected to the root categorization assumption in DM and proposed a generalized root syntax instead, according to which any functional item that simultaneously manifests lexical-encyclopedic properties (i.e., any semifunctional item) can be analyzed as a root-supported head. Schematically, this is simply the merger of a non-Root category and a root at a separate derivational layer, which is labeled by the non-Root. I illustrated the mechanism of root support by a comprehensive investigation of the semifunctional items in Chinese, showing that there is a close connection between root support and analyticity.

After examining the categorial flexibility at the first abstraction layer of the SCS, I turned to study that at the second abstraction layer in Chapter 5; namely, the layer of single categorial sequences. I suggested that various mergeme-bearing categories are organized by their mergeme values into partially ordered sets (acategorial categories, on the other hand, are outside the ladder of abstraction). The second-layer UG template is the combinatorial-scope-induced partial

ordering, and the flexibility lies in that this ordering can be either plainly partial or furthermore total; in the former scenario we obtain incomparable categories:

(3) Two categories c and c' are incomparable iff they have identical mergeme values.

While the predominant view in GB/minimalism (especially in the cartography branch) is that hierarchies of projections are totally ordered, I have demonstrated the potential existence of incomparable categories by looking into the "flavored" categories in DM. Flavored categories behave rather differently from nonflavored categories, and this difference matches that between incomparable and comparable categories deduced from the order-theoretic properties of c-sequences.

Faced with the conflict between legality and "bizarreness" concerning incomparable categories, I looked into them further and found out that albeit UG-allowed, incomparable categories are disfavored by non-UG or third factors. This implies that many proposed flavors do not really comply with virtual conceptual necessity and are hence nonminimalist in spirit. I showed what a minimalist application of category flavoring might look like by a case study of the historical development of resultative verbal compounds in Chinese. I reduced the three verbalizer flavors in previous studies to one (i.e., v_{BE}) and showed that even this one had only emerged under a series of coincidental language-specific conditions. What is more, its emergence qua a new categorizer brought about global impact on the language way beyond its motivating construction. Since neither historical motivation nor global impact is mentioned in most current applications of flavored *v*s, I suspect that they are either merely expository devices or not really categorizers. At any rate, we should be cautious in the usage of flavored categories in syntactic analyses.

Inquiries into syntactic categories in the generative tradition usually stop at the c-sequence layer. There has been copious research on the nature of single categories and on their interaction in extended projections; however, it has seldom been asked what further organization there might be in the syntactic category ontology beyond single categories and their hierarchical organizations. In Chapter 6 I took this question seriously and proposed UG templates and formal flexibility types for the third and fourth abstraction layers in the SCS; namely, the layer of a single granularity level and that of the granularity level space (GLS; i.e., the space of all possible granularity levels). In this context I defined a granularity level as an entire inventory of mergeme-bearing categories, which in the simplest assumption is just a set of somehow related c-sequences. My investigation at these higher layers are centered around three questions:

(4) a. How can a linguistically meaningful inter-c-sequence connection be constructed that can adequately and consistently formalize the cross-major-part-of-speech parallelism?

b. What is an adequate description of the GLS organization that is both comparatively abstract and sufficiently informative?

c. Is it possible to have a unified theory for the global interconnection of the SCS?

Since the abstractness and complexity of these questions go beyond the power of feature theory and (basic) order theory, I have approached them from the perspective of mathematical Category theory. Category theory was initially proposed as a bridge between algebra and topology but has had profound interdisciplinary influence since its invention. Its application in linguistics has so far been limited to categorial grammars and natural language processing, and Chapter 6 is an attempt to introduce its results and methods into minimalism. My Categorification of the SCS has the following key components:

(5) a. A single c-sequence can be viewed as a poset Category.

b. A single granularity level can be viewed as a Category of posets. It is associated with a zero object.

c. The GLS can be viewed as a Category of Categories of posets. It has a terminal object.

In particular, I have identified the zero object in (5b) as a universal spine in the sense of Wiltschko (2014) and the terminal object in (5c) as the defective category from Chapter 3. The universal spine is important because it is connected to all c-sequences at the same granularity level via an Adjunction, which is an important notion in both Category theory per se and in my application of it. The specific type of Adjunction that plays a vital role in the SCS Category is epi-Adjunction. In the above Categorical setting, I have identified the template at the third abstraction layer as the universal spine and the flexibility as the various ways the spine can be instantiated (as c-sequences). Similarly, the template at the fourth abstraction layer is the general strategy of category subtyping, and the flexibility lies in the various ways the syntactic categorial space can be divided (which yield varied granularity levels). In addition, epi-Adjunction may be an overarching or "super" template at the two c-sequence-based abstraction layers.

Finally, the above abstraction layers all reflect the ontological organization of mergeme-bearing categories, excluding acategorial categories. To incorporate acategorial categories into the picture we need one last abstraction layer, that of the entire SCS. In fact this layer is closer to the level of concern working syntacticians are more familiar with, as its focus is not on how syntactic categories are ontologically connected but on how they are combined into meaningful utterances. This is also the level of concern assumed in the mainstream linguistics application of Category theory (e.g., that based on Lambek calculus). Since this layer is not my focus, I have only included it for completeness. Its template is just the basic combinatorial operation (Merge in minimalism)—perhaps together with a small set of derivational rules—and its flexibility lies in the familiar parametric variation (which gives rise to varied particular grammars). In conclusion, the template-flexibility pairs I have proposed for the SCS are summarized in Table 7.1, which has been repeated several times throughout the dissertation.

Table 7.1 Templates and flexibility types in the syntactic category system (= Table 2.1)

Level of concern	Abstraction layer	Template	Flexibility
Individual	single category	mergeme-based feature set	incomplete mergemes
	single c-sequence	scope-based partial ordering	incomparable categories
Global	single granularity level	universal spine	varied c-sequences
	all granularity levels	category subtyping	varied granularity levels
Supraglobal	entire SCS	syntactic rules	parametric variation

In a dissertation as ambitious as this one, a lot of detail has inevitably had to be left out to make the main theme clear and to conform to the word limit. I have listed the chapter-specific remaining issues in the respective chapter summaries and here would like to briefly discuss two general directions of future research. First, Category theory could be applied to studying feature systems. In Chapter 2 I introduced several semantic models for linguistic features as terms in a (semi)formal language, and these models can be readily Categorified, with their potential interconnections being formalized as Functors and natural transformations. Second, I suggested near the end of Chapter 6 that the mergeme is a channel connecting the ontological and derivational aspects of syntactic categories; this might provide a new angle to studying the syntax-lexicon interface. Viewing the lexicon as a rich information web and the mergeme as a "localizer," we could let most (if not all) of the ontologically oriented information about syntactic categories (including their intrinsic interpretable or "categorial" features) stay in the lexicon and only feed the syntax with just enough information for the derivation to proceed. The potential outcome of this direction is a half-derivational–half-representational theory of natural language grammar, which may prove useful in the integration of insights from different grammatical theories (e.g., minimalism, HPSG, categorial grammars).

The *Merriam-Webster Collegiate Dictionary 11th Edition* defines *category* as (i) "any of several fundamental and distinct classes to which entities or concepts belong," and (ii) "a division within a system of classification." The role of syntactic categories in linguistics has long surpassed this basic taxonomic functionality of categories, and sometimes our mindset is so focused on particular technicalities of particular theories building on and manipulating categories that we temporarily forget or purposefully leave aside the foundational question of what a category is. Mathematicians (at least Categorists) are excited to studying the commonalities across mathematical subfields in terms of categories, but is it not the same situation in linguistic subfields? We study different aspects of human language in different theoretical schools, but categories (instead of features) seem to be the "standard currency" (to reuse Corbett's term) we share not only among ourselves but also with experts from other fields. In Chapter 1 I began my inquiry with a quotation from Cohen & Lefebvre, and now I would also like to finish with it:

> Categorization is the mental operation by which the brain classifies objects and events. This operation is the basis for the construction of our knowledge of the world. It is the most basic phenomenon of cognition, and consequently the most fundamental problem of cognitive science. (Cohen & Lefebvre 2005: 2)

From a categorially oriented perspective, therefore, the differences among competing (or non-competing) linguistics theories may turn out not to be that different after all. And with this remark I conclude my dissertation.

Appendix A

Examples of modifier-head compound verbs in current English

3D-print	bottle-feed	double-cross	hen-peck	power-nap
age-date	spoon-feed	dream-solve	high-tail	price-mark
air-quote	breast-feed	drop-kick	hitch-hike	proof-read
arm-twist	force-feed	dry-clean	hog-tie	radio-locate
art-edit	break-dance	facebook-stalk	home-teach	reverse-engineer
copy-edit	brow-beat	fact-check	home-work	ring-bark
gene-edit	bulk-download	spell-check	home-school	road-haul
guest-edit	bulk-purchase	spot-check	hood-wink	second-guess
ass-kick	butt-dial	fast-talk	house-break	shadow-ban
axe-murder	drunk-dial	fat-shame	jail-break	shadow-box
baby-sit	chain-drink	fellow-feel	humble-brag	shape-shift
dog-sit	chain-smoke	fellow-travel	hunger-strike	short-change
house-sit	charcoal-broil	time-travel	ice-skate	short-sell
back-comb	flame-broil	field-strip	ill-treat	shrink-wrap
back-date	cherry-pick	fine-tune	ill-use	sight-read
backseat-drive	finger-pick	finger-paint	jam-pack	sight-translate
drug-drive	hand-pick	spray-paint	job-share	sky-rocket
drunk-drive	cliff-hang	fist-bump	jump-start	sleep-walk
night-drive	cold-call	flat-hunt	kick-start	slow-bake
bar-tend	cold-email	head-hunt	jump-scare	slow-clap
batch-cook	computer-generate	job-hunt	karate-chop	sonar-scan
batch-download	crash-land	fly-tip	keel-haul	sonar-sweep

batch-produce	soft-land	free-associate	kerb-crawl	spell-bind
beauty-sleep	force-land	free-write	late-cancel	sprint-clean
binge-sleep	force-fit	ghost-write	lip-sync	stage-manage
binge-watch	cross-classify	hand-write	lip-read	strip-search
binge-eat	cross-cut	hand-deliver	man-spread	tailor-make
hate-eat	cross-fertilize	hand-wave	mass-produce	tail-spin
hate-drink	cross-hatch	hand-craft	mercy-kill	talent-scout
shame-eat	cross-reference	hand-wash	road-kill	tap-dance
bitch-slap	custom-produce	machine-wash	mind-read	tape-record
black-wash	day-dream	gift-wrap	muck-rake	test-fire
white-wash	deep-fry	globe-trot	name-drop	tongue-lash
blow-dry	shallow-fry	group-chat	parachute-jump	trash-talk
freeze-dry	stir-fry	video-chat	queue-jump	turbo-charge
rough-dry	Kentucky-fry	hag-ride	part-privatize	type-set
spin-dry	desk-reject	hang-glide	peer-review	valet-park
sun-dry	dive-bomb	hard-boil	pinky-swear	water-torture
tumble-dry	precision-bomb	hard-code	pistol-whip	window-shop
book-keep	double-check	head-shrink	potty-train	wire-tap

Appendix B

Historical Chinese documents

1. *SJI*, 1c. BCE. *Shǐ Jì* 史記 “Records of the Grand Historian”
2. *SGZ*, 3c. *Sān Guó Zhì* 三國志 “Records of the Three Kingdoms”
3. *AYWZ*, late 3c. *Ā Yù Wáng Zhuàn* 阿育王傳 “Narrative of Ashoka”
4. *SSXY*, 5c. *Shì Shuō Xīn Yǔ* 世說新語 “A New Account of the Tales of the World”
5. *DHBW*, 9c. *Dūn Huáng Biàn Wén* 敦煌變文 “Dunhuang Transformation Texts”

Appendix C

Order-theoretic notions

In the following definitions, X is a poset and A a subset of X (which inherits the partial order).

Definition C.1 (upper/lower bound). An element $x \in X$ is an *upper bound* for A if for every $a \in A$ we have $a \leq x$ (i.e., $A \leq x$). Dually, x is a *lower bound* for A if $x \leq A$.

Definition C.2 (maximum/minimum). An element $x \in X$ is a *greatest element* or *maximum* of A if it is an upper bound of A and belongs to A. Dually, x is a *least element* or *minimum* of A if it is a lower bound of A and belongs to A.

Definition C.3 (join/meet). The least element in the set of upper bounds for A, if it exists, is called a *least upper bound* or *join* of A, denoted by $\vee A, \cup A, \sqcup A$, and the like. Dually, the greatest element in the set of lower bounds for A, if it exists, is called a *greatest lower bound* or *meet* of A, denoted by $\wedge A, \cap A, \sqcap A$, and the like.

Definition C.4 (semilattice). X is a *join-semilattice* if each pair of elements in it has a join. Dually, X is a *meet-semilattice* if each pair of elements in it has a meet. X is a *lattice* if it is both a join-semilattice and a meet-semilattice.

Definition C.5 (chain). If for every $x, y \in A$ we have $x \leq y$ or $y \leq x$ (i.e., A has no incomparable elements), then A is a *chain* in X. If X itself is a chain, it is called a *totally ordered set*.

Definition C.6 (antichain). A is an *antichain* if any two distinct elements in it are incomparable. If X itself is an antichain, it is said to be *discrete*.

Appendix D

Size-related notions in Category theory

In the definition of Category (Definition 6.3.1.1), *collection* is used instead of *set* when speaking of objects and arrows. This is because in a general Category both the object and the arrow collections may be larger than sets can handle.

Definition D.1 (small/large Category)**.** When the collection of objects and the collection of arrows of a Category are both sets, the Category is *small*; otherwise, the Category is *large*.

Example D.1. The Category **Set** and Categories of structured sets (e.g., **Pos** and **Pre**) are large. The Category **Cat** of all small Categories is also large. All finite Categories are small.

Definition D.2 (locally small Category)**.** In a Category $\mathscr{C}$ there can be multiple arrows between any pair of objects $\langle A, B\rangle$. When the collection of arrows from A to B is a set (called a *hom-set*), the Category is *locally small*. The hom-set from A to B in $\mathscr{C}$ is denoted by $\mathscr{C}(A, B)$.

Example D.2. Any small Category is locally small. Many large Categories are locally small too, such as **Set**, **Pos**, and **Pre**. An example of a non-locally-small Category is the Category of all locally small Categories and Functors between them.

Appendix E

Special Categories and Functors

Definition E.1 (opposite Category). The *opposite Category* $\mathscr{C}^{\mathrm{op}}$ of a Category $\mathscr{C}$ has the same objects as $\mathscr{C}$, while an arrow $f : C \to D$ in $\mathscr{C}^{\mathrm{op}}$ is an arrow $f : D \to C$ in $\mathscr{C}$. In other words, $\mathscr{C}^{\mathrm{op}}$ is $\mathscr{C}$ with all arrows formally turned around.

Definition E.2 (identity Functor). A Functor with the same Category as source and target is an *endofunctor*. Among the endofunctors on $\mathscr{C}$ is the *identity Functor* $\mathrm{Id}_{\mathscr{C}}$.

Definition E.3 (inclusion Functor). Given a Subcategory $\mathscr{S}$ of $\mathscr{C}$, the inclusion $\mathscr{S} \to \mathscr{C}$ which sends each object and each arrow of $\mathscr{S}$ to itself in $\mathscr{C}$ is an *inclusion Functor*.

Definition E.4 (constant Functor). Let $\mathscr{C}, \mathscr{D}$ be nonempty Categories and D be a $\mathscr{D}$-object. A *constant Functor* $\Delta_D : \mathscr{C} \to \mathscr{D}$ sends every $\mathscr{C}$-object to D and every $\mathscr{C}$-arrow to 1_D.

Definition E.5 (bifunctor). A Functor of two variables, such as $S : \mathscr{B} \times \mathscr{C} \to \mathscr{D}$, is a *bifunctor*. The source of a bifunctor is a *product Category*.

Definition E.6 (thinning Functor). Suppose $\mathscr{J}$ is a thin Category and let $\mathscr{C}$ be a "fattened" Category with the same objects as $\mathscr{J}$ but extra arrows. A *thinning Functor* from $\mathscr{C}$ to $\mathscr{J}$ takes objects to themselves and maps every arrow from A to B in $\mathscr{C}$ to the unique such arrow in $\mathscr{J}$.

Definition E.7 (contravariant Functor). A *contravariant* Functor $F : \mathscr{C} \to \mathscr{D}$ assigns to each $f : A \to B$ in $\mathscr{C}$ an arrow $F(f) : F(B) \to F(A)$ in $\mathscr{D}$ such that $F(1_A) = 1_{F(A)}$ while $F(g \circ f) = F(f) \circ F(g)$.

Note E.1. A normal (i.e., noncontravariant) Functor is *covariant*. The term *Functor* used by itself usually means covariant Functor. A contravariant Functor $G : \mathscr{C} \to \mathscr{D}$ from $\mathscr{C}$ may be replaced by a covariant Functor $\overline{G} : \mathscr{C}^{\mathrm{op}} \to \mathscr{D}$ from $\mathscr{C}$'s opposite Category (Definition $E.1$).

Definition E.8 (hom-functor). A *hom-functor* $\mathscr{C}(-,-) : \mathscr{C}^{\mathrm{op}} \times \mathscr{C} \to \mathbf{Set}$ for a locally small Category $\mathscr{C}$ is a bifunctor sending a pair of $\mathscr{C}$-objects $\langle A, B \rangle$ to their hom-set $\mathscr{C}(A, B)$ and a pair of $\mathscr{C}$-arrows $\langle f : A' \to A, g : B \to B' \rangle$ to the **Set**-arrow between $\mathscr{C}(A, B)$ and $\mathscr{C}(A', B')$.

Note E.2. A hom-functor is contravariant in its first argument (usually reformulated via an opposite Category) and covariant in its second argument. The action of a hom-functor on arrows effectively sends a $\mathscr{C}$-arrow $h : A \to B$ to another $\mathscr{C}$-arrow $g \circ h \circ f : A' \to B'$.

Bibliography

Acedo-Mattelán, Víctor & Jaume Mateu. 2014. From syntax to roots: A syntactic approach to root interpretation. In Artemis Alexiadou, Hagit Borer & Florian Schäffer (eds.), *The roots of syntax and the syntax of roots*, 33–56. Oxford: Oxford University Press.

Ackema, Peter & Ad Neeleman. 2004. *Beyond morphology: Interface conditions on word formation* (reprinted edn.). Oxford: Oxford University Press.

Acquaviva, Paolo. 2009. Roots and lexicality in distributed morphology. In Alexandra Galani, Daniel Redinger & Norman Yeo (eds.), *YPL2 special issue: Fifth York-Essex Morphology Meeting*, University of York: Department of Language and Linguistic Science.

Acquaviva, Paolo. 2014a. Distributing roots: Listemes across components in distributed morphology. *Theoretical Linguistics* 40(3/4), 277–286.

Acquaviva, Paolo. 2014b. Roots, concepts, and word structure: On the atoms of lexical semantics. In Franz Rainer, Francesco Gardani, Hans Christian Luschützky & Wolfgang Dressler (eds.), *Morphology and meaning*, 49–70. Amsterdam: John Benjamins.

Acquaviva, Paolo. 2019. Categorization as noun construction: Gender, number, and entity types. In Éric Mathieu, Myriam Dali & Gita Zareikar (eds.), *Gender and noun classification*, 41–63. Oxford: Oxford University Press.

Adams, Valerie. 2001. *Complex words in English*. London: Routledge.

Adger, David. 2003. *Core syntax: A minimalist approach*. Oxford: Oxford University Press.

Adger, David. 2010. A minimalist theory of features. In Anna Kibort & Greville Corbett (eds.), *Features: Perspectives on a key notion in linguistics*, 185–218. Oxford: Oxford University Press.

Adger, David & Peter Svenonius. 2011. Features in minimalist syntax. In Cedric Boeckx (ed.), *The handbook of linguistic minimalism*, 27–51. Oxford: Oxford University Press.

Ahlers, Timo. 2010. Komplexe C^0-phobe Verben des Deutschen [Complex C^0-phobic verbs of German]. University of Vienna masters thesis.

Ahlers, Timo. 2011. Why immobile verbs in Germanic SOV are mobile in Germanic SVO languages: Formation of complex verbs. Paper presented at Comparative Germanic Syntax Workshop 26, June 23–24, Amsterdam.

Ait-Kaci, Hassan. 1984. A lattice theoretic approach to computation based on a calculus of partially ordered type structures. University of Pennsylvania dissertation.

Alcázar, Asier & Mario Saltarelli. 2014. *The syntax of imperatives*. Cambridge: Cambridge University Press.

Aldridge, Edith. 2012. PPs and applicatives in late Archaic Chinese. *Studies in Chinese Linguistics* 33(3), 139–164.

Alexiadou, Artemis. 2014. Roots don't take complements. *Theoretical Linguistics* 40(3/4), 287–297.

Alexiadou, Artemis & Terje Lohndal. 2017. The structural configurations of root categorization. In Leah Bauke & Andreas Blümel (eds.), *Labels and roots*, 203–231. Berlin: De Gruyter Mouton.

Amblard, Maxime. 2011. Minimalist grammars and minimalist categorial grammars: Toward inclusion of generated languages. In Sylvain Pogodalla, Myriam Quatrini & Christian Retoré (eds.), *Logic and grammar: Essays dedicated to Alain Lecomte on the occasion of his 60th birthday*, 61–80. Berlin: Springer.

Amblard, Maxime, Alain Lecomte & Christian Retoré. 2010. Categorial minimalist grammars: From generative syntax to logical form. *Linguistic Analysis* 36(1), 273–306.

Arsenijevic, Boban & Joanna Ut-Seong Sio. 2009. Mediating merge: Evidence from the Cantonese *ge*. Paper presented at GLOW in Asia VII, Feb 24–27, Hyderabad.

Awodey, Steve. 2010. *Category theory* (2nd edn.). Oxford: Oxford University Press.

Baez, John. 2001. Higher-dimensional algebra and Planck-scale physics. In Craig Callender & Nick Huggett (eds.), *Physics meets philosophy at the Planck length*, 177–195. Cambridge: Cambridge University Press.

Baker, James. 2018. Split intransitivity: Thematic roles, case, and agreement. University of Cambridge dissertation.

Baker, Mark. 1988. *Incorporation: A theory of grammatical function changing*. Chicago and London: The University of Chicago Press.

Baker, Mark. 2003. *Lexical categories: Verbs, nouns and adjectives*. Cambridge: Cambridge University Press.

Baker, Mark. 2008. *The syntax of agreement and concord*. Cambridge: Cambridge University Press.

Barendregt, Henk. 1991. Introduction to generalized type systems. *Journal of functional programming* 1(2), 125–154.

Barendregt, Henk. 1992. Lambda calculi with types. In S. Abramsky, Dov M. Gabby & T. S. E. Maibaum (eds.), *Background: Computational structures*. Handbook of Logic in Computer Science, vol. 2, 118–309. Oxford: Clarendon Press.

Barr, Michael & Charles Wells. 1990. *Category theory for computing science*. London: Prentice Hall.

Bauer, Laurie. 1983. *Word-formation in English*. Cambridge: Cambridge University Press.

Bauer, Laurie. 2009. Typology of compounds. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of compounding*, 343–357. Oxford: Oxford University Press.

Bauer, Laurie. 2017. *Compounds and compounding*. Cambridge: Cambridge University Press.

Bauer, Laurie & Antoinette Renouf. 2001. A corpus-based study of compounding in English. *Journal of English Linguistics* 29(2), 101–123.

Bauke, Leah. 2014. *Symmetry breaking in syntax and the lexicon*. Amsterdam: John Benjamins.

Bauke, Leah & Tom Roeper. 2017. How Unlabeled Nodes work: Morphological derivations and the subcomponents of UG operations. In Leah Bauke & Andreas Blümel (eds.), *Labels and roots*, 233–262. Berlin: De Gruyter Mouton.

Baxter, William & Laurent Sagart. 2014. *Old Chinese: A new reconstruction*. Oxford: Oxford University Press.

de Belder, Marijke. 2013. Collective mass affixes: When derivation restricts functional structure. *Lingua* 126, 32–50.

de Belder, Marijke. 2017. The root and nothing but the root: Primary compounds in Dutch. *Syntax* 20(2), 138–169.

de Belder, Marijke & Jeroen van Craenenbroeck. 2015. How to merge a root. *Linguistic Inquiry* 46(4), 625–655.

Benincà, Paola & Cecilia Poletto. 2004. Topic, Focus, and V2: Defining the CP sublayers. In Luigi Rizzi (ed.), *The structure of CP and IP*, 52–75. Oxford: Oxford University Press.

Berwick, Robert. 2011. All you need is Merge: Biology, computation, and language from the bottom up. In Anna Di Sciullo & Cedric Boeckx (eds.), *The biolinguistic enterprise: New perspectives on the evolution and nature of the human language faculty*, 461–91. Oxford: Oxford University Press.

Berwick, Robert & Noam Chomsky. 2016. *Why only us: Language and evolution*. Cambridge MA: MIT Press.

Biberauer, Theresa. 2008. Doubling vs. omission: Insights from Afrikaans negation. In Sjef Barbiers, Olaf Koeneman, Marika Lekakou & Margreet van der Ham (eds.), *Microvariation in syntactic doubling*, 103–140. Bingley: Emerald.

Biberauer, Theresa. 2011. In defence of lexico-centric parametric variation: Two 3rd factor-constrained case studies. Paper presented at Workshop on Formal Grammar and Syntactic Variation: Rethinking Parameters, Oct 28, Madrid.

Biberauer, Theresa. 2016. Going beyond the input: 3 factors and syntactic innovation. Paper presented at Linguistics Association of Great Britain Annual Meeting, Sep 6–9, York.

Biberauer, Theresa. 2017a. Factors 2 and 3: A principled approach. In Chenchen Song & James Baker (eds.), *Cambridge occasional papers in linguistics*, vol. 10, 38–65. Theoretical and Applied Linguistics, University of Cambridge.

Biberauer, Theresa. 2017b. The Final-over-Final Condition and particles. In Michelle Sheehan, Theresa Biberauer, Ian Roberts & Anders Holmberg (eds.), *The Final-Over-Final Condition: A syntactic universal*, 187–296. Cambridge MA: MIT Press.

Biberauer, Theresa. 2018. On the significance of the periphery: A 21st century generative perspective. Paper presented at 15th Workshop on Syntax, Semantics and Phonology, June 29, Barcelona.

Biberauer, Theresa, Anders Holmberg & Ian Roberts. 2014. A syntactic universal and its consequences. *Linguistic Inquiry* 45(2), 169–225.

Biberauer, Theresa & Ian Roberts. 2010. Subjects, tense and verb-movement. In Theresa Biberauer, Anders Holmberg, Ian Roberts & Michelle Sheehan (eds.), *Parametric variation: Null subjects in minimalist theory*, 263–303. Cambridge: Cambridge University Press.

Biberauer, Theresa & Ian Roberts. 2015. Rethinking formal hierarchies: A proposed unification. In James N. Chancharu, Xuhui Freddy Hu & Moreno Mitrović (eds.), *Cambridge occasional papers in linguistics*, vol. 7, 1–31. Theoretical and Applied Linguistics, University of Cambridge.

Biberauer, Theresa & Ian Roberts. 2017. Parameter setting. In Adam Ledgeway & Ian Roberts (eds.), *The Cambridge handbook of historical syntax*, 134–162. Cambridge: Cambridge University Press.

Bird, Steven. 1995. *Computational phonology: A constraint-based approach*. Cambridge: Cambridge University Press.

Bird, Steven & Ewan Klein. 1994. Phonological analysis in typed feature systems. *Computational linguistics* 20(3), 455–491.

Boeckx, Cedric. 2011. Some reflections on Darwin's problem in the context of Cartesian biolinguistics. In Anna Di Sciullo & Cedric Boeckx (eds.), *The biolinguistic enterprise: New perspectives on the evolution and mature of the human language faculty*, 42–64. Oxford: Oxford University Press.

Boisseau, Guillaume & Jeremy Gibbons. 2018. What you needa know about Yoneda: Profunctor optics and the Yoneda lemma (Functional Pearl). *Proceedings of the ACM on Programming Languages* 2(ICFP), 84.1–27.

Borer, Hagit. 2005a. *In name only*. Structuring Sense, vol. 1. Oxford: Oxford University Press.

Borer, Hagit. 2005b. *The normal course of events*. Structuring Sense, vol. 2. Oxford: Oxford University Press.

Borer, Hagit. 2013. *Taking form*. Structuring Sense, vol. 3. Oxford: Oxford University Press.

Bowers, John. 2018. *Deriving syntactic relations*. Cambridge: Cambridge University Press.

Bošković, Željko. 2016. What is sent to spell-out is phases, not phasal complements. *Linguistica* 56(1), 25–66.

Bošković, Željko. 2018. On movement out of moved elements, labels, and phases. *Linguistic Inquiry* 49(2), 247–282.

Bradley, Tai-Danae. 2018. What is applied category theory? Manuscript (arXiv:1809.05923), City University of New York.

Brannan, David, Matthew Esplen & Jeremy Gray. 2012. *Geometry* (2nd edn.). Cambridge: Cambridge University Press.

Brody, Michael. 2002. On the status of representations and derivations. In Samuel Epstein & T. Daniel Seely (eds.), *Derivation and explanation in the minimalist program*, 19–41. Oxford: Blackwell.

Butt, Miriam. 2010. The light verb jungle: Still hacking away. In Mengistu Amberber, Brett Baker & Mark Harvey (eds.), *Complex predicates in cross-linguistic perspective*, 48–78. Cambridge: Cambridge University Press.

Caha, Pavel. 2009. The nanosyntax of case. Universitetet i Tromsø dissertation.

Cardelli, Luca. 1988. A semantics of multiple inheritance. *Information and Computation* 76, 138–164.

Cardelli, Luca. 2004. Type systems. In Allen Tucker (ed.), *Computer science handbook* (2nd edn.), chap. 97. Chapman & Hall/CRC.

Carnap, Rudolf. 1956. *Meaning and necessity: A study in semantics and modal logic*. Chicago: University of Chicago Press.

Carpenter, Bob. 1992. *The logic of typed feature structures*. Cambridge: Cambridge University Press.

Chao, Yuan-Ren. 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.

Chen, Cuizhu. 2009. Hànyǔ rénchēng dàicí kǎolùn 汉语人称代词考论 [An examination of Chinese pronouns]. Central China Normal University dissertation.

Chen, Lishi. 2017. Wǎngluò yǔjìng xià xuéshēng shǐyòng biāodiǎn fúhào zhǔguānhuà qíngkuàng de diàochá yánjiū 网络语境下学生使用"标点符号"主观化情况的调查研究 [An investigation on the subjectiveness of students' usage of "punctuation marks" in internet context]. *Modern Chinese* 7, 112–115.

Cheng, Eugenia. 2015. *Cakes, custard, and category theory: Easy recipes for understanding complex mathematics*. London: Profile Books.

Cheng, L.-S. Lisa, C.-T. James Huang & C.-C. Jane Tang. 1996. Negative particle questions: A dialectal comparison. In James Black & Virginia Motapanyane (eds.), *Microparametric syntax and dialect variation*, 41–78. Amsterdam: John Benjamins.

Chomsky, Noam. 1957. *Syntactic structure*. The Hague: Mouton.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge MA: MIT Press.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, Noam. 1986. *Barriers*. Cambridge MA: MIT press.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 89–156. Cambridge MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A life in language*, 1–52. Oxford: Oxford University Press.

Chomsky, Noam. 2004. Beyond explanatory adequacy. In Adriana Belletti (ed.), *Structures and beyond*. The Cartography of Syntactic Structures, vol. 3, 104–131. Oxford: Oxford University Press.

Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36(1), 1–22.

Chomsky, Noam. 2007. Approaching UG from below. In Uli Sauerland & Hans-Martin Gärtner (eds.), *Interfaces + recursion = language? Chomsky's Minimalism and the view from syntax-semantics*, 1–30. Berlin: Mouton de Gruyter.

Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos Otero & Maria Zubizarreta (eds.), *Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud*, 134–166. Cambridge MA: MIT Press.

Chomsky, Noam. 2013. Problems of projection. *Lingua* 130, 33–49.

Chomsky, Noam. 2015. Problems of projection: Extensions. In Elisa Di Domenico, Cornelia Hamann & Simona Matteini (eds.), *Structures, strategies and beyond: Studies in honour of Adriana Belletti*, 1–16. Amsterdam: John Benjamins.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English.* New York: Harper & Row.

Church, Alonzo. 1940. A formulation of the simple theory of types. *The journal of symbolic logic* 5(2), 56–68.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford: Oxford University Press.

Cinque, Guglielmo & Luigi Rizzi. 2008. The cartography of syntactic structures. *Studies in linguistics* 2, 42–58.

Cinque, Guglielmo & Luigi Rizzi. 2010. The cartography of syntactic structures. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 65–78. Oxford: Oxford University Press.

Clark, Eve. 1993. *The lexicon in acquisition*. Cambridge: Cambridge University Press.

Clark, John, Colin Yallop & Janet Fletcher. 2007. *An introduction to phonetics and phonology* (3rd edn.). Oxford: Blackwell.

Claudi, Ulrike & Bernd Heine. 1986. On the metaphorical base of grammar. *Studies in Language* 10(2), 297–335.

Clements, George. 1985. The geometry of phonological features. *Phonology* 2(1), 225–252.

Clements, George. 2003. Feature economy in sound systems. *Phonology* 20(3), 287–333.

Clements, George & Elizabeth Hume. 1995. The internal organization of speech sounds. In John Goldsmith (ed.), *The handbook of phonological theory*, 245–306. Cambridge MA: Blackwell.

Cohen, Henri & Claire Lefebvre. 2005. Bridging the category divide. In Henri Cohen & Claire Lefebvre (eds.), *Handbook of categorization in cognitive science*, 1–15. Oxford: Oxford University Press.

Collins, Chris & Paul Postal. 2012. *Imposters: A study of pronominal agreement*. Cambridge MA: MIT Press.

Collins, Chris & Edward Stabler. 2016. A formalization of minimalist syntax. *Syntax* 19(1), 43–78.

Collins, Christopher. 2017. Merge(X, Y) = {X, Y}. In Leah Bauke & Andreas Blümel (eds.), *Labels and roots*, 47–68. Berlin: De Gruyter Mouton.

Collins, Jordan. 2005. A history of the theory of types with special reference to developments after the second edition of *Principia Mathematica*. McMaster University masters thesis.

Copestake, Ann. 2002. *Implementing typed feature structure grammars*. Stanford: Center for the Study of Language and Information.

Corbett, Greville. 2010. Features: Essential notions. In Anna Kibort & Greville G. Corbett (eds.), *Features: Perspectives on a key notion in linguistics*, 17–36. Oxford: Oxford University Press.

Corbett, Greville. 2012. *Features*. Cambridge: Cambridge University Press.

van Craenenbroeck, Jeroen. 2014. On diagnosing complement-taking roots. *Theoretical Linguistics* 40(3/4), 361–373.

Crole, Roy. 1993. *Categories for types*. Cambridge: Cambridge University Press.

Cuervo, María. 2003. Datives at large. Massachusetts Institute of Technology dissertation.

Davey, Brian & Hilary Priestley. 2002. *Introduction to lattices and order*. Cambridge: Cambridge University Press.

Déprez, Viviane. 2006. On the conceptual role of number. In Chiyo Nishida & Jean-Pierre Montreuil (eds.), *New perspectives on romance linguistics*, vol. 1, 67–82. Amsterdam: John Benjamins.

Di Sciullo, Anna-Maria & Edwin Williams. 1987. *On the definition of word*. Cambridge MA: MIT Press.

Dixon, Robert. 1977. *Where have all the adjectives gone?* 1–60. Amsterdam: John Benjamins.

Dougherty, Daniel. 1993. Closed categories and categorial grammar. *Notre Dame Journal of Formal Logic* 34(1), 36–49.

Douglas, Jamie. 2018. Māori subject extraction. *Glossa: A Journal of General Linguistics* 7(1), 110.1–34.

Dresher, B. Elan. 2009. *The contrastive hierarchy in phonology*. Cambridge: Cambridge University Press.

Durrell, Martin. 2011. *Hammer's German grammar and usage* (5th edn.). London: Routledge.

Ehresmann, Andrée Charles & Jean-Paul Vanbremeersch. 2007. *Memory evolutive systems: Hierarchy, emergence, cognition*. Amsterdam: Elsevier.

Eilenberg, Samuel & Saunders Mac Lane. 1945. General theory of natural equivalences. *Transactions of the American Mathematical Society* 58(2), 231–294.

Ellerman, David. 2007. Adjoints and emergence: Applications of a new theory of adjoint functors. *Axiomathes* 17(1), 19–39.

Ellis, John. 1993. *Language, thought, and logic*. Evanston, Illinois: Northwestern University Press.

Embick, David. 2015. *The morpheme: A theoretical introduction*. Berlin: De Gruyter Mouton.

Embick, David & Alec Marantz. 2008. Architecture and blocking. *Linguistic Inquiry* 39(1), 1–53.

Epstein, Samuel, Hisatsugu Kitahara & T. Daniel Seely. 2012. Structure building that can't be. In Myriam Uribe-Etxebarria & Vidal Valmala (eds.), *Ways of structure building*, 253–270. Oxford: Oxford University Press.

Erlewine, Michael. 2017. Low sentence-final particles in Mandarin Chinese and the Final-over-Final Constraint. *Journal of East Asian Linguistics* 26(1), 37–75.

Erné, Marcel. 2004. Adjunctions and Galois connections: Origins, history and development. In Klaus Denecke, Marcel Erné & Shelly Wismath (eds.), *Galois connections and applications*, 1–138. Dortrecht: Springer.

Fabb, Nigel. 1998. Compounding. In Andrew Spencer & Arnold Zwicky (eds.), *The handbook of morphology*, 66–83. Oxford: Blackwell.

Farmer, William. 2008. The seven virtues of simple type theory. *Journal of Applied Logic* 6(3), 267–286.

Feng, Shengli. 2010. Lùn yǔtǐ de jīzhì jí qí yǔfǎ shǔxìng 论语体的机制及其语法属性 [On mechanisms of register system and its grammatical property]. *Studies of the Chinese Language* 400–412.

Feng, Shengli. 2012. Yǔtǐ yǔfǎ: xíngshì-gōngnéng duìyìnglǜ de yǔyán tànsuǒ 语体语法:"形式-功能对应律"的语言探索 [Register grammar: A linguistic exploration of the form-function mapping]. *Contemporary Rhetoric* 6, 3–12.

Feng, Shengli. 2015. Light-verb syntax between English and Classical Chinese. In Audrey Li, Andrew Simpson & Wei-Tien Dylan Tsai (eds.), *Chinese syntax in a cross-linguistic perspective*, 229–250. Oxford: Oxford University Press.

Feng, Shengli. 2019. *Prosodic syntax in Chinese: Theory and facts*. London: Routledge.

Flemming, Edward. 2004. Contrast and perceptual distinctiveness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 232–276. Cambridge: Cambridge University Press.

Folli, Raffaella & Heidi Harley. 2005. Flavors of v. In Paula Kempchinsky & Roumyana Slabakova (eds.), *Aspectual inquiries*, 95. Dortrecht: Springer.

Folli, Raffaella & Heidi Harley. 2007. Causation, obligation, and argument structure: On the nature of little v. *Linguistic Inquiry* 38(2), 197–238.

Folli, Raffaella & Heidi Harley. in press. A head movement approach to Talmy's typology. *Linguistic Inquiry* .

Fong, Brendan & David Spivak. 2018. Seven sketches in compositionality: An invitation to applied category theory. Manuscript (2018-10-12 edn.), Massachusetts Institute of Technology.

Fortmann, Christian. 2007. Bewegungsresistente Verben [Movement-resistent verbs]. *Zeitschrift für Sprachwissenschaft* 26(1), 1–40.

Fowlie, Meaghan. 2013. Multiple multiple spellout. In Theresa Biberauer (ed.), *Challenges to linearization*, 129–169. Mouton de Gruyter.

Frampton, John & Sam Gutmann. 2000. Agreement is feature sharing. Manuscript, Northeastern University.

Frampton, John & Sam Gutmann. 2006. How sentences grow in the mind. In Boeckx Cedric (ed.), *Agreement systems*, 121–157. Amsterdam: John Benjamins.

Frascarelli, Mara & Roland Hinterhölzl. 2007. Types of topics in German and Italian. In Kerstin Schwabe & Susanne Winkler (eds.), *On information structure, meaning and form*, 87–116. Amsterdam: John Benjamins.

Freyd, Peter. 1964. *Abelian categories*. New York: Harper & Row.

Freywald, Ulrike & Horst Simon. 2007. Wenn die Wortbildung die Syntax stört: Über Verben, die nicht in V2 stehen können [When word building disturbs syntax: On verbs that cannot stand in V2]. In Maurice Kauffer & René Métrich (eds.), *Verbale Wortbildung im Spannungsfeld zwischen Wortsemantik, Syntax und Rechtschreibung*, 181–94. Tübingen: Stauffenburg.

Fromkin, Victoria, Robert Rodman, Peter Collins & David Blair. 1996. *An introduction to language* (3rd edn.). Sydney: Harcourt Brace.

Ganesalingam, Mohan. 2013. *The language of mathematics*. Heidelberg: Springer.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum & Ivan Sag. 1985. *Generalized phrase structure grammar*. Cambridge MA: Harvard University Press.

Gazdar, Gerald & Geoffrey Pullum. 1982. *Generalized phrase structure grammar: A theoretical synopsis*. Indiana University Linguistics Club.

German, James & Laurent Prévot. 2016. Sentence-final particles in Singapore English: Are they pragmatic or phonological? In *Speech prosody 8*, 883–887. Boston University.

Gillibert, Jean & Christian Retoré. 2014. Category theory, logic and formal linguistics: Some connections, old and new. *Journal of Applied Logic* 12, 1–13.

del Gobbo, Francesca. 2014. Classifiers. In C.-T. James Huang, Y.-H. Audrey Li & Andrew Simpson (eds.), *The handbook of Chinese linguistics*, 26–48. Malden: John Wiley & Sons.

Gödel, Kurt. 1944. Russell's mathematical logic. In Paul Schilpp (ed.), *The philosophy of Bertrand Russell*, 123–153. Northwestern University, Evanston and Chicago.

Goldblatt, Robert. 2006. *Topoi: The categorial analysis of logic* (revised edn.). Mineola: Dover.

Goldsmith, John. 1979. *Autosegmental phonology*. Cambridge MA: MIT Press.

Greenberg, Joseph. 1960. A quantitative approach to the morphological typology of language. *International journal of American linguistics* 26(3), 178–194.

Grimshaw, Jane. 1991. Extended projection. Manuscript, Brandeis University.

Grimshaw, Jane. 2000. Locality and extended projection. In Peter Coopmans, Martin Everaert & Jane Grimshaw (eds.), *Lexical specification and insertion*, 115–134. Amsterdam: John Benjamins.

Grimshaw, Jane. 2005. *Words and structure*. Stanford: Center for the Study of Language and Information.

Gupta, Anthea. 1992. The pragmatic particles of Singapore colloquial English. *Journal of Pragmatics* 18(1), 31–57.

ten Hacken, Pius. 2016. *The semantics of compounding*. Cambridge: Cambridge University Press.

Haegeman, Liliane. 2014. West Flemish verb-based discourse markers and the articulation of the speech act layer. *Studia Linguistica* 68(1), 116–139.

Haegeman, Liliane & Virginia Hill. 2013. The syntacticization of discourse. In Raffaella Folli, Cristina Sevdali & Robert Truswell (eds.), *Syntax and its limits*, 370–390. Oxford: Oxford University Press.

Haider, Hubert. 2010. *The syntax of German*. Cambridge: Cambridge University Press.

Hale, Ken & S. Jay Keyser. 2002. *Prolegomenon to a theory of argument structure*. MIT Press.

Hall, Daniel. 2007. The role and representation of contrast in phonological theory. University of Toronto dissertation.

Halle, Morris. 1969. How not to measure length of lexical representations and other matters. *Journal of Linguistics* 5(2), 305–308.

Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Ken Hale & S. Jay Keyser (eds.), *Essays in linguistics in honor of Sylvain Bromberger* (The View from Building 20), 111–176. Cambridge MA: MIT Press.

Halle, Morris & Alec Marantz. 1994. Some key features of distributed morphology. In Andrew Carnie, Heidi Harley & Tony Bures (eds.), *MIT working papers in linguistics*, vol. 21, 275–288. Department of Linguistics, MIT.

Harbour, Daniel. 2008. *Morphosemantic number: From Kiowa noun classes to UG number features*. Dortrecht: Springer.

Harbour, Daniel. 2011. Valence and atomic number. *Linguistic inquiry* 42(4), 561–594.

Harbour, Daniel. 2013. "Not plus" isn't "not there": Bivalence in person, number, and gender. In Ora Matushansky & Alec Marantz (eds.), *Distributed morphology today: Morphemes for Morris Halle*, 135–150. Cambridge MA: MIT Press.

Harbour, Daniel. 2016. *Impossible persons*. Cambridge MA: MIT Press.

Harley, Heidi. 1994. Hug a tree: Deriving the morphosyntactic feature hierarchy. In Andrew Carnie, Heidi Harley & Tony Bures (eds.), *MIT working papers in linguistics*, vol. 21, 289–320. Department of Linguistics, MIT.

Harley, Heidi. 1995. Subjects, events, and licensing. Massachusetts Institute of Technology dissertation.

Harley, Heidi. 2008. On the causative construction. In Shigeru Miyagawa (ed.), *Handbook of Japanese linguistics*, 20–53. Oxford: Oxford University Press.

Harley, Heidi. 2009a. Compounding in distributed morphology. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of compounding*, 129–144. Oxford: Oxford University Press.

Harley, Heidi. 2009b. The morphology of nominalizations and the syntax of *v*P. In Anastasia Giannakidou & Monika Rathert (eds.), *Quantification, definiteness, and nominalization*, 321–343. Oxford: Oxford University Press.

Harley, Heidi. 2014. On the identity of roots. *Theoretical Linguistics* 40(3/4), 225–276.

Harley, Heidi. 2017. The "bundling" hypothesis and the disparate functions of little v. In Roberta D'Alessandro, Irene Franco & Ángel Gallego (eds.), *The verbal domain*, 3–28. Oxford: Oxford University Press.

Harley, Heidi & Rolf Noyer. 1999. Distributed morphology. *Glot International* 4(4), 3–9.

Harley, Heidi & Elizabeth Ritter. 2002. Person and number in pronouns: A feature-geometric analysis. *Language* 78(3), 482–526.

Harnad, Stevan. 2005. To cognize is to categorize: Cognition is categorization. In Henri Cohen & Claire Lefebvre (eds.), *Handbook of categorization in cognitive science*, 19–43. Oxford: Oxford University Press.

He, Zhensheng. 2010. Lùn bìngliè liáncí hé gēn tóng yǔ jí yǐjí de biǎoyì gōngnéng 论并列连词"和、跟、同、与、及、以及"的表义功能 [On the semantics of the coordinative conjunctions *hé*, *gēn*, *tóng*, *yǔ*, *jí*, and *yǐjí*]. *Journal of Language and Literature Studies* 10A, 70–71.

Heim, Johannes, Hermann Keupdjio, Zoe Wai-Man Lam, Adriana Osa-Gómez, Sonja Thoma & Martina Wiltschko. 2016. Intonation and particles as speech act modifiers: A syntactic analysis. *Studies in Chinese Linguistics* 37(2), 109–129.

Hicks, Glyn. 2009. *Tough*-constructions and their derivation. *Linguistic Inquiry* 40(4), 535–566.

Higginbotham, James. 1998. On higher-order logic and natural language. In *Proceedings of the British Academy*, vol. 95, 1–27.

Hill, Virginia. 2007. Vocatives and the pragmatics-syntax interface. *Lingua* 117(12), 2077–2105.

Hopper, Paul & Elizabeth Traugott. 2003. *Grammaticalization* (2nd edn.). Cambridge: Cambridge University Press.

Hornstein, Norbert. 2009. *A theory of syntax: Minimal operations and universal grammar*. Cambridge: Cambridge University Press.

Hornstein, Norbert & Jairo Nunes. 2008. Adjunction, labeling, and bare phrase structure. *Biolinguistics* 2(1), 57–86.

Hu, Mingyang. 1981. Běijīnghuà de yǔqì zhùcí hé tàncí 北京话的语气助词和叹词 [Sentence-final particles and interjections in Beijing Mandarin]. *Studies of the Chinese Language* 5(6).

Hu, Xuhui. 2018. *Encoding events: Functional structure and variation*. Oxford: Oxford University Press.

Hu, Zengyi & Ke Chao. 1986. *Èwēnkè yǔ jiǎnzhì* 鄂温克语简志 [*A description of the Evenki language*]. Beijing: Publishing House of Minority Nationalities.

Huang, C.-T. James. 2006. Resultatives and unaccusatives: A parametric view. *Bulletin of the Chinese Linguistic Society of Japan* 2006(253), 1–43.

Huang, C.-T. James, Y.-H. Audrey Li & Yafei Li. 2009. *The syntax of Chinese*. Cambridge: Cambridge University Press.

Huang, C.-T. James & Masao Ochi. 2004. Syntax of the hell: Two types of dependencies. In Moulton Keir & Matthew Wolf (eds.), *Proceedings of the 34th Conference of the North Eastern Linguistic Society*, 279–294.

Huang, Cheng-Teh James. 2015. On syntactic analyticity and parametric theory. In Audrey Li, Andrew Simpson & Wei-Tien Dylan Tsai (eds.), *Chinese syntax in a cross-linguistic perspective*, 1–48. Oxford: Oxford University Press.

Ito, Junko & Armin Mester. 2003. *Japanese morphophonemics: Markedness and word structure*. Cambridge MA: MIT Press.

Jacobson, Pauline. 2012. Direct compositionality. In Markus Werning, Wolfram Hinzen & Edouard Machery (eds.), *The Oxford handbook of compositionality*, 109–128. Oxford: Oxford University Press.

Jakobson, Roman, C. Gunnar Fant & Morris Halle. 1952. *Preliminaries to speech analysis*. Cambridge MA: MIT Press.

Jakobson, Roman & John Lotz. 1949. Notes on the French phonemic pattern. *Word* 5(2), 151–158.

Jespersen, Otto. 1965. *Morphology*, vol. 4. London/Copenhagen: George Allen & Unwin / Ejnar Munksgaard.

Jiang, Shaoyu. 2013. Xiānqín hànyǔ de dòngbīn guānxi hé jíwùxìng 先秦漢語的動賓關係和及物性 [Verb-Object relationships and transitivity in pre-Qin Chinese]. *Bulletin of Chinese Linguistics* 7(2), 59–113.

Johnson, Kyle. 2003. Towards an etiology of adjunct islands. *Nordlyd* 31(1), 187–215.

Kan, Zhehua. 2007. Causative verb and causative constructions at the syntax-semantics interface. Shanghai Jiao Tong University dissertation.

Kay, Paul. 1966. Comment on Benjamin N. Colby's ethnographic semantics: A preliminary survey. *Current Anthropology* 7(1), 20–23.

Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge MA: MIT Press.

Kayne, Richard. 2005. *Movement and silence*. Oxford: Oxford University Press.

Kayne, Richard. 2016. The silence of heads. *Studies in Chinese linguistics* 37(1), 1–37.

Kelly, Justin. 2013. The syntax-semantics interface in distributed morphology. Georgetown University dissertation.

Kiparsky, Paul. 1982. From cyclic phonology to lexical phonology. In Harry van der Hulst & Norval Smith (eds.), *The structure of phonological representations*, vol. 1, 131–157. Dordrecht: Foris.

Koopman, Hilda. 1995. On verbs that fail to undergo V-second. *Linguistic Inquiry* 26(1), 137–163.

Kornai, András. 2008. *Mathematical linguistics*. London: Springer.

Kornfeld, Laura. 2009. IE, Romance: Spanish. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of compounding*, 436–452. Oxford: Oxford University Press.

Kracht, Marcus. 2007. Introduction to linguistics. Lecture notes for LING 20, UCLA.

Kratzer, Angelika. 1996. Severing the external argument from its verb. In Johan Rooryck & Laurie Zaring (eds.), *Phrase structure and the lexicon*, 109–137. Dortrecht: Kluwer.

Kratzer, Angelika & Irene Heim. 1998. *Semantics in generative grammar*. Oxford: Blackwell.

Lamarre, Christine. 1995. Běifāng guānhuà lǐ biǎoshì kěnéng de dòngcí cíwěi le 北方官话里表示可能的动词词尾"了" [On the potential suffix *le* in Northern Mandarin]. *Studies of the Chinese Language* 4, 267–278.

Lamarre, Christine. 2001. Cóng pǔtōnghuà lǐ gēn dé yǒuguān de jǐgè géshì qù tàntǎo fāngyán lèixíngxué 从普通话里跟"得"有关的几个格式去探讨方言类型学 [From several *de*-related constructions in Standard Mandarin to dialect typology]. *Studies in Language and Linguistics* 2, 7–18.

Lambek, Joachim. 1988. Categorial and categorical grammars. In Richard Oehrle, Emmon Bach & Deirdre Wheeler (eds.), *Categorial grammars and natural language structures*, 297–317. Dortrecht: Kluwer.

Lambek, Joachim. 2008. *From word to sentence: A computational algebraic approach to grammar*. Milan: Polimetrica S.a.s.

Lambek, Joachim. 2011. Compact monoidal categories from linguistics to physics. In Bob Coecke (ed.), *New structures for physics*, 467–487. Berlin: Springer.

Lambek, Joachim & Philip Scott. 1986. *Introduction to higher order categorical logic*. Cambridge: Cambridge University Press.

Landman, Fred. 2000. *Events and plurality: The Jerusalem lectures*. Dortrecht: Springer.

Landry, Elaine. 2017. *Categories for the working philosopher*. Oxford: Oxford University Press.

Larson, Richard. 1988. On the double object construction. *Linguistic Inquiry* 19(3), 335–391.

Lawvere, F. William. 1964. An elementary theory of the category of sets. *Proceedings of the National Academy of Sciences of the United States of America* 52(6), 1506–1511.

Lawvere, F. William & Stephen Schanuel. 1986. *Categories in continuum physics*. Berlin: Springer.

Lawvere, F. William & Stephen Schanuel. 2009. *Conceptual mathematics: A first introduction to categories* (2nd edn.). Cambridge: Cambridge University Press.

Leben, William. 1973. Suprasegmental phonology. Massachusetts Institute of Technology dissertation.

Lecomte, Alain. 2005. Categorial grammar for Minimalism. In Claudia Casadio, Philip Scott & Robert Seely (eds.), *Language & grammar: Studies in mathematical linguistics and natural language*, 163–168. Stanford: Center for the Study of Language and Information.

Lecomte, Alain. 2008. Semantics in minimalist categorial grammars. In Philippe de Groote (ed.), *Proceedings of the 13th Conference on Formal Grammar*, 41–60. Stanford: Center for the Study of Language and Information.

Leinster, Tom. 2016. *Basic category theory* (corrected edn.). Cambridge: Cambridge University Press (originally). ArXiv:1612.09375.

Levin, Theodore, Paulina Lyskawa & Rodrigo Ranero. 2018. Microvariation in the realization of Eastern Mayan number agreement. Paper presented at Cambridge Comparative Syntax 7, May 10–12, Cambridge.

Li, Charles & Sandra Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Li, Yunfeng. 2017. Xiàndài hànyǔ yǔqìcí o yo lo de huàyǔ gōngnéng jí qí zhǔguānxìng yánjiū 现代汉语语气词"哦、哟、咯"的话语功能及其主观性研究 [On the discourse functionality and subjectivity of the Modern Chinese sentence-final particles *o*, *yo*, and *lo*]. Anhui University masters thesis.

Lian, Chin-Fa. 2015. Xiàndài mǐnnányǔ wú de duōchóng gōngnéng: cóng jiēcéng jiégòu rùshǒu 現代閩南語「無」的多重功能: 從階層結構入手 [Polyfunctionality of *bo* in Modern Southern Min: A case of its interpretation in tandem with hierarchical positions]. *Language and Linguistics* 16(2), 169–186.

Libben, Gary. 2006. Why study compound processing? An overview of the issues. In Gary Libben & Gonia Jarema (eds.), *The representation and processing of compound words*, 1–22. Oxford: Oxford University Press.

Lieber, Rochelle & Pavol Štekauer. 2009. Introduction: Status and definition of compounding. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of compounding*, 4–18. Oxford: Oxford University Press.

Lim, Lisa. 2007. Mergers and acquisitions: On the ages and origins of Singapore English particles. *World Englishes* 26(4), 446–473.

Lin, Jimmy. 2004. Event structure and the encoding of arguments: The syntax of the Mandarin and English verb phrase. MIT dissertation.

Liu, Cheng-Hui. 2000. Gǔ hànyǔ dòngcí de fùhéhuà yǔ shǐchénghuà 古漢語動詞的複合化與使成化 [On verb compounding and and the emergence of the *shicheng* construction]. *Chinese Studies* 18, 231–260.

Liu, Xunning. 1985. Xiàndài hànyǔ cíwěi le de láiyuán 现代汉语词尾“了”的来源 [The origin of Modern Chinese suffix *le*]. *Dialect* 2, 128–133.

Loke, Kit-ken. 1997. The grammaticalisation and regrammaticalisation of Chinese numeral classifier morphemes. *Journal of Chinese Linguistics* 25(1), 1–20.

Lowenstamm, Jean. 2008. On little n, √, and types of nouns. In Jutta Hartmann, Veronika Hegedűs & Henk van Riemsdijk (eds.), *Sounds of silence: Empty elements in syntax and phonology*, 105–144. Amsterdam: Elsevier.

Lü, Shuxiang. 1979. *Hànyǔ yǔfǎ fēnxī wèntí* 汉语语法分析问题 [*Problems in the analysis of Chinese grammar*]. Beijing: Commercial Press.

Lü, Shuxiang. 1999. *Xiàndài hànyǔ bābǎi cí* 现代汉语八百词 [*800 words in Modern Chinese*] (enlarged edn.). Beijing: Commercial Press.

Lundberg, George. 1929. *Social research*. New York: Longmans, Green and Co.

Ma, Xiao. 1992. Xiàndài hànyǔ zhōng de A_1la + A_2la + $A_n$la géshì 现代汉语中的“A_1啦 + A_2啦 + A_n啦”格式 [On the A_1*la* + A_2*la* + A_n*la* construction in Modern Chinese]. *Journal of Huaiyin Teachers College (Social Sciences Edition)* 14(1), 85–88.

Mac Lane, Saunders. 1998. *Categories for the working mathematician* (2nd edn.). New York: Springer.

Magnan, Francois & Gonzalo Reyes. 1994. Category theory as a conceptual tool in the study of cognition. In John Macnamara & Gonzalo Reyes (eds.), *The logical foundations of cognition*, 57–90. Oxford: Oxford University Press.

Marantz, Alec. 1995. *Cat* as a phrasal idiom: Consequences of late insertion in distributed morphology. Manuscript, Massachusetts Institute of Technology.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In *UPenn working papers in linguistics*. 2, vol. 4, 201–225. Philadelphia: Penn Graduate Linguistics Society.

Marantz, Alec. 2001. Words. Paper presented at 20th West Coast Conference on Formal Linguistics, Feb 23–25, University of Southern California.

Marantz, Alec. 2013. Verbal argument structure: Events and participants. *Lingua* 130, 152–168.

Marchand, Hans. 1969. *The categories and types of present-day English word-formation: A synchronic-diachronic approach*. München: Beck.

Mardale, Alexandru. 2011. Prepositions as a semilexical category. *Bucharest Working Papers in Linguistics* 13(2), 57–73.

Martin, Johannes. 1986. *Data types and data structures*. London: Prentice-Hall International.

Matushansky, Ora & Alec Marantz. 2013. *Distributed morphology today: Morphemes for Morris Halle*. Cambridge MA: MIT Press.

May, Robert. 1985. *Logical form: Its structure and derivation*. Cambridge MA: MIT press.

Mazzola, Guerino & Maria Mannone. 2016. Global functorial hypergestures over general skeleta for musical performance. *Journal of Mathematics and Music* 10(3), 227–243.

McCarthy, John. 1988. Feature geometry and dependency: A review. *Phonetica* 45(2–4), 84–108.

McCawley, James. 1992. Justifying part-of-speech assignments in Mandarin Chinese. *Journal of Chinese Linguistics* 20(2), 211–246.

McIntyre, Andrew. 2002. Verb-second and backformations and scalar prefix verbs in German: The interaction between morphology, syntax and phonology. Manuscript, University of Leipzig.

Megerdoomian, Karine. 2002. Beyond words and phrases. University of Southern California dissertation.

Mei, Zulin. 1994. Tángdài sòngdài gòngtóngyǔ de yǔfǎ hé xiàndài fāngyán de yǔfǎ 唐代、宋代共同語的語法和現代方言的語法 [On the grammar of the common language in Tang and Song dynasties and the grammars of Modern Chinese dialects]. *Languages in China and Linguistics* 2, 61–97.

Meinke, Karl & John Tucker. 1993. *Many-sorted logic and its applications*. Chichester: John Wiley & Sons.

Mencken, Henry Louis. 2009. *The American language: An inquiry into the development of English in the United States* (reprinted, revised edn.). New York: Cosimo.

Merchant, Jason. 2017. On categorizers and selection. Paper presented at Cambridge Comparative Syntax 6, May 4–6, Cambridge.

Milewski, Bartosz. 2018. *Category theory for programmers*. Blurb.

Mitchell, John. 2003. *Concepts in programming languages*. Cambridge: Cambridge University Press.

Miyagawa, Shigeru. 1998. *(S)ase* as an elsewhere causative and the syntactic nature of words. *Journal of Japanese Linguistics* 16(1), 67–110.

Miyagawa, Shigeru. 2017. *Agreement beyond phi*. Cambridge MA: MIT Press.

Montague, Richard. 1974. Universal grammar. In Richmond Thomason (ed.), *Formal philosophy*, 222–246. New Haven and London: Yale University Press.

Moortgat, Michael. 2007. Symmetries in natural language syntax and semantics: The Lambek-Grishin calculus. In Daniel Leivant & Ruy de Queiroz (eds.), *International workshop on logic, language, information, and computation*, 264–284. Berlin: Springer.

Moortgat, Michael. 2009. Symmetric categorial grammar. *Journal of Philosophical Logic* 38(6), 681–710.

Moortgat, Michael. 2010. Symmetric categorial grammar: Residuation and Galois connections. *Linguistic Analysis* 36(1), 143–166.

Moshier, M. Drew & William Rounds. 1987. A logic for partially specified data structures. In *Proceedings of the 14th ACM SIGACT-SIGPLAN symposium on principles of programming languages*, 156–167. ACM.

Myler, Neil. 2016. *Building and interpreting possession sentences*. Cambridge MA: MIT Press.

Nederpelt, Rob & Herman Geuvers. 2014. *Type theory and formal proof: An introduction*. Cambridge: Cambridge University Press.

Nedjalkov, Igor. 1997. *Evenki*. London: Routledge.

Neeleman, Ad & Hans van der Koot. 2005. Syntactic haplology. In Martin Everaert & Henk van Riemsdijk (eds.), *The Blackwell companion to syntax*, vol. 4, 685–710. Oxford: Blackwell.

Neeleman, Ad & Hans van de Koot. 2010. Theoretical validity and psychological reality of the grammatical code. In Martin Everaert, Tom Lentz, Hannah De Mulder, Øystein Nilsen & Arjen Zondervan (eds.), *The linguistics enterprise: From knowledge of language to knowledge in linguistics*, 183–212. Amsterdam: John Benjamins.

Nguyễn, Đình Hoà. 1997. *Vietnamese*. Amsterdam: John Benjamins.

Nishiyama, Kunio. 2016. The theoretical status of *ren'yoo* (stem) in Japanese verbal morphology. *Morphology* 26(1), 65–90.

Nübling, Damaris & Renata Szczepaniak. 2008. On the way from morphology to phonology: German linking elements and the role of the phonological word. *Morphology* 18(1), 1–25.

Nunes, Jairo & Juan Uriagereka. 2000. Cyclicity and extraction domains. *Syntax* 3(1), 20–43.

van Oostendorp, Marc. 2005. Obligatory contour principle. Lecture notes for Phonological Theory II, Leiden University.

Orgun, Cemil. 1996. Sign-based morphology and phonology with special attention to Optimality Theory. University of California, Berkeley dissertation.

Oseki, Yohei. 2015. Eliminating pair-merge. In Ulrike Steindl, Thomas Borer, Huilin Fang, Alfredo García Pardo, Peter Guekguezian, Brian Hsu, Charlie O'Hara & Iris Chuoying Ouyang (eds.), *Proceedings of the 32nd West Coast Conference on Formal Linguistics*, 303–312. Somerville MA: Cascadilla Proceedings Project.

Pak, Jin-Hee. 2014. Wàiguórén hànyǔ qīngdòngcí xídé qíngkuàng kǎochá 外国人汉语轻动词习得情况考察 [An investigation of foreigners' acquisition of Chinese light verbs]. Fudan University masters thesis.

Panagiotidis, Phoevos. 2015. *Categorial features*. Cambridge: Cambridge University Press.

Partee, Barbara. 2004. *Compositionality in formal semantics: Selected papers of Barbara Partee*. Oxford: Blackwell.

Partee, Barbara, Alice ter Meulen & Robert Wall. 1993. *Mathematical methods in linguistics* (softcover reprint edn.). Dortrecht: Kluwer.

Paul, Waltraud. 2014. Why particles are not particular: Sentence-final particles in Chinese as heads of a split CP. *Studia Linguistica* 68(1), 77–115.

Paul, Waltraud. 2015. *New perspectives on Chinese syntax*. Berlin: De Gruyter Mouton.

Pesetsky, David & Esther Torrego. 2007. The syntax of valuation and the interpretability of features. In Simin Karimi, Vida Samiian & Wendy K. Wilkins (eds.), *Phrasal and clausal architecture: Syntactic derivation and interpretation*, 262–294. Amsterdam: John Benjamins.

Phillips, Steven & William Wilson. 2014. A category theory explanation for systematicity: Universal constructions. In Paco Calvo & John Symons (eds.), *The architecture of cognition: Rethinking Fodor and Pylyshyn's systematicity challenge*, 227–249. Cambridge, MA: MIT Press.

Phillips, Steven & William Wilson. 2016. Systematicity and a categorical theory of cognitive architecture: Universal construction in context. *Frontiers in Psychology* 7, 1139.1–14.

Piedeleu, Robin, Dimitri Kartsaklis, Bob Coecke & Mehrnoosh Sadrzadeh. 2015. Open system categorical quantum semantics in natural language processing. In Lawrence Moss & Pawel Sobocinski (eds.), *6th Conference on Algebra and Coalgebra in Computer Science*, 270–289. Wadern: Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

Pierce, Benjamin. 1991. *Basic category theory for computer scientists*. Cambridge MA: MIT press.

Pitts, Andrew. 2001. Categorical logic. In Samson Abramsky, Dov Gabbay & Tom Maibaum (eds.), *Logic and algebraic methods*. Handbook of Logic in Computer Science, vol. 5, 39–128. Oxford: Oxford University Press.

Pitts, Andrew. 2018. Category theory. Course at University of Cambridge, Oct–Dec.

Plag, Ingo. 2006. Morphology in pidgins and creoles. In Keith Brown (ed.), *Encyclopedia of language and linguistics* (2nd edn.), 304–308. Oxford: Elsevier.

Pollard, Carl. 2004. Higher-order categorial grammar. In Michael Moortgat (ed.), *International conference on categorial grammars*, 340–361.

Pollard, Carl & Ivan Sag. 1987. *Information-based syntax and semantics*, vol. 1. Stanford: Center for the Study of Language and Information.

Pollard, Carl & Ivan Sag. 1994. *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.

Pollock, Jean-Yves. 1989. Verb movement, universal grammar, and the structure of IP. *Linguistic inquiry* 20(3), 365–424.

Porter, Harry. 1988. A logic-based grammar formalism incorporating feature-structures and inheritance. Oregon Graduate Institute of Science and Technology dissertation.

Preller, Anne. 2014. Natural language semantics in biproduct dagger categories. *Journal of Applied Logic* 12(1), 88–108.

Pylkkänen, Liina. 2008. *Introducing arguments*. Cambridge MA: MIT Press.

Ramchand, Gillian. 2008. *Verb meaning and the lexicon: A first phase syntax*. Cambridge: Cambridge University Press.

Ramchand, Gillian. 2018. *Situations and syntactic structures: Rethinking auxiliaries and order in English*. Cambridge MA: MIT Press.

Ramchand, Gillian & Peter Svenonius. 2014. Deriving the functional hierarchy. *Language Sciences* 46, 152–174.

Refsing, Kirsten. 1986. *The Ainu language*. Aarhus: Aarhus University Press.

Richards, Norvin. 2001. *Movement in language: Interactions and architectures*. Oxford: Oxford University Press.

Richards, Norvin. 2010. *Uttering trees*. Cambridge MA: MIT Press.

Ritter, Elizabeth & Martina Wiltschko. 2019. Nominal speech act structure: Evidence from the structural deficiency of impersonal pronouns. *Canadian Journal of Linguistics* 1–21. doi:10.1017/cnj.2019.10. First view.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of grammar: A handbook of generative syntax*, 281–337. Dortrecht: Kluwer.

Rizzi, Luigi & Guglielmo Cinque. 2016. Functional categories and syntactic theory. *Annual Review of Linguistics* 2, 139–163.

Rizzi, Luigi & Ian Roberts. 1989. Complex inversion in French. *Probus* 1(1), 1–30.

Roberts, Ian. 2007. *Diachronic syntax*. Oxford: Oxford University Press.

Roberts, Ian. 2010. *Agreement and head movement: Clitics, incorporation, and defective goals*. Cambridge MA: MIT Press.

Roberts, Ian. 2011. Head movement and the minimalist program. In Cedric Boeckx (ed.), *The Oxford handbook of linguistic Minimalism*, 195–219. Oxford: Oxford University Press.

Roberts, Ian. 2017. Introduction. In Ian Roberts (ed.), *The Oxford handbook of universal grammar*, 1–34. Oxford: Oxford University Press.

Roberts, Ian. 2019. *Parameter hierarchies and universal grammar*. Oxford: Oxford University Press.

Roberts, Ian & Anna Roussou. 2003. *Syntactic change: A minimalist approach to grammaticalization*. Cambridge University Press.

Roy, Isabelle, Bridget Copley & Lorraine McCune. 2016. The development of Merge and recursion in the transition from single words to sentences: Tracking the development of syntactic complexity in early child language. Manuscript, University Paris8/CNRS and Rutgers University.

Roy, Isabelle, Bridget Copley & Lorraine McCune. 2017. The development of Merge and recursion in early child language: On the functional/lexical divide. In *Workshop on the emergence of features*, Tromsø.

Rubin, Edward. 2003. Determining pair-merge. *Linguistic Inquiry* 34(4), 660–668.

Russell, Bertrand. 1903. *The principles of mathematics*. Cambridge: Cambridge University Press.

Russell, Bertrand & Alfred North Whitehead. 1997. *Principia mathematica to* 56*. Cambridge: Cambridge University Press.

Rydeheard, David & Rod Burstall. 1988. *Computational category theory*. New York: Prentice Hall.

Sagey, Elizabeth. 1986. The representation of features and relations in non-linear phonology. Massachusetts Institute of Technology dissertation.

Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt, Brace.

Schifano, Norma. 2018. *Verb movement in Romance: A comparative study*. Oxford: Oxford University Press.

Schröder, Bernd. 2016. *Ordered sets: An introduction with connections from combinatorics to topology* (2nd edn.). Cham: Birkhäuser.

Shao, Arick. 2012. Set-theoretic paradoxes. Math Mentor Program at University of Toronto, note.

Shi, Yuzhi. 2002. *The establishment of Modern Chinese grammar: The formation of the resultative construction and its effects* (English edn.). Amsterdam: John Benjamins.

Shi, Yuzhi. 2003. *The establishment of Modern Chinese grammar: The formation of the resultative construction and its effects* 现代汉语语法系统的建立——动补结构的产生及其影响 (Chinese edn.). Beijing: Beijing Language and Culture University Press.

Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.

Shieber, Stuart Merrill, Hans Uszkoreit, Fernando Pereira, Jane Robinson & Mabry Tyson. 1983. The formalism and implementation of PATR-II. In Barbara Grosz & Mark Stickel (eds.), *Research on interactive acquisition and use of knowledge (final report of SRI project 1894)*, 39–79. California: SRI International.

Shlonsky, Ur. 2010. The cartographic enterprise in syntax. *Language and linguistics compass* 4(6), 417–429.

Siddiqi, Daniel. 2010. Distributed morphology. *Language and Linguistics Compass* 4(7), 524–542.

Simmons, Harold. 2011. *An introduction to category theory*. Cambridge: Cambridge University Press.

Simpson, Andrew. 2014. Sentence-final particles. In C.-T. James Huang, Y.-H. Audrey Li & Andrew Simpson (eds.), *The handbook of Chinese linguistics*, 156–179. Malden: John Wiley & Sons.

Smith, Carlota. 1991. *The parameter of aspect* (2nd edn.). Kluwer.

Smith, J. David, Alexandria Zakrzewski, Jennifer Johnson, Jeanette Valleau & Barbara Church. 2016. Categorization: The view from animal cognition. *Behavioral Sciences* 6(2), 12.1–24.

Smith, Peter. 2015. Feature mismatches: Consequences for syntax, morphology and semantics. University of Conneticut dissertation.

Smith, Peter. 2018. Category theory: A gentle introduction. Manuscript (2018-01-29 edn.), University of Cambridge.

Song, Chenchen. 2015. Structural variation in Chinese compound verbs: A comparative study of Standard Mandarin and Dongying dialect. University of Cambridge masters thesis.

Song, Chenchen. 2016a. Blocked particle verb movement in German and Hungarian: A unified analysis. Paper presented at Linguistics Association of Great Britain Annual Meeting, Sep 6–9, York.

Song, Chenchen. 2016b. A minimalist study of complex verb formation: Cross-linguistic patterns and variation. PhD First-Year Report, University of Cambridge.

Song, Chenchen. 2016c. Rivisiting Chinese potential constructions: A cross-dialectal study. Paper presented at Newcastle Postgraduate Conference, Mar 18, Newcastle.

Song, Chenchen. 2017a. Categorizing verb-internal modifiers. Manuscript, University of Cambridge.

Song, Chenchen. 2017b. Emergent [V] flavors and minimized flexibility of lexical categories. Paper presented at Cambridge Comparative Syntax 6, May 4–6, Cambridge.

Song, Chenchen. 2017c. Prosodically-driven morphosyntactic change? Revisiting the history of Chinese disyllabic words. In Chenchen Song & James Baker (eds.), *Cambridge occasional papers in linguistics*, vol. 10, 152–196. Theoretical and Applied Linguistics, University of Cambridge.

Song, Chenchen. 2018a. Root-supported categories and their morphosyntactic characteristics. Talk at Cambridge SyntaxLab, May 1, Cambridge.

Song, Chenchen. 2018b. Severing telicity from result. *Journal of East Asian Linguistics* 27(3), 275–307.

Speas, Peggy & Carol Tenny. 2003. Configurational properties of point of view roles. In Anna Di Sciullo (ed.), *Syntax and semantics*. Asymmetry in Grammar, vol. 1, 315–345. Amsterdam: John Benjamins.

Spivak, David. 2014. *Category theory for the sciences*. Cambridge MA: MIT Press.

Stabler, Edward. 1997. Derivational minimalism. In Christian Retoré (ed.), *Logical aspects of computational linguistics*, 68–95. Berlin: Springer.

Starke, Michal. 2009. Nanosyntax: A short primer to a new approach to language. *Nordlyd* 36(1), 1–6.

Starke, Michal. 2011. Towards an elegant solution to language variation: Variation reduces to the size of lexically stored trees. Manuscript, Tromsø University.

Sun, Xixin. 1999. *Jìndài hànyǔ yǔqìcí* 近代汉语语气词 [*Interjections in Modern Chinese*]. Beijing: Language and Culture Press.

Suñer, Margarita. 1994. V-movement and the licensing of argumental wh-phrases in Spanish. *Natural Language & Linguistic Theory* 12(2), 335–372.

Svenonius, Peter. 2014. Generalized applicatives: Reassessing the lexical-functional divide. *Theoretical Linguistics* 40(3/4), 439–446.

Svenonius, Peter. 2019. Syntactic features. In *Oxford research encyclopedia of linguistics*, 1–31. Oxford: Oxford University Press.

Sybesma, Rint & Boya Li. 2007. The dissection and structural mapping of Cantonese sentence final particles. *Lingua* 117(10), 1739–1783.

Talmy, Leonard. 1972. Semantic structures in English and Atsugewi. University of California, Berkeley dissertation.

Tan, Liya & Haihong Chen. 2017. Wǎngluò liúxìngyǔ dī de duōgōngnéng yòngfǎ qiǎnxī 网络流行语"滴"的多功能用法浅析 [The multifunctional usage of online catch words *Di*]. *Journal of Southwest Petroleum University (Social Sciences Edition)* 19(6), 95–100.

Tang, Sze-Wing. 2015. A generalized syntactic schema for utterance particles in Chinese. *Lingua Sinica* 1(1), 3.

Tang, Ting-Chi. 2015. *Duìbǐ fēnxī yánjiū rùmén* 對比分析研究入門（上） [*An introduction to comparative analytical research (Part I)*]. Taipei: Zhiliang Press.

Tham, Shiao Wei. 2015. Resultative verb compounds in Mandarin. In *The Oxford handbook of Chinese linguistics*, 306–322. Oxford: Oxford University Press.

Tomalin, Marcus. 2006. *Linguistics and the formal sciences: The origins of generative grammar*. Cambridge: Cambridge University Press.

Traugott, Elizabeth & Ekkehard König. 1991. The semantics-pragmatics of grammaticalization revisited. In Elizabeth Traugott & Bernd Heine (eds.), *Approaches to grammaticalization*, vol. 2, 189–218. Amsterdam: John Benjamins.

Travis, Lisa. 1984. Parameters and effects of word order variation. Massachusetts Institute of Technology dissertation.

Tsoulas, George. 2017. Syntactic features. In Ian Roberts (ed.), *The Oxford handbook of universal grammar*, 183–218. Oxford: Oxford University Press.

Tsujimura, Natsuko. 2014. *An introduction to Japanese linguistics* (3rd edn.). Malden: John Wiley & Sons.

Usuki, Takeshi. 2007. Manner conflation in v/√root partition. *Journal of the Graduate School of Fukuoka University* 1–30.

Vikner, Sten. 2005. Immobile complex verbs in Germanic. *The Journal of Comparative Germanic Linguistics* 8(1), 85–117.

Wang, Chyan-An Arthur. 2014. Mandarin pseudo-imposters. In Chris Collins (ed.), *Cross-linguistic studies of imposters and pronominal agreement*, 173–195. Oxford: Oxford University Press.

Wang, Hao. 1952. Logic of many-sorted theories. *The Journal of Symbolic Logic* 17(2), 105–116.

Wang, Li. 1989. *Hànyǔ yǔfǎ shǐ* 漢語語法史 [*A history of Chinese grammar*]. Beijing: Commercial Press.

Wang, Li. 2004. *Hànyǔ shǐgǎo* 漢語史稿 [*A manuscript on the history of the Chinese language*] (2nd edn.). Beijing: Zhonghua Book Company.

Weatherall, James. 2016. Are Newtonian gravitation and geometrized Newtonian gravitation theoretically equivalent? *Erkenntnis* 81(5), 1073–1091.

Williams, Alexander. 2014. Causal VVs in Mandarin. In C.-T. James Huang, Y.-H. Audrey Li & Andrew Simpson (eds.), *The handbook of Chinese linguistics*, 311–341. Malden: John Wiley & Sons.

Wiltschko, Martina. 2014. *The universal structure of categories: Towards a formal typology*. Cambridge: Cambridge University Press.

Wiltschko, Martina. 2016. What particle syntax can teach us about the particularities of syntax. Paper presented at workshop Linguistic Variation in the Interaction between Internal and External Syntax, Feb 8–9, Utrecht.

Wiltschko, Martina. 2017. Ergative constellations in the structure of speech acts. In Jessica Coon, Diane Massam & Lisa Travis (eds.), *The Oxford handbook of ergativity*, 419–442. Oxford: Oxford University Press.

Wiltschko, Martina. 2018. Response particles beyond answering. In Laura Bailey & Michelle Sheehan (eds.), *Word order and syntactic structure*. Order and structure in syntax, vol. 1, 241–279. Berlin: Language Science Press.

Wiltschko, Martina & Johannes Heim. 2016. The syntax of confirmationals. In Gunther Kaltenböck, Evelien Keizer & Arne Lohmann (eds.), *Outside the clause: Form and function of extra-clausal constituents*, 305–340. Amsterdam: John Benjamins.

Xia, Xianpei. 2018. Nǚzǐ zìchēng chénqiè de yuánliú 女子自称"臣妾"的源流 [On the origin of the female self-appellation *chénqiè*]. *Journal of Changsha University of Science and Technology (Social Science)* 33(5), 118–122.

Xiao, Hong. 2006. Hànyǔ duōdòngtóngbīn jùshì de fāzhǎn 汉语多动同宾句式的发展 [The development of the multiple-verb-common-object structure in Chinese]. *Studies in Language and Linguistics* 26(4), 111–114.

Xue, Bei. 2009. Bìngliè liáncí yǔ jí chāyì tànyuán 并列连词"与""及"差异探源 [On the origin of the differences between the coordinative conjunctions *yǔ* and *jí*]. *Journal of Suzhou College of Education* 26(2), 51–54.

Ye, Jianwei. 2007. Ègǎo xiànxiàng shànbiàn dào fěnsī wénhuà de chuánbōxué yánjiū—yǐ línwénhuà wéilì 恶搞现象嬗变到粉丝文化的传播学研究——以淋文化为例 [A communication-scientific study of the evolution from kuso phenomena to fan cultures: The case of Lin culture]. Zhejiang University masters thesis.

Yin, Hailiang. 2013. Shāndōng lèlíng fāngyán zhùcí le 山东乐陵方言助词"了" [On the particle *le* in Leling (Shandong) Mandarin]. *Journal of Shijiazhuang University* 15(4), 81–86.

Zeller, Jochen. 2002. *Particle verbs and local domains*. Amsterdam: John Benjamins.

Zhang, Niina Ning. 2013. *Classifier structures in Mandarin Chinese*. Berlin: De Gruyter Mouton.

Zhang, Yujin. 2001. *Jiǎgǔwén yǔfǎxué* 甲骨文语法学 [*The syntax of oracle bone scripts*]. Xuelin Publishing House.

Zhu, Dexi. 1982. *Yǔfǎ jiǎngyì* 语法讲义 [*Lectures notes on Chinese grammar*]. Beijing: Commercial Press.

Zuo, Linxia. 2004. Gǔhànyǔ zhōng de gòngbīn jiégòu 古汉语中的共宾结构 [The shared accusative structure in Ancient Chinese]. *Journal of Xiaogan University* 24(4), 43–46.

Zwart, Jan-Wouter. 2006. Local agreement. In Cedric Boeckx (ed.), *Agreement systems*, 317–339. Amsterdam: John Benjamins.

Zwart, Jan-Wouter. 2007. Layered derivations. Advanced Core Training in Linguistics lecture, Mar 9, London.

Zwart, Jan-Wouter. 2009. Prospects for top-down derivation. *Catalan Journal of linguistics* 8, 161–187.

Zwart, Jan-Wouter. 2011. Recursion in language: A layered-derivation approach. *Biolinguistics* 5(1-2), 043–056.