Ways of Structure
Building

Myriam Uribe-Etxebarria,
Vidal Valmala
Ways of Structure Building
32 Negative Indefinites
by Doris Penka

33 Events, Phrases, and Questions
by Robert Truswell

34 Dissolving Binding Theory
by Johan Rooryck and Guido Vanden Wyngaerd

35 The Logic of Pronominal Resumption
by Ash Asudeh

36 Modals and Conditionals
by Angelika Kratzer

37 The Theta System
Argument Structure at the Interface
edited by Martin Everaert, Marijana Marelj, and Tal Siloni

38 Sluicing in Cross-Linguistic Perspective
edited by Jason Merchant and Andrew Simpson

39 Telicity, Change, and State
A Cross-Categorial View of Event Structure
edited by Violeta Demonte and Louise McNally

40 Ways of Structure Building
edited by Myriam Uribe-Etxebarria and Vidal Valmala

For a complete list of titles published and in preparation for the series, see pp 424–425.
This page intentionally left blank
Ways of Structure Building

Edited by
MYRIAM URIBE-ETXEBARRIA
AND VIDAL VALMALA

OXFORD UNIVERSITY PRESS
Contents

General Preface xi
Acknowledgements xii
Notes on Contributors xiii
Abbreviations xvii

1 Overview 1–16
   Myriam Uribe-Etxebarria and Vidal Valmala

Part I. Merge and beyond

2 Constituent Structure Sets II 19–43
   Dirk Bury and Hiroyuki Uchida
   2.1 Introduction 19
   2.2 Constituent structure sets (CSS) 24
   2.3 Representational collapsibility 29
   2.4 German V2 in CSS (Bury 2003, 2005) 30
   2.5 Wh-extraction 33
   2.6 PF linearization 38
   2.7 CSS in comparison 39
   2.8 Conclusion 43

3 A Parallel Merge Solution to the Merchant/Johnson Paradox 44–65
   Barbara Citko
   3.1 Parallel Merge: a new way of structure building 44
   3.2 Voice mismatches in VP ellipsis, pseudogapping, and
gapping constructions 46
   3.3 Small conjunct account 50
   3.4 A Parallel Merge approach to gapping 62
   3.5 Conclusion 65

4 Evidence for Multidominance in Spanish Agentive Nominalizations 66–92
   Antonio Fábregas
   4.1 The empirical problem and its relevance 66
   4.2 Empirical data: affixes that cancel argument structure 68
   4.3 How a multidominance account can explain these cases 74
   4.4 Why a multidominance account is necessary 83
   4.5 Summary and conclusions 92
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Clitic Placement and Multidominance</td>
<td>93–113</td>
</tr>
<tr>
<td>Martina Gračanin-Yuksek</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>93</td>
</tr>
<tr>
<td>5.2</td>
<td>Clitic je in non-MD structures</td>
<td>94</td>
</tr>
<tr>
<td>5.3</td>
<td>Q&amp;Qs in Croatian</td>
<td>97</td>
</tr>
<tr>
<td>5.4</td>
<td>German <em>Subjektlücke in finiten Sätzen</em></td>
<td>105</td>
</tr>
<tr>
<td>5.5</td>
<td>Constraint on Sharing</td>
<td>107</td>
</tr>
<tr>
<td>5.6</td>
<td>Conclusion</td>
<td>111</td>
</tr>
<tr>
<td>6</td>
<td>Sideward Movement: Triggers, Timing, and Outputs</td>
<td>114–142</td>
</tr>
<tr>
<td>Jairo Nunes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.1</td>
<td>Introduction</td>
<td>114</td>
</tr>
<tr>
<td>6.2</td>
<td>Some examples of sideward movement</td>
<td>119</td>
</tr>
<tr>
<td>6.3</td>
<td>Preventing overgeneration</td>
<td>121</td>
</tr>
<tr>
<td>6.4</td>
<td>Sideward movement and noncanonical phonetic realization of copies</td>
<td>128</td>
</tr>
<tr>
<td>6.5</td>
<td>Wh-movement and adjunct control</td>
<td>134</td>
</tr>
<tr>
<td>6.6</td>
<td>Morphological sideward movement</td>
<td>138</td>
</tr>
<tr>
<td>6.7</td>
<td>Concluding remarks</td>
<td>142</td>
</tr>
<tr>
<td>7</td>
<td>Unconventional Mergers</td>
<td>143–166</td>
</tr>
<tr>
<td>Mark de Vries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.1</td>
<td>Introduction</td>
<td>143</td>
</tr>
<tr>
<td>7.2</td>
<td>Internal and external remerge</td>
<td>145</td>
</tr>
<tr>
<td>7.3</td>
<td>Parenthetical Merge</td>
<td>153</td>
</tr>
<tr>
<td>7.4</td>
<td>Combining Parenthetical Merge and <em>external remerge</em> in amalgams</td>
<td>159</td>
</tr>
<tr>
<td>7.5</td>
<td>Conclusions</td>
<td>164</td>
</tr>
<tr>
<td>Part II</td>
<td>Triggers and constraints</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>“Lasnik-Effects” and String-Vacuous ATB Movement</td>
<td>169–205</td>
</tr>
<tr>
<td>Jun Abe and Norbert Hornstein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.1</td>
<td>Introduction</td>
<td>169</td>
</tr>
<tr>
<td>8.2</td>
<td>The ATB movement analysis of Right-Node-Raising: Sabbagh’s (2007) approach</td>
<td>170</td>
</tr>
<tr>
<td>8.3</td>
<td>Proposal</td>
<td>181</td>
</tr>
<tr>
<td>8.4</td>
<td>Against the Right/Left Edge Restriction</td>
<td>194</td>
</tr>
<tr>
<td>8.5</td>
<td>Speculations on some further consequences</td>
<td>200</td>
</tr>
<tr>
<td>8.6</td>
<td>Conclusion</td>
<td>205</td>
</tr>
</tbody>
</table>
9 Disharmony, Antisymmetry, and the Final-over-Final Constraint 206–244

Theresa Biberauer and Michelle Sheehan

9.1 Introduction 206
9.2 The Final-over-Final Constraint 208
9.3 FOFC and the LCA 213
9.4 CP extraposition 216
9.5 Accounting for CP extraposition 221
9.6 Conclusion 243

10 Don’t Feed Your Movements When You Shift Your Objects 245–252

Željko Bošković

10.1 Introduction 245
10.2 Don’t feed your movements 246
10.3 Icelandic object shift 249

11 Structure Building That Can’t Be 253–270

Samuel David Epstein, Hisatsugu Kitahara, and T. Daniel Seely

11.1 Introduction 253
11.2 Valuation-induced Transfer: a problem 257
11.3 Resolving the before/after problem 259
11.4 Deducing cyclic Transfer 261
11.5 Consequences 266

12 Specificity-driven Syntactic Derivation 271–296

Antje Lahne

12.1 Introduction 271
12.2 Theoretical background 272
12.3 Deriving Minimal Link Condition effects 279
12.4 Deriving anti-MLC effects 289
12.5 Outlook 294

13 Structure Building From Below: More on Survive and Covert Movement 297–329

Winfried Lechner

13.1 Introduction 297
13.2 Defining repulsion 300
13.3 Quantifier scope restrictions 303
13.4 TSP and movement 304
13.5 Analyzing scope restrictions with TSP 312
13.6 A note on the density of movement paths 327
13.7 Conclusion 329
## Contents

14 On Transparent Adjuncts in Japanese  
*Yoichi Miyamoto*  
14.1 Introduction  
14.2 Adjunct Condition effects  
14.3 Comparative deletion and floating quantifiers in Japanese  
14.4 The proposal  
14.5 Concluding remarks  

15 Feature-Splitting Internal Merge: The Case of Tough-constructions  
*Miki Obata and Samuel David Epstein*  
15.1 Introduction  
15.2 Framing the issues  
15.3 Phase-head complement feature symmetry  
15.4 Feature-splitting: separation of [Case] and [Phi]  
15.5 Discussion and outstanding issues  

**References**  
**Language Index**  
**Subject Index**
General Preface

The theoretical focus of this series is on the interfaces between subcomponents of the human grammatical system and the closely related area of the interfaces between the different subdisciplines of linguistics. The notion of “interface” has become central in grammatical theory (for instance, in Chomsky’s recent Minimalist Program) and in linguistic practice: work on the interfaces between syntax and semantics, syntax and morphology, phonology and phonetics, etc. has led to a deeper understanding of particular linguistic phenomena and of the architecture of the linguistic component of the mind/brain.

The series covers interfaces between core components of grammar, including syntax/morphology, syntax/semantics, syntax/phonology, syntax/pragmatics, morphology/phonology, phonology/phonetics, phonetics/speech processing, semantics/pragmatics, and intonation/discourse structure, as well as issues in the way that the systems of grammar involving these interface areas are acquired and deployed in use (including language acquisition, language dysfunction, and language processing). It demonstrates, we hope, that proper understandings of particular linguistic phenomena, languages, language groups, or inter-language variations all require reference to interfaces.

The series is open to work by linguists of all theoretical persuasions and schools of thought. A main requirement is that authors should write so as to be understood by colleagues in related subfields of linguistics and by scholars in cognate disciplines.

There has always been a tension within generative grammar between explanations which rely on how the syntax interfaces with external systems such as the lexicon, or spell-out, and those which rely on the nature of the computational system itself. In the current volume, leading researchers demonstrate that closer inspection of how structure is built results in new solutions to problems which were previously thought to be best understood in terms of interface processes. Taking as their topics some of the most challenging phenomena in syntax, the authors propose novel ways of defining the core structure building algorithm; these approaches both cast new light on old problems and sharpen the issue of exactly how to resolve the tension between the computational system and the interpretive components of human language.

David Adger
Hagit Borer
Acknowledgements

We would like to thank the anonymous referees who provided valuable and detailed comments on the chapters in this volume. We are also indebted to Julia Steer, our editor at Oxford University Press, for her guidance and support throughout the publishing process, as well as to Bryan Leferman, who helped us proof reading the manuscript. Alejo Alcaraz and José Ignacio Martín also deserve a special mention for providing us with logistic support while preparing the manuscript.

Finally, we would like to acknowledge the Basque Government and the University of the Basque Country (UPV-EHU) (grants GIC07/144-IT-210-07, HM-2009-1-1, and UFi11/14) and the Spanish Ministry of Science and Innovation (grants FFI2008-04786/FILO and FFI2011-29218/FILO) for their financial support.
Notes on Contributors


Theresa Biberauer is a Senior Research Associate in the Department of Theoretical and Applied Linguistics at the University of Cambridge, and Senior Lecturer Extraordinary in the General Linguistics Department at Stellenbosch University in South Africa. Her research to date has focused on a range of synchronic and diachronic topics, notably word-order variation, subject-related phenomena, negation and doubling phenomena. She is the editor of The Limits of Syntactic Variation (Benjamins, 2008) and, along with Anders Holmberg, Ian Roberts, and Michelle Sheehan, a co-author of Parametric Variation (CUP, 2010).

Željko Bošković is Professor in the Department of Linguistics at the University of Connecticut. His research interests include theoretical and comparative syntax. He is the author of The Syntax of Nonfinite Complementation: An Economy Approach (1997) and On the Nature of the Syntax-Phonology Interface: Cliticization and Related Phenomena (2001).

Dirk Bury is an Honorary Research Fellow at the Department of Linguistics at UCL. His work combines a restrictive theory of constituent structure with a flexible view of clause structure, where empty functional heads are not lexical items, but are licensed derivationally or configurationally. He used the idea that moved material can project to analyze verb movement and free relatives. He has also worked on the structure of verb-second and verb-initial languages, and on adjunction.

Barbara Citko is an Associate Professor of Linguistics at the University of Washington in Seattle. She specializes in syntactic theory, and has worked on the syntax of questions, relative clauses, coordinate structures, and various types of multidominant constructions. Her published work includes one book (Symmetry in Syntax: Merge, Move Labels, published in 2011 by Cambridge University Press), as well as articles in journals such as Linguistic Inquiry, Natural Language and Linguistic Theory, Syntax, and Journal of Slavic Linguistics.
SAMUEL DAVID EPSTEIN is a Professor in the Linguistics Department at the University of Michigan. He is co-author (with E. Groat, R. Kawashima, and H. Kitahara) of *A Derivational Approach to Syntactic Relations*, and (with T. Daniel Seely) of *Derivations in Minimalism*, and co-editor (with N. Hornstein) of *Working Minimalism* (1999). He is co-founder (with S. Flynn) of the journal *Syntax*.

ANTONIO FÁBREGAS is currently Full Professor of Spanish Linguistics at the University of Tromsø, where he is also affiliate to the Center ASTL. He took his PhD in Linguistics at the Ortega y Gasset Research Institute (2005). His work concentrates on the syntactic analysis of morphological processes. He has published on topics such as relational adjectives, nominalizations, Aktionsart, irregularity and allomorphy, compounding, manner of motion verbs, infinitives, gerunds, and prepositions.

MARTINA GRAČANIN-YUKSEK is an Assistant Professor of Linguistics at the Middle East Technical University, Ankara, Turkey. Her research interests fall into two broad categories: investigations of multidominance on the one hand and investigations in Slavic linguistics on the other. Within these, she has focused on the syntax of (multiple coordinated) wh-questions, relative clauses, free relatives, and double-object construction, as well as on the issues in linearization of multidominant structures. She has published work in *Linguistic Inquiry* and *Syntax*, among others.

NORBERT HORNSTEIN teaches Linguistics at the University of Maryland (College Park).

HISATSUGU KITAHARA is Professor at the Institute of Cultural and Linguistic Studies at Keio University. His research area is minimalist syntax, specifically a derivational approach to phrase structure. He is also interested in foundational issues concerning the field of generative grammar. He is the author of *Elementary Operations and Optimal Derivations* (MIT Press, 1997).

ANTJE LAHNE is a lecturer in Linguistics at the University of Konstanz. Her research interests are theoretical and comparative syntax, formal morphology, morphophonology, and the syntax-morphology interface. Her PhD thesis deals with the formal modeling of reflexes of successive-cyclic movement.

WINFRIED LECHNER is Assistant Professor for German linguistics and theoretical linguistics at the Department for German Language and Literature, University of Athens. His main academic interests are located in the areas of syntax and semantics, focusing on the interaction between these two components. In recent and current work, he has been investigating the grammatical principles determining the logical syntax of scope, the cross-linguistic typology of different, reconstruction effects, opacity at the syntax-semantics interface, and the architecture of the grammar.

YOICHI MIYAMOTO is an Associate Professor in Linguistics at the Graduate School of Language and Culture, Osaka University. His research interests relate to comparative
syntax of Japanese and English, along with first and second language acquisition with a special focus on the syntax-semantics interface. His publications are in Journal of Language and Linguistics, Canadian Journal of Linguistics, and Journal of East Asian Linguistics, amongst others.

Jairo Nunes is Professor of Linguistics at the Universidade de São Paulo. His research interests involve syntactic theory and comparative syntax. He is the author of Linearization of Chains and Sideward Movement (MIT Press, 2004), co-author of Understanding Minimalism (Cambridge University Press, 2005) and Control as Movement (Cambridge University Press, 2010), and co-editor of The Copy Theory of Movement (John Benjamins, 2007) and Minimalist Essays on Brazilian Portuguese Syntax (John Benjamins, 2009). He is also co-editor of Probus: International Journal of Latin and Romance Linguistics (Mouton de Gruyter).

Miki Obata completed her PhD thesis at the University of Michigan, Ann Arbor, in 2010. She is an Assistant Professor in Linguistics at Mie University, Japan. Her research interest is theoretical syntax and psycholinguistics. She has published work in Syntax (co-authored with Samuel David Epstein), NELS proceedings, and other edited volumes.

T. Daniel Seely is Professor of Linguistics at Eastern Michigan University. His work in syntax has appeared in Linguistic Inquiry and Syntax. He is organizer and editor of “Geometric and Thematic Structure in Binding” (1996), the first LINGUIST List online conference, and he is co-editor (with Samuel David Epstein) of Derivation and Explanation in the Minimalist Program (2002) and co-author (with Samuel David Epstein) of Derivations in Minimalism (2006).

Michelle Sheehan is a Research Associate in the Department of Theoretical and Applied Linguistics at the University of Cambridge. She has worked on a variety of topics related to syntactic theory, cross-linguistic variation, linearization, and the syntax-semantics interface. She is a co-author of the Cambridge University Press volume Parametric Variation and has authored and co-authored papers in a number of journals and edited collections on topics such as null subjects, word order, island effects, and extraposition.

Hiroyuki Uchida is a postdoctoral research fellow at the Human-Level Intelligence Laboratory at Rensselaer Polytechnic Institute. His research interests are logic, algebraic structures, proof theories, and their application in natural language structures. He is especially interested in the logical and algebraic properties of the syntactic structures and operations in natural language data. He has published work in Lecture Notes in Artificial Intelligence, Linguistic Analysis, as well as in edited volumes.

Myriam Uribe-Etxebarria is Associate Professor at the Department of Linguistics and Basque Studies, University of the Basque Country. Her research interest focuses on
the syntax-semantics interface, with particular attention to the grammar of temporal relations, wh-questions, and focus. She is the author of *Interface Licensing Conditions on Negative Polarity Items: A Theory of Polarity and Tense Interactions* (UPV-EHU, 2004), and co-editor of *Theoretical Issues at the Morphology-Syntax Interface* (ASJU, UPV-EHU, 1997), *Romance Linguistics: Theoretical Perspectives* (John Benjamins, 1998), and *Noun Phrases and Nominalization in Basque: Syntax and Semantics* (John Benjamins, in press). She has published her work in various journals and edited volumes.

**Vidal Valmala** is an Associate Professor of English Language and Linguistics at the University of the Basque Country. Among his research topics are ellipsis and the syntactic realization of topic and focus. He has published articles on floating quantification, fragments, VP ellipsis, and comparative correlatives.

**Mark de Vries** is an Associate Professor of Linguistics at the University of Groningen. His research interests include the theoretical foundations of syntax as well as comparative syntax. Construction types he has worked on are relative clauses, appositions, coordination, and dislocation, among other things. Currently, he is leading a research project on parenthesis and ellipsis. He has published articles in *Linguistic Inquiry, Biolinguistics*, and various other journals and edited volumes.
**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>accusative</td>
</tr>
<tr>
<td>ACE</td>
<td>Adjunct Condition effect</td>
</tr>
<tr>
<td>Agr</td>
<td>agreement</td>
</tr>
<tr>
<td>AgrOP</td>
<td>agreement object phrase</td>
</tr>
<tr>
<td>ARC</td>
<td>appositive relative clause</td>
</tr>
<tr>
<td>Asp</td>
<td>aspect</td>
</tr>
<tr>
<td>ATB</td>
<td>Across-the-Board</td>
</tr>
<tr>
<td>Aux</td>
<td>auxiliary</td>
</tr>
<tr>
<td>BHR</td>
<td>Biberauer, Holmberg, and Roberts</td>
</tr>
<tr>
<td>BP</td>
<td>Brazilian Portuguese</td>
</tr>
<tr>
<td>BSN</td>
<td>Biberauer, Sheehan, and Newton</td>
</tr>
<tr>
<td>C</td>
<td>complementer</td>
</tr>
<tr>
<td>CED</td>
<td>Condition on Extraction Domain</td>
</tr>
<tr>
<td>CH</td>
<td>chain</td>
</tr>
<tr>
<td>Ch</td>
<td>Computational system of human language</td>
</tr>
<tr>
<td>CI</td>
<td>Conceptual-Intentional</td>
</tr>
<tr>
<td>CL</td>
<td>classifier</td>
</tr>
<tr>
<td>COSH</td>
<td>Constraint On Sharing</td>
</tr>
<tr>
<td>CP</td>
<td>complementer phrase</td>
</tr>
<tr>
<td>CRP</td>
<td>Case-Resistance Principle</td>
</tr>
<tr>
<td>CS</td>
<td>complement stranding</td>
</tr>
<tr>
<td>CSS</td>
<td>constituent structure set</td>
</tr>
<tr>
<td>D</td>
<td>determiner</td>
</tr>
<tr>
<td>DAT</td>
<td>dative</td>
</tr>
<tr>
<td>DO</td>
<td>direct object</td>
</tr>
<tr>
<td>DP</td>
<td>determiner phrase</td>
</tr>
<tr>
<td>EC</td>
<td>Extension Condition</td>
</tr>
<tr>
<td>ECP</td>
<td>Empty Category Principle</td>
</tr>
<tr>
<td>EF</td>
<td>edge feature</td>
</tr>
<tr>
<td>EKS</td>
<td>Epstein, Kitahara, and Seely</td>
</tr>
<tr>
<td>EM</td>
<td>External Merge</td>
</tr>
<tr>
<td>EP</td>
<td>extended projection</td>
</tr>
<tr>
<td>EPP</td>
<td>Extended Projection Principle</td>
</tr>
</tbody>
</table>
Abbreviations

FOFC Final-over-Final Constraint
FQ floating quantifier
FUT future
GB Government and Binding
GEN genitive
Ger gerund
GerP gerund phrase
GSP Generalized Specificity Principle
I inflection
IC immediate containment (Chapter 2)
IC interrupting clause (Chapter 7)
ID immediate dominance
IM Internal Merge
InitP initiator phrase
IO indirect object
LAC Label Accessibility Condition
LCA Linear Correspondence Axiom
LCF Law of Conservation of Features
LCR Law of Conservation of Relations
LDA long distance agreement
LF logical form
LI lexical item
LNR Left-Node-Raising
MD multidominance / multidominant
MLC Minimal Link Condition
MME Maximize Matching Effects
NEG negation
NOM nominative
NP noun phrase
NS narrow syntax
NTC No-Tampering Condition
O object
OV object-verb order
ParP parenthetical phrase
PD proper dominance
PF phonetic form
PIC Phase Impenetrability Condition
PL plural
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pol</td>
<td>polarity (head)</td>
</tr>
<tr>
<td>PP</td>
<td>prepositional phrase</td>
</tr>
<tr>
<td>ProcP</td>
<td>process phrase</td>
</tr>
<tr>
<td>PS</td>
<td>phrase structure</td>
</tr>
<tr>
<td>P-stranding</td>
<td>preposition stranding</td>
</tr>
<tr>
<td>Q-float</td>
<td>quantifier float</td>
</tr>
<tr>
<td>QP</td>
<td>quantifier phrase</td>
</tr>
<tr>
<td>QR</td>
<td>quantifier raising</td>
</tr>
<tr>
<td>QUOT</td>
<td>quotative</td>
</tr>
<tr>
<td>RC</td>
<td>reflexive containment</td>
</tr>
<tr>
<td>RD</td>
<td>reflexive dominance</td>
</tr>
<tr>
<td>RM</td>
<td>relativized minimality</td>
</tr>
<tr>
<td>RNR</td>
<td>Right-Node-Raising</td>
</tr>
<tr>
<td>RRC</td>
<td>Right Roof Constraint</td>
</tr>
<tr>
<td>S</td>
<td>subject</td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>SLF</td>
<td>Subject Lacking in Finite clauses</td>
</tr>
<tr>
<td>SM</td>
<td>sensorimotor</td>
</tr>
<tr>
<td>SMS</td>
<td>sensory-motor system</td>
</tr>
<tr>
<td>SMT</td>
<td>strong minimalist thesis</td>
</tr>
<tr>
<td>SO</td>
<td>syntactic object</td>
</tr>
<tr>
<td>Spec</td>
<td>specifier</td>
</tr>
<tr>
<td>T</td>
<td>tense</td>
</tr>
<tr>
<td>TP</td>
<td>tense phrase</td>
</tr>
<tr>
<td>TSP</td>
<td>The Survive Principle</td>
</tr>
<tr>
<td>UTAH</td>
<td>Universal Theta-role Assignment Hypothesis</td>
</tr>
<tr>
<td>V</td>
<td>verb</td>
</tr>
<tr>
<td>VO</td>
<td>verb-object order</td>
</tr>
<tr>
<td>VP</td>
<td>verb phrase</td>
</tr>
<tr>
<td>VPE</td>
<td>verb phrase ellipsis</td>
</tr>
</tbody>
</table>
This page intentionally left blank
Overview

MYRIAM URIBE-ETXEBARRIA AND VIDAL VALMALA

This volume offers the reader a wide and updated view of some of the most important approaches to the following key issues in contemporary syntactic theory: What are the operations available for (syntactic) structure building in natural languages? What are the triggers behind those structure building operations? Which constraints operate on the structure building operations available? These questions have been at the heart of the debate in linguistic theory since the beginning of the generative enterprise.

As witnessed by the large number of (in many cases conflicting) proposals that have emerged in recent years when trying to provide answers to these questions, there is no unified view in the field. This is why we considered that a volume gathering some of the most representative and influential alternatives currently available would be both useful and necessary, and hence the title of this volume: *Ways of Structure Building*. We consider that this type of project is also especially relevant in the present context of deep and rapid changes within the generative approaches to the study of the language faculty, and we hope that it will be helpful for those trying to gain a deeper understanding of the driving forces which have motivated those changes. The fact that the chapters in this volume introduce new data and paradigms into the discussion makes it interesting for all researchers, independently of their adherence to a particular theoretical approach.

All the contributions in this book aim at providing new answers to the three fundamental interrelated questions raised above on the basis of a detailed discussion of a wide range of phenomena (gapping, Right-Node-Raising, comparative deletion, Across-the-Board movement, *tough*-constructions, nominalizations, scope interactions, wh-movement, A-movement, Case and agreement relations, among others), using evidence from a rich variety of languages (for instance Brazilian Portuguese, Bulgarian, Croatian, English, German, Icelandic, Japanese, and Spanish).

A common feature of many of the proposals presented in this volume is that they clearly illustrate the shift in the locus of the explanation of linguistic phenomena that
characterizes contemporary linguistic theory. A shift, in many cases, from a model which relied on properties of systems external to narrow syntax (such as the lexicon or the PF component) to one which relies on properties of the structure building mechanisms available.

Government and Binding, for instance, appealed to both empty categories in the lexicon (PRO, pro, or null operators) and PF ellipsis/deletion processes when trying to account for a number of elliptical constructions like Right-Node-Raising, control infinitives, parasitic gaps, etc. The consequence was an enrichment of the mental lexicon with "null categories" and an enrichment of the PF component with ellipsis/deletion operations of different types. In recent years, however, there has been an attempt to provide an account of many of those elliptical constructions on the basis of alternative approaches to structure building like Parallel Merge and sideward movement which dispense with many of the null categories and ellipsis/deletion operations of Government and Binding. Some illustrations of this change of locus follow.

Tough-constructions and parasitic gaps were typically analyzed in Government and Binding as consisting of movement of a null operator to Spec/CP of the embedded clause. As a result of this shift, some authors now analyze the former in terms of (im)proper movement (Obata and Epstein, this volume) and the latter have been proposed to involve sideward movement (Nunes 1995, 2001, 2004, this volume). Also, in situ approaches to Right-Node-Raising constructions, which crucially relied on ellipsis/PF deletion within Government and Binding, have an alternative Parallel Merge solution in more recent proposals (Wilder 1999, Bachrach and Katzir 2009, among others). Adjunct control infinitives which involved a PRO subject in Government and Binding have been reinterpreted as instances of sideward movement (Hornstein 1999, 2001). Also gapping, for which ellipsis/PF deletion was proposed in Government and Binding, has been argued to involve Across-the-Board movement (Johnson 2009) or Parallel Merge (Citko, this volume) in recent studies.

The chapters of this book will thus offer the reader a picture of many of the tools which, on the basis of recent research on structure building, have allowed scholars to find new solutions to old problems of syntactic theory. In what follows we briefly introduce the most prominent properties of the proposals put forth in this book, grouping the different chapters so as to show the reader how they compete or complement each other.

The emergence of Kayne’s (1994) Linear Correspondence Axiom—i.e. asymmetric c-command maps onto linear precedence—and the Minimalist Program (Chomsky 1995b and subsequent work) paved the way for the development of recent proposals which introduce two "new" structure building operations: Parallel Merge by which an element is simultaneously merged with two independent syntactic objects—already advocated for in seminal work by McCawley (1982) and Goodall (1987)—and sideward movement by which an element moves to an independent syntactic tree (Nunes 1995,
2001, 2004). The possibility of developing a "theory of Parallel Merge" re-emerged in the late 1990s (Wilder 1999, Citko 2005, Gärtner 1997, among others) as a consequence of the existence of the Linear Correspondence Axiom as a restrictor of its generative power. Similarly, the possibility of sideward movement, which also requires some version of the Linear Correspondence Axiom to avoid overgeneration, emerged as a natural solution with the abandonment of D-structure and the conception of movement as Copy and Merge in Minimalism.

The relation between Parallel Merge and sideward movement is a complex one; while some view them as notational variants of the same structure building operation, others consider them competing hypotheses. It is true that they share a number of properties: (i) both are built on the basis of numerations which contain one single occurrence of the parallel-merged/sideward-moved element; (ii) as has been already mentioned, both strategies crucially rely on some version of the Linear Correspondence Axiom to avoid overgeneration; and (iii) both require the abandonment of the part of the theta criterion that prohibits situations in which an argument has more than one thematic role, independently argued for in Hornstein (1999) and subsequent work.

However, Parallel Merge and sideward movement differ in important respects. (i) In Parallel Merge there is one occurrence of the shared constituent in the syntax, whereas in sideward movement there is more than one copy of the shared element in the syntax, which is relevant when explaining phenomena related to the pronunciation of multiple copies. (ii) Sideward movement leads to a convergent derivation only if an additional copy of the shared constituent is merged in a position which c-commands both the original position of the sideward-moved element and the position to which it has been sideward-moved, whereas in some approaches to Parallel Merge it does not require subsequent movement of the shared constituent, thus allowing "in situ" Parallel Merge, an operation which has been used in "in situ" approaches to Right-Node-Raising (see Wilder 1999, among others). (iii) Leaving aside Across-the-Board constructions, adjuncts cannot undergo sideward movement, as copying the sideward-moved element is a last resort operation which is possible only if triggered by some feature like a theta role in a predicate, whereas nothing prevents the possibility of (parallel-)merging an adjunct with two phrases. The consequence is that whereas they "compete" concerning the explanation of certain phenomena, their empirical coverage is not exactly the same.

In Sideward Movement: Triggers, Timing, and Outputs, Nunes discusses both the status of sideward movement as a structure building mechanism within Minimalism and some of its specific achievements, providing additional arguments which support it and extend its domain of application to the morphological component.

Concerning the status of sideward movement, Nunes shows that once movement is re-analyzed in Minimalism as Copy and Merge, sideward movement does not differ from regular "upward" movement and thus comes as a natural
consequence—i.e. it is actually not a new structure building operation but simply a descriptive artifact. He further shows that sideward movement is constrained by the same conditions which restrict movement in the general case plus independently needed linearization constraints like the Linear Correspondence Axiom, and so no extra requirements are needed to prevent overgeneration.

Sideward movement is crucially constrained by the economy principle of Last Resort on the operation Copy: Copy can only apply if required. As we have already mentioned, this implies that arguments can undergo sideward movement but, leaving aside Across-the-Board constructions, adjuncts cannot, as nothing—i.e. no feature(s)—requires the creation of a copy of an adjunct. Like in “upward” movement, the phase-based computation requires that sideward movement take place from more to less embedded domains. As noted above, an important consequence is that in certain contexts sideward movement is not affected by island conditions due to the fact that it takes place between two independent root objects, i.e. before the “adjunct-to-be” has attached to the matrix and has become an adjunct.

As we have mentioned, sideward movement is only possible if there is subsequent “upward” movement of the sideward-moved phrase to a position which c-commands both the copy in the original position and the one in the sideward-moved position. This is a consequence of the application of (some version) of the Linear Correspondence Axiom; if chain formation, which requires c-command, does not take place, chain reduction becomes impossible and thus the undeleted copies prevent the linearization of the structure.

In the chapter entitled A Parallel Merge Solution to the Merchant/Johnson Paradox, Citko focuses on gapping, an elliptical construction that has been analyzed in terms of ellipsis (Sag 1976) and Across-the-Board movement (Johnson 2009). Assuming Johnson’s conclusion that gapping cannot be treated in terms of movement followed by vP ellipsis, she also argues that Johnson’s Across-the-Board movement analysis itself cannot explain the impossibility of voice mismatches between the clause containing the gapped constituent(s) and the antecedent clause. This is a restriction also noticed by Merchant (2008) for pseudogapping, which he analyzes as resulting from movement of the remnants to a focus position followed by vP ellipsis.

In order to solve the dilemma, Citko revives the Parallel Merge analysis to gapping suggested in Goodall (1987) and Citko (2006a) and proposes that the phenomenon arises when V undergoes Parallel Merge. Where there is only one (shared) V, she argues, only one little v is needed, the consequence being that little v, which encodes voice information, also has to undergo Parallel Merge in these contexts due to an economy condition which states that if only one v is required, then only one (shared) v is possible. The impossibility of voice mismatches thus derives from the fact that there is only one multidominated voice-related head in gapping configurations.

In Clitic Placement and Multidominance, Gračanin-Yuksek challenges the generally accepted idea that only multidominant structures which are in compliance with
some version of the Linear Correspondence Axiom—i.e. which can be linearized on the basis of asymmetric c-command—are possible. She proposes that the Linear Correspondence Axiom is not the crucial factor that constrains multidominant structures on the basis of two types of constructions: Croatian questions in which two fronted coordinated wh-phrases are followed by the auxiliary clitic “je,” which unlike other auxiliary clitics in Croatian follow pronominal clitics, and German coordination of V2 clauses with a single subject. In both cases she argues in favor of multidominant structures and shows that the application of standard revised versions of the Linear Correspondence Axiom do not predict the order of elements attested.

Specifically, Gračanin-Yuksek provides evidence that in both multidominant and non-multidominant structures there are sometimes elements which follow material that c-command, contrary to what we would expect from the application of the Linear Correspondence Axiom. She takes these facts to indicate that the linear order does not entirely depend on syntactic structure. She further shows that all the grammatical cases of Parallel Merge which enter into conflict with the Linear Correspondence Axiom-based approach to linearization comply with the Constraint on Sharing that she advocates.

In Evidence for Multidominance in Spanish Agentive Nominalizations, Fábregas extends the empirical coverage of the Parallel Merge approach to the domain of word formation processes. The phenomenon he focuses on is agentive nominalization in Spanish, in particular, nominalization formed by the suffixation of –dor (roughly equivalent to the English suffix –er). His goal is to derive the properties of this type of derived nominals and to establish the conditions under which remerge is allowed or disallowed.

With the exception of some recent work (Marvin 2007, Alexiadou and Schäffer 2008, or Baker and Vinokurova 2009, among others), agentive nominalizations have received less attention than event nominalizations. However, this type of nominalization presents some intriguing properties that make them particularly interesting for the study of complex word formation as well as for the study of the interaction between morphosyntax and semantics. In the case under analysis, Fábregas shows that in the affixation operation that creates agentive nominals with –dor, the suffix –dor performs a dual role simultaneously. On the one hand, it seems to act as an argument of the verbal base: it requires the verbal stem to allow for a particular thematic position which turns out to be unavailable in the resulting nominalization; roughly speaking, it is as if –dor “absorbed” this thematic position, which Fábregas identifies as the initiator of the event (Ramchand 2008). On the other hand, this suffix is responsible for the change in grammatical category of the derived word.

Under the solution that Fábregas advocates, the nominalizer affix –dor occupies the specifier position of an Initiator Phrase, in the sense of Ramchand. This explains why –dor behaves as an argument of the verb: –dor is interpreted as the initiator of the
process denoted by the complement of Initiator Phrase, Proc(ess)P, which immediately
derives why no other initiator is available in this type of derived nominal. Fábregas
further argues that this affix has a category feature $N$ (“noun”); this N feature, not being
dominated by any nominal functional structure in its base position, is still syntactically
active and needs to project as soon as possible. The fact that the N feature is not
licensed in the base position of –dor (the specifier of InitP) triggers, under this author’s
view, remerge of this suffix. This remerge operation projects the suffix –dor as the head
of a noun phrase that takes the verbal structure as its complement. This explains why
this suffix changes the category of the base. Crucially, remerge is only possible when –
dor does not have a complement, and is therefore ambiguous between a maximal and a
minimal category (in the spirit of Bare Phrase Structure). This immediately prevents a
DP in an argument position from remerging, correctly predicting that verbal phrases
which contain DP arguments in Spec-Init will not be able to become agentive
nominalizations.

Fábregas discusses the advantages of his analysis over competing alternatives
and shows how it can capture the differences between suffixes with properties similar
to –dor, and other nominalizing suffixes like –ción. Under his proposal, the latter
never introduces an argument of a verb, thus remerge is not necessary. This also
explains why this type of suffix, in contrast with –dor, does not require a theme vowel
in the derived complex nominal. A conclusion of his analysis is that the characteriz-
ing properties of complex word formation involving the type of nominalization
analyzed in his chapter can be derived entirely in terms of syntactic operations.

Parenthetical constructions pose a number of problems for all theoretical ap-
proaches but have not received much attention in the literature. In Unconventional
Mergers, de Vries focuses on the analysis of parenthetical constructions which he
shows pose a challenge for any theory of structure building due to the fact that they
exhibit two contradictory properties: they are syntactically integrated into the host
clause but are opaque for the establishment of c-command relations between ele-
ments in the host clause and elements contained in the parenthetical. In order to
solve this tension, de Vries proposes the introduction of Parenthetical Merge, a new
primitive structure building operation. The idea is implemented in the following way:
parentheses are embedded in a parenthetical phrase headed by a “parenthetical
specifying coordinator” head which triggers the application of Parenthetical Merge
which in turn triggers “opacity” for the establishment of c-command relations with
elements outside the parenthetical.

De Vries also discusses the structure of “amalgams,” an intriguing type of paren-
thetical construction characterized by the existence of a constituent “shared” by
(a head of) the host clause and (a head of) the parenthetical. He shows that
these parenthetical constructions are special in that the “shared” constituent enters
into c-command relations with elements both in the parenthetical and in the host
clause. He analyzes amalgams as resulting from the application of two types of
structure building operations: Parenthetical Merge of the amalgam to the host clause plus Parallel Merge –which he calls “external remerge”–of the shared constituent. This explains, he argues, the special status of the shared constituent as far as c-command relations are concerned.

In Constituent Structure Sets II, Bury and Uchida present a proposal to structure building which differs in important respects from those developed within Minimalism. Bury and Uchida propose a non-tree-based structure representation system called “constituent structure sets” and outline some of its main properties. The system is less expressive than common tree-based alternatives, and syntactic copying is only possible in one particular structural configuration. Their hypothesis goes as follows. Consider the numeration set in (1).

(1) The numeration set, $N = \{H, A\}$

Assuming the common numeration set $N$ in (1) above, in a labeled tree like (2a) it is possible to repeat a given item in $N$ on different nodes. Bury and Uchida consider that, while syntactic theories using labeled trees normally restrict the number of such duplications through additional constraints, an alternative and more restrictive system should be pursued which prevents overgeneration without the necessity of additional constraints. The result of this, they argue, would be a system where the notion of tree nodes could be dispensed with, which allows definition of a reflexive dominance ($RD$) relation directly between the items in $N$, rather than between tree nodes decorated with the items in $N$. This leads to a representation system similar to Brody’s (2000) telescope trees as in (2b).

(2) a. Labeled tree:  
     \[ \begin{array}{c}
     H \\
     \downarrow \\
     A \ H
     \end{array} \]

b. Telescope: 
   \[
     \begin{array}{c}
     H \\
     \downarrow \\
     A \\
     \downarrow \\
     H \ A
     \end{array}
   \]

c. $\wp(N) = \{\{H,A\},\{H\},\{A\},\emptyset\}$

d. CSS A: \{H,A\}

e. CSS B: \{H,A\}

With a finite numeration set, it is possible to have only a finite number of telescope structures and each structure is provably finite, a computationally attractive result. But telescope structures cannot duplicate any item of the numeration set and need to duplicate items in the lexicon to deal with certain data, increasing the number of items in the numeration set. An unwelcome consequence is that this makes the system less predictive.

This leads Bury and Uchida to propose the following alternative. For each structure $S$, they define a relation analogous to reflexive dominance between subsets of $N$. Starting with the power set of $N$, i.e. $\wp(N)$ in (2c), $\wp(N)$ is partially ordered with
regard to (reflexive) set containment. If the properties of reflexive dominance are attributed to reflexive containment, then it is possible to get hierarchical relations as in (2d)–(2e). They claim that, as in telescope structures, no independent notion of tree nodes is required, since no two nodes are occupied by the same set of items. If $N$ is finite, then $\mathcal{P}(N)$ is finite. This enables their system to have only a finite number of structures, where each structure is finite; still, their constituent structure sets can duplicate an item in $N$ within a strict limit. Bury and Uchida provide some algorithms to calculate the head of each set in (2d–e), to derive $<H,\{H,A\}>, <H,\{H\}>$ and $<A,\{A\}>$ (i.e. $\langle \text{head,}\{\text{head, \ldots }\} \rangle$). Then, (2d) will represent the syntactic copying of $H$ (i.e. $H$ will head two sets in (2d)) and (2e) will represent a non-copy structure. With the assumption that each set $x$ in $\mathcal{P}(N)$ has exactly one head if $x$ appears in the structure, the number of sets that one item $a$ in $N$ can head in each $S$ is provably finite and hence the number of copies of $a$ is finite.

Bury and Uchida then discuss German V2 data that their system’s restricted duplication of the “head” item can explain neatly. They also briefly consider the implications of adopting constituent structure sets in their treatment of extraction data. Finally, they put their proposal into a historical perspective, making a brief comparison to other structure representations proposed in theoretical syntax, such as dependency graphs and particular set-based structure representations proposed in Minimalism.

As the authors notice, their definition of constituent structure sets results in a system that cannot represent certain common copy/projection structures and multidominance structures, making their proposal incompatible with some of the approaches to phrasal movement defended in this volume (in particular with multidominance systems as well as with (certain versions of) the copy theory of movement).

In the last decade there have been important innovations concerning the feature composition of syntactic heads which have clear consequences for the ability of heads to enter into syntactic relations like Agree, Case valuation and movement, and also for the timing of these operations in the process of structure building. Several of the contributions reflect some of the latest developments in this area and deal directly with this issue and its consequences for structure building.

The goal of the chapter entitled Structure Building That Can’t Be, by Epstein, Kitahara, and Seely, is to present a new deduction of cyclic transfer which overcomes some conceptual and empirical problems that Chomsky’s “valuation-induced” account (Chomsky 2007, 2008) has to face. Their proposal relies on strict adherence to the Strong Minimalist Thesis (optimal satisfaction of the interfaces) and 3rd factor considerations, and is developed within a derivational approach to syntactic relations (Epstein et al. 1998, Epstein 1999, Epstein and Seely 2002, 2006).

They assume Chomsky’s (2007, 2008) feature-inheritance analysis, according to which the phase-heads $C$ and $v$ (inherently) bear phi-features, which they transmit
to the head of their respective complements, T and V; once endowed with these
features, T and V can then raise a local DP to their specifier via their edge feature.
Epstein, Kitahara, and Seely argue that this operation, which can be characterized as a
“counter cyclic” internal merge, is problematic as it necessarily incorporates an
additional “replacement” process which goes against the Law of Conservation of
Relations, an economy condition that states that in narrow syntax syntactic relations
among existing terms cannot be altered throughout a derivation. If “replacement”
processes are not possible, then countercyclic Internal Merge (subject raising) cannot
infix Spec/TP into the C-rooted object, but rather creates a “two peaked” object: two
distinct but intersecting set theoretic objects which share a term but no single root.
Since there is no single root that (reflexively) contains all terms, the Label Accessibil-
ity Condition (a 3rd factor efficiency consideration) will prevent these “two peaked”
intersecting objects from undergoing Merge, and the derivation halts. In order to
continue, one of these intersecting “peaks” (namely the phase-head complement)
must be eliminated, forcing Transfer to apply. This deduces cyclic Transfer: “always
and only when a phase head is externally merged and phi-features are inherited by
the head of the phase-head complement.”

As further evidence of the superiority of their approach over previous explana-
tions, Epstein, Kitahara, and Seely show that their system can also provide an analysis
of some object agreement facts in Icelandic which were problematic for the timing
of Transfer as formulated in Chomsky (2007, 2008) (see Bošković this volume
for related discussion). Finally, they also show that they can derive the “invisibility”
of Spec-T to C as a property of the derivational approach to syntactic relations they
defend here and have advocated in previous work (see references above): C is not in
a relation with Spec-T because at the point of the derivation when C was merged to T,
the Spec-T position had not yet been created.

An interesting feature of this chapter, which distinguishes it from alternative
approaches to structure building defended in this volume, is its conception of how
Merge applies. Epstein, Kitahara, and Seely defend the view that, while Merge is a
core part of the human faculty of language, the way in which Merge operates is
constrained by 3rd factor considerations. The basic idea is that narrow syntax
executes Merge with minimal search, so that computational complexity is greatly
reduced. Under this view, merger of two roots (External Merge) is always preferable
to merger of one root and its subpart (Internal Merge) because finding roots involves
no search. Consequently, all applications of External Merge will necessarily precede
any application of Internal Merge within the construction of each phase. Since, by
the Label Accessibility Condition, narrow syntax has access only to the highest label
of the root, only External Merge of two roots is possible. Thus, under Epstein,
Kitahara, and Seely’s analysis, narrow syntax cannot “reach” inside the root and
“pull out” its term for External Merge. However, after narrow syntax exhausts
External Merge, search into the root will be permitted: these operations include
feature inheritance, Agree, and Internal Merge. This conception of how Merge applies
draws a dividing line between the approach defended by Epstein, Kitahara, and Seely
and alternative approaches like sideward movement in Hornstein (2001) and Nunes
(2001) and the multidominance analysis in Citko (2005) and subsequent work.

The new approaches to structure building have also led to important changes in
our conception of the paths of movement, both with regard to the locality conditions
and triggers of movement as well as with regard to the type of chains allowed.

Concerning the type of chains derived by movement, under the traditional
approach, A-movement was assumed to feed A’-movement in wh-chains involving
arguments. In this regard, movement of the wh-subject to Spec/CP, for instance, was
assumed to be preceded by movement of the wh-phrase from its base generated
position to Spec/TP and from there to Spec/CP. This classical view has been
challenged in work by McCloskey (2000) and Holmberg and Hróarsdóttir (2003).
On the basis of the discussion of quantifier float (QF) in West Ulster English,
McCloskey concludes that in contexts involving quantifier float under wh-movement
the subject-wh moves directly to Spec/CP, without being preceded by movement
to Spec/TP. The analysis of agreement facts in the context of wh-movement in
Icelandic led Holmberg and Hróarsdóttir to the same conclusion. These complex
facts have been recently reinterpreted under what has been dubbed The Parallel
Movement Analysis (TPMA) (Hiraiwa 2005, Chomsky 2008). The major tenet of the
TPMA is that the wh-element moves directly to Spec/CP, but it also moves in parallel
to Spec/TP. This results in the possibility of filling up Spec/TP despite the absence
of a feeding relation between the movement of the wh- to TP and the movement of
this wh-element from TP to the higher Spec/CP position.

In his contribution to this volume, Don’t Feed Your Movements When You Shift
Your Objects, Bošković, adopting Chomsky’s (2008) analysis, provides further sup-
port for The Parallel Movement Analysis on the basis of object shift in Icelandic (see
Epstein, Kitahara, and Seely this volume for related discussion). This language allows
quantifier float under object shift but disallows it in contexts involving movement
to Spec/CP. Bošković shows that this contrast is amenable to an analysis in terms of
parallel movement. Under his proposal, the contrasts in the possibility of quantifier-
floating are straightforwardly derived if object shift does not feed wh-movement. The
solution he presents also provides him with a test to weigh the advantages and
disadvantages of two of the leading proposals put forth to account for object shift
in this language. His conclusion is that an account of object shift where it involves
movement to a position above Spec/vP/SpecAgroP (see, among others, Bošković
1997, 2004a,b, Hiraiwa 2001, Svenonius 2001) is to be preferred to the analysis where
the final landing site of Icelandic object shift is the accusative case position (either
Spec/vP or Spec/AgroP).

Tough-constructions constituted one of the most intricate puzzles of Government
and Binding. They cannot be built on the basis of movement from the object position
of the embedded clause to the subject position of the matrix clause because this implies violations of generally assumed conditions on movement: the Chain Condition (the chain obtained had two case positions) and the improper movement condition if movement proceeds via the intermediate Spec/CP. On the other hand, Chomsky’s (1981) object PRO analysis and the generally adopted null operator approach proposed in Chomsky (1982) have the unwelcome result that the subject of tough-constructions has to be assumed to be merged in a non-theta position.

In Feature-Splitting Internal Merge: the Case of tough-constructions, Obata and Epstein propose a “proper improper movement” analysis of this construction that avoids the above-mentioned shortcomings.

Obata and Epstein first develop a novel theory of the feature composition of T and V which has the following properties: (i) the feature-inheritance relation takes place between C and T and between ν and V (see also Epstein, Kitahara, and Seely, this volume); (ii) finite T bears [uPhi], whereas non-finite (Control) T bears [uCase]. In this respect Obata and Epstein’s system clearly departs from other approaches like Chomsky’s (2001, 2007, 2008) where both finite and non-finite T bear identical [uPhi]. (iii) [uPhi] values phonologically realized Case, whereas [uCase] values Null Case (PRO). There are thus three types of T and V depending on the existence and the types of features of C and ν respectively. If there is no C/ν, no Case-valuing features are inherited by T/V, which thus cannot value Case. If T and V have [uCase], they value null case (PRO). If T and V have [uPhi], they value nominative and accusative Case respectively. This means that, unlike in traditional approaches to the distribution of PRO, in Obata and Epstein’s system PRO is possible in the complement position of V.

Obata and Epstein further adopt Feature-Splitting Internal Merge, a novel type of structure building operation proposed in Obata and Epstein (2008, 2011) by which a single element can be split into two syntactic objects allowing its features to split into two landing sites. As a consequence of feature-splitting and the theory of feature composition they develop, an element can enter into multiple Case-checking relations and a non-phonologically specified Case like Null Case on PRO can be “revised” to a phonologically specified Case in the course of the derivation.

With this background, they can explain the possibility of improper movement in tough-constructions and the impossibility of traditional improper movement configurations as follows. In impossible cases of improper movement, the improperly moved DP is attracted to the edge of the embedded CP only with the [iQ] feature, so that the matrix T probe cannot value its [uPhi] and the derivation crashes. In “proper improper movement” tough-constructions, on the contrary, the moved DP is attracted to the edge of CP with an [iPhi] feature, which is accessible to the matrix T, which can thus value its [uPhi].

Until recently, there has been general agreement in the generative tradition that structure building operations which involve movement are sensitive to locality
effects, which are thus typically used as “detectors” of movement, be it overt (before spell-out) or covert (after spell-out and therefore with no impact on PF). Disagreement exists, however, in two areas. The first involves a number of issues related to the relation between movement and its (non-)PF effects; one issue is whether or not overt movement can be (string-)vacuous,—i.e. with no effect on linear order. A perhaps more controversial issue is whether or not covert movement exists. The second area of debate is related to the identification of the locus of the explanation of locality effects on movement, as well as the (unexpected) lack of locality effects in certain contexts.

Minimalist theorizing has provided a new ground for the discussion of many of these issues. There is a line of thought, for instance, according to which the distinction between “overt” and “covert” movement should not be conceived of as reflecting a difference in “timing,” i.e. whether they differ with respect to the point in the derivation in which they take place, but as a difference in the PF effects that they trigger, i.e. whether it is the head or the tail of the chain that is pronounced (see Fox and Nissenbaum 1999, among others). Under this conception, the overt vs. covert movement distinction is thus an epiphenomenon, and the question of whether there is “overt” (string-)vacuous movement should be looked at through different lenses. As for the locus of the application of the principles of locality, opacity (island) effects have been argued to be the result of well-formedness conditions on the PF component (see Hornstein, Lasnik, and Uriagereka 2003, Lasnik 2001, Merchant 2001).

Regarding the role of locality effects as detectors of movement, in the debate concerning how Right- and Left-Node-Raising structures are built, the absence of locality effects has strongly militated against the Across-the-Board movement analysis of the construction (Ross 1967, among others) and in favor of the in situ approach (Wilder 1999, among others).

In “Lasnik-Effects” and String-Vacuous ATB Movement, Abe and Hornstein, assuming both a rightward Across-the-Board movement approach to Right-Node-Raising and the above-mentioned PF-based conception of the overt vs. covert movement distinction, propose that in Right-Node-Raising structures movement of the shared constituent is sometimes “covert” (it is the foot of the chain that is pronounced), whereas at other times it is “overt” (it is the head of the chain that is pronounced). They further argue that string-vacuous movement is always “covert” due to a requirement that “overt” movement have a PF effect, so that string-vacuous Right-Node-Raising is “covert” movement whereas non-string-vacuous Right-Node-Raising is “overt” movement.

They show that string-vacuous Right-Node-Raising is not sensitive to locality (island effects) whereas non-string Right-Node-Raising is, and explain this contrast on the basis of a PF approach to locality. Moved elements extracted from islands are marked with “*”, a non-legitimate PF object. The absence of island effects in string-vacuous Right-Node-Raising results from the fact that the copy in the head of the
chain is deleted, thus eliminating the offending “*”. In non-string-vacuous Right-
Node-Raising, on the other hand, it is the foot of the chain that is deleted, resulting in a non-convergent PF representation due to the fact that “*” survives.

Much of the work carried out during the 1980s within the Government and Binding model tried to give a principled account of the existence of barriers and islands for movement. Huang’s (1982) analysis in terms of the Condition of Extraction Domains, which crucially relied on the +/- complement nature of the constituent which constitutes the extraction domain, set the basis of what would become the classical approach to islands in terms of government: only those constituents that are governed will be transparent for extraction (Chomsky 1986, Lasnik and Saito 1992, among others). Under this view, adjuncts would always be inherent islands as, characteristically, they are never governed.

In recent years, however, we have seen alternative approaches to islands, and to the islandhood of adjuncts, which do not rely on the notion of government, but rather on the way in which the structure is built. The so-called Late Insertion of Adjuncts Hypothesis (Lebeaux 1988, Stepanov 2001), for instance, argues that the reason why adjuncts are islands is because they enter the derivation later than arguments. Other works have proposed a radically different view where adjuncts are not islands during the whole process of structure building, making it possible to extract out of the adjunct at certain well-defined steps of the process: before this constituent merges with the structure and becomes an island. As discussed above, this view has been successfully pursued within the sideward movement approach, represented in this book by Nunes’ chapter. In the chapter entitled On Transparent Adjuncts in Japanese, Miyamoto defends an alternative approach which explains, in terms other than sideward movement, why some adjuncts are exempted from islandhood during the process of structure building.

Based on the analysis of Japanese comparative deletion, Miyamoto shows that there is an asymmetry in the possibility of extracting a constituent out of an adjunct in this language: while extraction out of an adjunct which is a subject-oriented secondary predicate is always prohibited, it is sometimes possible when this secondary predicate is object-oriented. These facts lead him to conclude that adjuncts are not inherent islands. Miyamoto proposes that adjuncts can be exempted from their island status whenever they enter into an Agree relation with an element in the matrix clause. Adopting the syntactic approach to inner aspect by MacDonald (2006, 2008a,b, and 2009), Miyamoto argues that the reason why object-oriented secondary predicates allow extraction from within is that they enter into an Agree relation with Aspect; this is not possible for subject-oriented secondary predicates, as they are too high in the structure to establish any feature relation with this functional head. As further evidence for this aspectual feature-based approach, Miyamoto discusses some asymmetries in the possibility of extraction out of the secondary predicate observed between predicates which exhibit an unergative/unaccusative alternation.
The results he presents support the conclusions in Demonte (1988) and Borgonovo and Neeleman (2000), who on the basis of data from Spanish, English, and Dutch, also show that not all adjuncts are barriers for movement.

Under one of the most influential approaches, movement is defined in terms of *Attract*: the attracting element (the probe) can attract the moving element (the goal) when there is an affinity between the set of features of these two elements, that is when both sets of features are compatible. There is however an alternative approach to movement, where movement is conceived precisely as a consequence of the incompatibility, or feature mismatch, between the set of feature $F_a$ and $F_β$ on two nodes $α$ and $β$. This is the view that has been put forth under what is known as The Survive Principle (see Stroik 1999, 2009, Putnam 2009, Lechner 2009; for earlier implementations of push chains in syntax, see, among others, Moro 2000, van Riemsdijk 1997).

In the chapter entitled *Structure Building From Below: More on Survive and Covert Movement*, Lechner contributes to this line of research by providing a new formulation of The Survive Principle, and by looking for criteria which can help distinguish The Survive Principle from standard theories which define movement in terms of feature attraction. Lechner provides a detailed account of some complex scope restrictions found in English, comparing the different scope possibilities displayed by a variety of quantificational configurations. The first configuration involves contexts with three quantifiers in an asymmetric c-command relation (in particular, the contrasts exhibited by constructions involving ditransitive predicates in the Double Object construction (IO DO frame) and the DP PP frame (DO to IO). Then he analyzes triples of quantifiers, two of which are in a containment relation (inverse linking). Finally, he also examines the possible and impossible construals exhibited by VP fronting in contexts where three quantifiers are involved. He shows how his approach, based on the new definition of The Survive Principle he provides, can explain the possible and impossible readings of all these cases, motivating the tucking-in ordering that is needed to account for them.

Lechner’s analysis provides support for the following conclusions. First, the triggers for movement should be extended to semantic properties such as the logical type of an expression. Second, movement paths are dense: all movement proceeds in the smallest possible step (density being the property that most radically distinguishes The Survive Principle from Attract-based models, according to Lechner). Third, a decision procedure on the order of movement that relies on some concept of closeness—such as the Minimal Link Condition (MLC) (Chomsky 1995b, 2000, 2001)—is bound to fail. Finally, the analysis supports a single output model of the grammar.

As mentioned above, the Minimal Link Condition is one of the most influential formulations of locality. The Minimal Link Condition requires that movement and Agree/Case relations be established between a probe and the closest available goal. We have just seen that Lechner’s chapter provides arguments against the MLC based
on a variety of contexts involving three quantificational elements. In the chapter entitled *Specificity-driven Syntactic Derivation*, Lahne defends the same view and argues against appealing to the Minimal Link Condition to derive locality effects. She differs from Lechner, however, in framing her analysis within the standard “compatibility of features” approach. She adopts the assumption that each phrase is a locality domain (Manzini 1994, Epstein and Seely 2002, Müller 2004, Lahne 2008c), and criticizes the Minimal Link Condition because, as pointed out by Müller (2006a), this condition overlaps with the Phase Impenetrability Condition (Chomsky 2001) in deriving the locality effects due to Superiority. She further argues that the Minimal Link Condition does not make the right predictions for syntactic constellations involving equidistance between a probe and a goal in contexts with more than one goal.

She provides an alternative approach to locality according to which syntactic derivations are driven by a specificity principle on Merge, the General Specificity Principle (an extension of the Maximize Matching effects in Chomsky 2001): when more than one constituent is a candidate for an agreement or displacement operation, the operation will involve the more specific constituent (where specificity is to be understood not in its formal semantic sense, but in its morphological sense). What follows from this is a system that derives locality effects in a new way: not the closest available goal is chosen, but the most specific one. She shows that her system can account not only for the syntactic constellations that the Minimal Link Condition was designed to explain (Superiority, Equidistance, A-over-A, Two-Edge case), but also for other configurations that the Minimal Link Condition had problems in dealing with and which involve anti-Minimal Link Condition effects (in particular, order preserving movement and antisuperiority effects). The resulting system is able to unify a wide variety of locality effects that were accounted for by independent and, in certain cases, partially overlapping principles, and provides support for the hypothesis that the concept of intervention by closeness in its classical interpretation should be abandoned. While the relevance of specificity was well known in morphology (the Subset Principle, Halle 1997), if Lahne’s analysis is correct, it suggests that specificity might also play an important role in grammatical operations in general.

There are two key issues in contemporary linguistic theory concerning the relation between syntactic structure and word order. A traditional concern has been whether the structure building mechanism responsible for cross-linguistic variation in the relative order of heads and their complements is based on parametric variation related to the properties of External Merge, i.e. the existence of a head-parameter, or whether it is based on parametric variation related to the presence of features which trigger Internal Merge, i.e. there is a basic universal word order, so that the orders that do not conform to it must be derived by movement. The second is how the hierarchical structure resulting from the application of External and Internal Merge is
related to linear order. A related and important question is that which deals with the appropriate explanation of “disharmonic” word orders (the presence of mixed head-complement and complement-head orders within the same language) and of the limits of “disharmony.”

In *Disharmony, Antisymmetry and the Final-over-Final Constraint*, Biberauer and Sheehan deal with these two issues, concentrating on the analysis of the descriptive generalization which imposes the most stringent restriction on disharmony: the Final-over-Final-Constraint, i.e. the fact that no structure can exist with a head final phrase dominating a head-initial one. The effects of the Final-over-Final Constraint become apparent in the impossibility of having head-initial CPs in head-final VPs universally: *[VP [CP C TP] V]}. In order to avoid this pattern, languages resort to “extraposition” of the CP, resulting in [VP V [CP C TP]] surface order.

Assuming with Kayne (1994) that the universal order is Spec-Head-Complement as well as some version of the Linear Correspondence Axiom, Biberauer and Sheehan argue that complement-head orders are derived when the relevant head X is endowed with a \(^\) feature (also called an edge or EPP feature in the literature) that triggers movement of the complement of X to the specifier of X.

As for the effects of the Final-over-Final Constraint, Biberauer and Sheehan propose that the unattested *[VP [CP C TP] V] is impossible due to a violation of the Linear Correspondence Axiom, i.e. it is an unlinearizable structure, and that the surface “extraposed” order [VP V [CP C TP]] actually involves a more complex structure in which the relation between V and CP is mediated by a little n which takes CP as its complement. After movement of nP to Spec-V triggered by \(^\) on V, “scattered deletion” applies as a last resort strategy in order to avoid a representation which is not in compliance with the Linear Correspondence Axiom, causing CP to spell out in its first-merge position: [VP [nP [CP C TP]] [V V^ [nP [CP C TP]]]]. Their analysis allows them to explain the intriguing constraints on word order patterns observed cross-linguistically.

Although some of the chapters in this volume deal with more than one of the questions raised at the beginning of this introduction, for presentational purposes they are divided into two major sections. Those in the first section, *Merge and beyond*, provide different answers to the first question by presenting alternatives to the traditional conception of how Merge operates. The chapters in the second section, *Triggers and constraints*, put forth some specific proposals that analyze what the triggers and constraints on structure building operations are, i.e. the other two questions introduced above.

We hope that the chapters which follow will help the reader gain a deeper understanding of the complex mechanisms that underlie the process of structure building in natural languages.
Part I

Merge and beyond
This page intentionally left blank
Constituent Structure Sets II*

DIRK BURY AND HIROYUKI UCHIDA

2.1 Introduction

The expressive power of a grammar has been a central concern of formal linguists since Chomsky initiated this line of research in the 1950s. Finding the appropriate level of restrictiveness for a grammar is essential if it is to provide a plausible foundation for a theory of language acquisition or for a predictive theory of language variation, as well as for computational applications.

There are different ways of limiting the expressive power of a grammar. Broadly one can distinguish between restrictions of the base component, which defines what structures are available, and restrictions on the transformational component, which defines what operations can be performed on those structures (Chomsky 1965, McCawley 1968). Restrictions of the transformational component include the simplification of the shape of transformational rules, which culminated in the single, simple rule move-alpha (Chomsky 1981), or the requirement that all movements be triggered (Chomsky 1995b). Restrictions of the base component include constraints on the shape of possible base rules, such as X-bar theory (Chomsky 1970, Jackendoff 1977), and attempts to derive the properties of possible structures from deeper principles, such as the LCA (Kayne 1994).

The proposal we develop in this chapter involves a particular way of restricting the expressive power of the base component, which in turn has significant ramifications

---

* The work presented here has benefited from the input of many people. We are grateful to Ad Neeleman, Andrew Carnie, Annabel Cormack, Dick Hudson, Hans van de Koot, Klaus Abels, Lutz Marten, Marcus Kracht, Matt Reeve, Michelle Sheehan, Natasha Kurtonina, Nathan Klinedinst, Neil Smith, Nick Allott, Nick Cassimatis, Pam Macdonald, Reiko Vermeulen, Rob Truswell, and Thiago Galery, as well as to the audiences at Ways of Structure Building (Vitoria-Gasteiz), Trees & Beyond (Brussels), Mathematics of Language 10 (UCLA), and LAGB’07 (King’s College London). We would also like to thank the anonymous reviewers and the editors of this volume for their comments and suggestions. Any errors and unclarities that remain are the authors’ fault.
for the analysis of certain movement phenomena that have traditionally been treated in terms of transformations.

Although our proposal, like those mentioned above, aims to reduce the expressive power of the grammar, it achieves this goal in a fundamentally different way. Instead of proposing a particular grammatical theory, we propose a novel structure representation system called **constituent structure sets** (CSS), which replaces labeled trees as representations of natural language syntactic structures. Since our representation system can potentially be used in different syntactic theories, our proposal is theory-neutral. However, since our structure representation system has much weaker expressive power than labeled tree representations, adopting our system automatically rules out certain theoretical options, such as phrasal copying, as we will show later.

In contrast, the X-bar schema mentioned above can be considered as a tool that can be used within a theory that uses labeled tree structures in order to filter out undesirable structures that in principle are legitimate examples of labeled tree diagrams. Similarly, various theoretical tools have been used to exclude undesirable syntactic structures that are possible according to the algebraic properties of labeled tree structures, such as the use of feature checking for restricting syntactic copying (cf. Chomsky 1995b). Such theoretical constraints can avoid the over-generation of the grammar as a whole, even with labeled trees being used as part of the theory. However, achieving restrictiveness in this way generally increases the computational complexity of the grammar so that the recognition time becomes unmanageable (cf. Stabler 2004). Using a representation system which has much weaker expressive power and which has the finiteness properties built into its basic algebraic properties is one way of decreasing the complexity of the grammar.

As we implied above, our structure representation system is not based on trees and it is less expressive than common tree-based alternatives. In particular, in our system, syntactic copying, which plays a crucial role in both analyses of movement and, under the guise of projection, analyses of constituency (for discussion, cf. Speas 1990, Bury 2003, Boeckx 2008), is only possible in one particular structural configuration. Specifically, the system we propose is incompatible with certain current approaches to phrasal movement, namely with multidominance systems (cf. Gärtner 2002, Citko 2005, this volume, Graćanin-Yuksek 2007, this volume, for recent proposals), and with (certain versions of) the Copy theory of movement (cf. Chomsky 1995b, Nunes 2004, among many others).

We now provide an intuitive explanation of our structure representation system in comparison to labeled tree structures on the one hand and what we call, adapting a term of Brody (2000), “telescope structures” on the other.

In order to compare the expressive powers of the three representation systems abstracted away from details inessential for our purpose, we keep two elements
constant. First, we assume that each syntactic structure comes with a finite numeration set \( \text{Cat} \), as in (1a) below and that the three representation systems share the same numeration set. For comparing the formal expressive powers, it is not essential exactly how the members of the numeration set are selected. However, in linguistic application, it will become important to provide independent criteria for selecting the members of the numeration set for each natural language structure. For concreteness, we adopt the idealized assumption that the numeration set contains the same number of items as the overt expressions that appear in the phonological string to be generated by the structure.\(^1\) For example, if the sentence that we want to generate is *Meg ran*, then the numeration set is \( \{V,D\} \), where \( V \) is for *run* and \( D \) is for *Meg*, as in (1a). Because of this isomorphism between the categories in the numeration set and the phonological words in the phonological string, we sometimes represent the numeration set using categories, such as \( \{V,D\} \) and sometimes represent the set using phonological words, such as \( \{\text{run, Meg}\} \).

Secondly, we assume that each structure is equipped with a particular kind of partial order, that is, a reflexive dominance relation as is familiar in tree structures.\(^2\) We keep the common properties of reflexive dominance, such as transitivity, reflectivity, and the presence of the maximal element (i.e. “rootedness” in the tree notation) constant for the three representation systems.\(^3\) In section 2.2, we define such properties as properties of relational structures. In this section, however, we simply use tree notations for convenience to make it visually clear that the three representation systems share the same properties of reflexive dominance, as we will see in (1)–(2). With these two assumptions, consider (1).

\(^1\) In the main sections, we show how we can relax this strict requirement in linguistic application, such as the inclusion of a T head that does not correspond to a separate overt word in a trivial way and a C head in the analysis of verb-initial clauses in German.

\(^2\) We consider reflexive dominance (RD) as the basic relation from which we can derive immediate dominance (ID) or its equivalent in CSS, if necessary. This is because ID is a special case of RD and can be derived from RD without a disjunctive condition. In contrast, deriving RD from ID requires a disjunctive condition and also expands the set of items that enter into the partial order. Moreover, ID cannot always be maintained via every P-morphic mapping between structures, whereas RD can, cf. Kurtonina (1994: 32). These considerations suggest that RD is more basic in relational structures, though ID might be more basic in a derivational grammar presentation, cf. Cornell (1998). For reflexive transitive closure in terms of immediate dominance relation, see Kepser (2006). An anonymous reviewer wondered why we do not base our structure on proper (i.e. irreflexive) dominance, but most of the arguments above apply to proper dominance (PD) as well. That is, we cannot derive RD from PD without adding a disjunctive clause and neither can we maintain PD via every P-morphic mapping. For further details, see Bury and Uchida (2005).

\(^3\) In fact, as we see shortly, the basic syntactic relation in CSS is based on the set containment relation. Since sets are not discrete items while tree nodes and category names are, some of the relational properties need to be defined in a slightly different way in our system. See section 2.2 for details. However, except for such minor differences, the basic properties of RD are maintained in our system as well.
(1)  

a. The numeration set, $Cat = \{V,D\}$ (or $Cat = \{\text{run, Meg}\}$)

b. a labeled tree:

c. Telescope A:

d. Telescope B:

In a labeled tree as in (1b), the reflexive dominance relation links tree nodes which in turn are decorated with items of the numeration set $Cat$.\(^4\) Since the notion of syntactic nodes is independent of the notion of categorial labels (which are the members of the numeration set $Cat$) in labeled trees, we can repeat the same item of $Cat$ in different syntactic nodes which are then linked by the reflexive dominance relation. For example, in (1b), $V$ is successively duplicated onto the mother nodes and $D$ is copied into a terminal node. If we assume that there can be a potentially infinite number of syntactic nodes, then, even if the numeration set is finite, we can generate a tree structure of an infinite size and we can also generate an infinite number of distinct syntactic structures. As we have indicated above, in practice, no syntactic theory makes use of the full expressive power of such labeled tree representations. However, as we have also indicated, filtering out the excessive copying possibilities that labeled trees can represent in principle tends to increase the complexity of the syntactic theory.

Thus, while it is easy to think of tools that could be added to a theory to restrict the availability of such syntactic copying, an (arguably more principled) way to avoid excessive duplication could come from a structure representation system that simply cannot express excessive duplication.

One way to develop such a restrictive system would be to abandon the notion of tree nodes and instead define a relation of reflexive dominance directly between the items in the numeration set, that is, between categories, rather than between tree nodes decorated with categories. A generative theory that does something like this is mirror theory (Brody 2000; see also Abels 2001, Adger et al. 2010). Brody assumes that structural representations contain only word-level categories, what he calls the telescope assumption.\(^5\) We adopt this term here and use it to refer to structures that only employ categories but that make no use of the concept of nodes as telescope structures. In such a system, given the numeration in (1a), we can then generate

---

\(^4\) Following the standard convention, we use italics for (meta-)variables. We use intuitive meta-variables such as $Cat$ for a numeration set (mnemonic for “categories”) or $RD$ for a reflexive dominance relation (note that $RD$ in non-italic capital is an abbreviation for “reflexive dominance,” not a variable).

\(^5\) As Brody notes, in this respect mirror theory is similar to Dependency Grammar; see section 2.7 for discussion.
representations like (1c). In fact, (1c) is the only telescope structure that we can generate with \( \text{Cat} \) in (1a) and with \( V \) as the maximal (or top) element. This is because one structure is equipped with exactly one partial order, that is, a reflexive dominance relation, and in each telescope structure, the structure’s unique reflexive dominance relation directly orders the items in \( \text{Cat} \). The only other telescope structure that we can have with \( \text{Cat} \) in (1a), that is, (1d), has \( D \) as the maximal element instead. Now, as is clear with these examples, for any finite numeration set, each telescope tree that we can generate with it is provably finite and we can generate only a finite number of distinct telescope structures. The former finiteness is obvious since each reflexive dominance relation directly orders the items of the numeration set in a telescope structure. That is, the number of the items that enter into the reflexive dominance relation is the same as the cardinality of the numeration set which is finite by assumption. Similarly, there is only a finite number of distinct ways of partially ordering a finite numeration set by way of a reflexive dominance relation. Such provable finiteness properties make the system computationally attractive. However, telescope structures cannot duplicate any item of the numeration set and as a result, if we do need to duplicate the same item in the numeration set for dealing with some linguistic phenomena, then we need to duplicate items in the lexicon, increasing the number of the items in the numeration set as the data require.\(^6\) We think that it is more restrictive to systematically constrain the duplication of a lexically provided item in the syntax, rather than duplicating an item in the numeration set in an arbitrary manner.

Since labeled tree structures arguably allow for too much copying, and telescope structures for too little, we propose a structure representation system that in terms of its expressive power lies between the two. We define a basic relation that is analogous to the reflexive dominance relation, but that holds between subsets of the numeration. That is, the basic relation of our system holds between sets of categories, not between nodes in a tree as in systems using labeled trees, or between individual categories as in a telescope system.

Given the numeration set \( \text{Cat} \) in (1a), we first form the power set of \( \text{Cat} \), i.e. \( \mathcal{P}(\text{Cat}) \), as shown in (2a).

\((2)\)

\[ \begin{align*}
\mathcal{P}(\text{Cat}) & = \{ \{V, D\}, \{V\}, \{D\}, \emptyset \} \\
& \\
\text{CSS A:} & \\
\{V, D\} & \{V, D\} \\
\{V\} & \{D\} \\
\{D\} & \end{align*} \]

\(^6\) See Uchida and Bury (2008) for an example of such an arbitrary expansion of the numeration set with a telescope structure.
$\wp(Cat)$ is partially ordered with regard to (reflexive) set containment (RC). If we attribute to RC the properties of RD, then we can generate hierarchical relations as in (2b)–(2c). Just like telescope structures, this system makes no reference to an independent notion of tree nodes, since no two nodes are occupied by the same set of items. If $Cat$ is finite, then $\wp(Cat)$ is finite. Thus we can have only a finite number of structures and each structure is finite. On the other hand, unlike telescope structures, the structure representation system that we propose can duplicate an item in $Cat$ within a strict limit. First, for each set of categories in (2b–c), we choose one of its members as its “head”. We then form a complex set that contains the head and the set of categories as its members, in the form of \{head, \{head, \ldots\}\}. We call such a complex set a treelet (see section 2.2 for the formal definition). For example, we can replace the sets of categories, \{V,D\}, \{V\} and \{D\}, in (2b–c) with the treelets, \{V,\{V,D\}\}, \{V,\{V\}\} and \{D,\{D\}\}, respectively. Then, (2b) will represent the syntactic copying of V (i.e. V will head the two sets \{V,D\} and \{V\} in (2b)) while (2c) will represent a non-copy structure. If we were to set the head of the maximal member of $\wp(Cat)$ as D, as in \{D,\{V,D\}\} instead, then (2b–c) would become the two distinct structures with the same hierarchical ordering as in Telescope B in (1d), that is, with D being higher than V in the structural hierarchy. (2b) would then become the copy structure for D (whether such a structure is useful in linguistic application is a separate issue, just as it is not clear if (1d) is useful in application). With the assumption that each set $x$ in $\wp(Cat)$ has exactly one head if $x$ appears in the structure, the number of sets that one item $a$ in $Cat$ can head in each syntactic structure is provably finite and hence the number of copies of $a$ is finite.

After defining our structure representation system in section 2.2, we show that our system cannot distinguish some common copy/projection structures from the non-copy/non-projection structures in section 2.3. Section 2.3 also shows that our system cannot represent multiple dominance structures as distinct. Section 2.4 discusses German V₂ data that we can explain neatly with the restricted duplication of an item in $Cat$ as above. In section 2.5, we sketch how we can deal with wh-extraction with our system. Section 2.6 sketches how to generate phonological strings. Section 2.7 compares our system to some other structure representation systems and considers how the proposed system relates to certain proposals in generative syntax. We also clarify where our proposal stands in the development of Chomskyan syntactic theories. Section 2.8 provides concluding remarks.

### 2.2 Constituent structure sets (CSS)

This section provides the exact definition of the proposed representation system. Unlike labeled tree structures, the system simply cannot represent the syntactic copying of one item in an infinite number of times. However, our representation system can still express a limited number of copies in the structure. We suggest that
this restricted amount of syntactic copying is useful in linguistic application. Most importantly, this restriction falls out from the basic definition of our representation system.

The foundational idea of our structure representation system is based on the representation system proposed in Bury (2003). Informally, this system replaces a traditional graphical syntactic tree with a set of “constituents,” where we call each such set a constituent structure set. Each constituent of such a syntactic structure, that is, each member of a constituent structure set, in turn is a set of categories taken from the numeration set for the syntactic structure, where one of those categories is designated as the head of this constituent, as we represented in the form of \{head, \{head, \ldots\}\} in section 2.1. As we emphasize repeatedly, a main merit of replacing each graphical tree by a set made out of the items taken from the numeration set is to provide a formal maximal bound to each syntactic structure.

Now we formally define our syntactic structures from top down. First, each syntactic structure is defined as in (3).

(3) \text{Structure := } \langle \text{Cat, } \text{CSS, } \text{RC} \rangle

\text{Cat} is the numeration set, or the set of categories, which is isomorphic to the set of selected lexical items for the structure in our analysis. In linguistic application, it will become important to restrict the membership of \text{Cat} by independent criteria. Thus, we assume that for each structure the number of items in \text{Cat} corresponds to the number of overt PF expressions. As we indicated in section 2.1, we do not really need to assume the isomorphism between the numeration set and the set of overt phonological items, although we do stick to it for presentation reasons. In actual linguistic analysis, we may assume that \text{Cat} can include T even if it does not host an overt word in the string to be generated, since T is linguistically well motivated, for example, in terms of the tense information that is often expressed as verbal morphology, the agreement relation between the verb and the subject (in the spec of T), and the special status of the external argument of the verb.

\text{CSS in (3) is mnemonic for “constituent structure set,” corresponding to the notion of a “set of constituents” as informally explained above. Formally, for each } a \in \text{Cat}, \text{the constituent structure set CSS contains at least one treelet } x, \text{ where each treelet } x \text{ is a complex set in the form of } \{a, Dx\}, \text{ following the restrictions in (4).}

(4) \text{Treelet:}

\text{a. In each structure, for each } a \in \text{Cat}, \text{ CSS contains at least one treelet } x \text{ such that } x=\{a, Dx\} \text{ and } \{a\} \subseteq Dx \subseteq \text{Cat.}

\text{b. In each structure, for each } x \in \text{CSS and for each } a \in \text{Cat}, \text{ if } a \text{ is the head of } x, \text{ then } \{a\} \subseteq Dx \subseteq \text{Cat.}
In words, each treelet \( x \in \text{CSS} \) has the head \( a \) and the dominance set \( D_x \). While each treelet corresponds to a constituent, the dominance set lists the categories that this constituent contains. Thus, if a category \( a \) heads a treelet, the dominance set in the treelet must contain \( a \) as a member, as stated in (4b). Also, remember the intuitive exposition of CSS at (2) in section 2.1 and notice that each dominance set \( D_x \) as in (4) corresponds to a member of \( \mathcal{P}(\text{Cat}) \) as in (2a), i.e. the power set of the numeration set \( \text{Cat} \), where each member of \( \mathcal{P}(\text{Cat}) \) is a subset of \( \text{Cat} \) by definition. In that exposition in section 2.1, the head of a selected member of \( \mathcal{P}(\text{Cat}) \) (i.e. a selected subset of \( \text{Cat} \)) was chosen from the member of that subset. Thus, from that perspective as well, the head of each treelet must be a member of its dominance set.

\( \text{RC} \) (mnemonic for “reflexive containment”) in (3) is a binary relation between treelets, which is analogous to reflexive dominance, though unlike reflexive dominance, which is defined between tree-nodes in labeled tree representations, \( \text{RC} \) is defined between treelets. As is clear from our exposition of CSS around (2) in section 2.1, the containment relation between treelets is isomorphic to the containment between the dominance sets of those treelets, as shown in (5).

(5) a. Reflexive Containment, \( \text{RC} \): \( \forall x, y \in \text{CSS}, (\text{RC}(x,y) \Leftrightarrow (D_x \supseteq D_y)) \)

b. Immediate Containment, \( \text{IC} \): \( \forall x, y \in \text{CSS} \).

\[
(\text{RC}(x,y) \land x \neq y \land \neg \exists z \in \text{CSS}, (\text{RC}(x,z) \land \text{RC}(z,y) \land z \neq x \land z \neq y))
\]

The basic relation of our structural representation system is reflexive containment (RC) following the restrictions in (5a), whereas (5b) defines the derived relation of immediate containment (IC) which is used in the LF and PF interpretations and which is also useful for clarifying some points in this section. As we see below, given the basic definition of RC in (5a) together with its antisymmetry property given in (7c), it follows that each treelet \( x \) corresponds to exactly one dominance set \( D_x \), which is unique to \( x \). Crucially, since \( \text{RC} \) is isomorphic to the reflexive set containment between the dominance sets as in (5a) and since the set of all the dominance sets that appear in CSS is a proper subset of \( \mathcal{P}(\text{Cat}) \) (i.e. the power set of the numeration set \( \text{Cat} \)), once we choose a finite numeration set, both the upper bound of each structure and the maximal number of distinct structures that we can generate with the numeration set is automatically set without any further restrictions.\(^7\) This is computationally attractive.

Starting with this general reflexive set-containment, we attribute to RC basically the same relational properties of reflexive dominance as is used in standard tree structures. First, each CSS has a unique maximal treelet with regard to \( \text{RC} \), as shown in (6).

\(^7\) This is basically the same finiteness property as we explained in the text following (2) in section 2.1.
(6) Maximal treelet, \( \exists x \in \text{CSS}. \forall y \in \text{CSS}. \ R_C(x,y) \)

For each CSS, its reflexive containment \( R_C \) is a partial order, as in (7).

(7) \ a. Reflexitivity: \( \forall x \in \text{CSS}. \ R_C(x,x) \)

\[R_C(x,y) \land R_C(y,z) \rightarrow R_C(x,z)\]

\ b. Transitivity: \( \forall x,y,z \in \text{CSS} \)

\[R_C(x,y) \land R_C(y,z) \rightarrow R_C(x,z)\]

\ c. Antisymmetry: \( \forall x,y \in \text{CSS} \)

\[R_C(x,y) \land R_C(y,x) \rightarrow x = y\]

\ d. Corollary A: \( \forall x,y \in \text{CSS}. \ (D_x = D_y) \rightarrow x = y\)

(7a) and (7b) follow from (5a) but we explicitly specify these conditions here as well. Because of antisymmetry in (7c) together with the definition of \( R_C \) as in (5a), it follows that a CSS cannot contain two treelets that have the same dominance set but have different heads, as specified in the Corollary A in (7d).

We also assume the maximally binary-branching constraint in (8a). Just as it is debatable whether trees should be at most binary branching or not, the status of the maximally binary-branching constraint in our representation system is provisional, but it plays some non-trivial role when we define the PF structures and the interpretation of our syntactic structures as such PF structures.

(8) \ a. Maximally Binary Branching:

\[\forall x,y,z \in \text{CSS}, \ (\{x' \in \text{CSS} \mid R_C(x',x) \land x' \neq x\} = \{y' \in \text{CSS} \mid R_C(y',y) \land y' \neq y\}\]

\[\{z' \in \text{CSS} \mid R_C(z',z) \land z' \neq z\} \rightarrow (x = y) \lor (x = z) \lor (y = z)\]

\ b. Unique Splittability: \( \forall x,y \in \text{CSS} \)

\[\{(x' \in \text{CSS} \mid R_C(x',x) \land x' \neq x\} = \{y' \in \text{CSS} \mid R_C(y',y) \land y' \neq y\} \land x \neq y\]

\[\rightarrow (D_x \cup D_y = \emptyset)\]

\ c. Inclusiveness:

\[\forall x \in \text{CSS}, \forall a \in \text{Cat}. \ (a \in D_x \rightarrow \exists y \in \text{CSS}. (R_C(x,y) \land \text{head}(y) = a))\]

The unique splittability constraint in (8b) prevents one CSS from containing two treelets such as \( [d, [d,e]] \) and \( [c, [c,e]] \) in which a category \( e \) appears in the dominance sets of two distinct treelets that are not ordered with regard to \( R_C \). Together with the other conditions that we have introduced so far, it follows that if there exists a treelet \( z = [a, D_z] \) such that \( IC(z, x) \) and \( IC(z, y) \), then the dominance set \( D_z \) must be the union of \( D_x \) and \( D_y \) and \( [a] \).

Inclusiveness in (8c) prohibits us from generating a CSS as in (9a).

(9) \ a. Undesirable treelet:

\[\{\text{can}, \{\text{can, read, Meg}\}; \{\text{read}, \{\text{read, can}\}; \{\text{Meg, [Meg]}\}}\]

\ b. (9a) should be:

\[\{\text{can}, \{\text{can, read, Meg}\}; \{\text{read}, \{\text{read}\}; \{\text{Meg, [Meg]}\}}\]

In (9a), the treelet headed by \textit{read} contains the modal \textit{can} in its dominance set even though \textit{can} does not head any treelet that is reflexively contained by that treelet. Instead, \textit{can} heads the maximal treelet that contains the treelet headed by \textit{read}.
in (9a), which should not be the case. Intuitively, when we scan a CSS starting with
treelets of smaller sizes (i.e. treelets whose dominance sets contain fewer members,
starting with “identity treelets” \{a, \{a\}\}), then it should be the case that any element
\(a \in \text{Cat}\) be introduced as the head of a treelet initially, in the form \{a, \{a, \ldots\}\} where
“\ldots” might be empty (i.e. as in an identity treelet), before being incorporated into the
dominance sets of the containing treelets. Inclusiveness in (8c) requires exactly that
to be the case and then the correct CSS for (9a) is (9b).

As a derived property of our structure representation system, for each CSS, Closure
in (10) is automatically satisfied by Reflexivity in (7a).

(10) Closure (satisfied by (7a)):
\[(\forall x \in \text{CSS}. \exists y \in \text{CSS}. \text{RC}(x,y)) \& (\forall y \in \text{CSS}. \exists x \in \text{CSS}. \text{RC}(x,y))\]

Also, because of the presence of the unique maximal treelet in each CSS as required
by (6) together with Unique Splittability in (8b), we have Upward Non-Branching in
(11) as a derived theorem.\(^8\)

(11) Upward Non-Branching:
\[\forall x, y, y' \in \text{CSS}. ((\text{RC}(y, x) \& \text{RC}(y', x)) \rightarrow \text{RC}(y, y') \lor \text{RC}(y', y))\]

Each CSS is a set and so is each dominance set \(D_x\). Thus, we interpret \(n\) occurrences
of one item in each set as one, as shown in (12).\(^9\)

(12) Denotational interpretation of sets:
\[\forall a \in \text{Cat}. \{a, a, a\} = \{a\}\]
\[\forall a, b \in \text{Cat}. \{(a, \{a, b\}; \{a, \{a, b\}\})) = \{(a, \{a, b\}\}) \ldots \text{etc.}\]

Given the restriction in (12), a CSS can still contain more than one treelet for one
category, say, \(a \in \text{Cat}\), such as \{a, \{a, b, c\}\}, \{a, \{a, b\}\} and \{a, \{a\}\}, where \(\text{Cat} = \{a, b, c\}\). As
we show in some detail in the next section, (12) means that the proposed structure
representation system cannot distinguish some of the copy structures that PS trees
can represent as distinct. This inability to express syntactic copying is not stipulated,
it follows from the basic properties of our representation system.

This section has formally defined our structural representations as constituent
structure sets (CSS). Given a finite set of categories (= numerated items), each
CSS is a set of all constituents with distinguished members which we called heads.
CSS are less expressive than labeled trees since CSS are not equipped with the
notion of syntactic nodes. On the other hand, CSS are more expressive than
“telescope structures,” that is, the relational structure in which the basic syntactic

\(^8\) See Uchida and Bury (2008) for proof.
\(^9\) As we discuss in section 2.7, resorting to this property of sets on its own is not a distinguishing feature
of our proposal, though this property was used from the start of the development of our system in Bury
(2003).
relation (i.e. reflexive dominance) directly pairs items of the numeration set. In CSS, the basic relation (reflexive containment) is defined between certain subsets of the numeration set, not directly between items of the numeration set and this leads to an extra amount of expressive power in comparison to such telescope structures. The next section makes such comparisons by using example structures.

2.3 Representational collapsibility

This section compares the expressive power of the representation system defined in section 2.2 with the expressive power of labeled tree representations, especially in terms of categorial copying. We also indicate that CSS are more expressive than “telescope structures” as we described them in section 2.1.

As we have seen in section 2.2, unordered sets cannot distinguish multiple occurrences of a category from one occurrence, as in \{a, a\}=\{a\}. Because of this, CSS cannot distinguish certain structures that PS trees can (cf. Bury 2003). Compare (13) with (14).

\[(13)\]

\[
\begin{array}{cccccc}
\text{a.} & \text{b.} & \text{c.} & \text{d.} & \text{e.} & \text{f.} \\
V & V & V & V & V & V \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
V & D & V & D & V & D \\
& D & V & D & V & D \\
\end{array}
\]

Constituent structure sets in (14a)–(14e) represent the trees in (13a)–(13e).

\[(14)\]

\[
\begin{array}{c}
\text{a.} \{\{V, \{V\}\}\} \\
\text{b.} \{\{V, \{V, V\}\}; \{V, \{V\}\}\} = \{\{V, \{V\}\}; \{V, \{V\}\}\} = \{\{V, \{V\}\}\} = (14a) \\
\text{c.} \{\{V, \{V, V, D\}\}; \{V, \{V\}\}; \{D, \{D\}\}\} \\
= \{\{V, \{V, D\}\}; \{V, \{V\}\}; \{D, \{D\}\}\} \\
\text{d.} \{\{V, \{V, V, V, D, D\}\}; \{V, \{V, V, D\}\}; \{V, \{V\}\}; \{D, \{D\}\}; \{D, \{D\}\}\} \\
= \{\{V, \{V, D\}\}; \{V, \{V, D\}\}; \{V, \{V\}\}; \{D, \{D\}\}\} = (14c) \\
\text{e.} \{\{V, \{V, V, V, D\}\}; \{V, \{V, V, D\}\}; \{V, \{V\}\}; \{D, \{D\}\}\} = (14d) = (14c) \\
\end{array}
\]

The two tree structures in (13a,b) collapse into one CSS, as shown in (14a,b). Thus, copying of V is possible only if the higher copy has a filled specifier, as in (13c) which corresponds to the distinct CSS in (14c). Also, in CSS, we cannot fill this specifier position by copying a category from a lower position in the tree as in (13d). In CSS, (13d) is equivalent to (13c), as is shown in (14c,d). Moreover, as (14c–e) show, CSS cannot distinguish the multiple dominance structure in (13e) from the copy-chain structure in (13d) or from the non-movement structure in (13c).
(cf. Kracht (2001) shows that copy chains and multiple dominance structures are formally equivalent).\textsuperscript{10}

For comparison, consider “telescope structures” as we described them in section 2.1. Each telescope structure can be represented as a relational structure \(<Cat, RD>\) where the unique syntactic relation \(RD\) (which is a reflexive dominance relation) directly links the items of the numeration set \(Cat\). Now, with \(Cat = \{V, D\}\), such a system cannot distinguish \((13c)\) from \((13f)\), whereas our CSS can. If we maintain the common properties of reflexive dominance and if we assume that \(V\) is the maximal element, as is the case both for \((13c)\) and for \((13f)\), then the only possible telescope structure is \(S = \{<V,D>, <V,V>, <D,D>, \}\), where \(<V,D>\) means that \(V\) reflexively dominates \(D\). This structure corresponds to \((13f)\). It is debatable whether \(S\) also covers \((13c)\), since in \((13c)\), only the higher node labeled with \(V\) reflexively dominates \(D\) and not the lower one with \(V\). In either case, however, we cannot have a telescope structure for \((13c)\) distinct from \(S\), which again is for \((13f)\). That is, \((13c)\) is either unrepresentable or equivalent to \((13f)\) in this relational structure. Thus, CSS are more expressive than telescope structures, i.e. relational structures that directly link items in \(Cat\).

Turning now to linguistic applications of CSS, self-attachment, or reprojection, of a head (i.e. copying of \(V\) as in \((13c)\)) is possible with a filled specifier, whereas movement/copying into a “specifier position” is not expressible. Thus, for A/A-bar movement phenomena, we must resort either to a base generation analysis or use of distinct categories/lexical items that are related by way of the semantics. In section 2.5, we sketch an analysis of wh-extraction in CSS following the former strategy.

This section has shown that our representation system is much weaker in expressive power than labeled trees while our system is more expressive than “telescope structures.” The next section provides a range of linguistic data in which the restricted ability of CSS to duplicate an item in the numeration set leads to a more explanatory analysis. That is, the V2 pattern in German.

\subsection*{2.4 German V2 in CSS (Bury 2003, 2005)}

\subsubsection*{Data}
In German main clauses, the fronted finite verb (i.e. \(V\)-\text{fin}, in \((15)\)) is preceded by a single phrasal constituent XP:

\begin{equation}
(15) \text{XP V-fin, } \{(\text{Subject}) \ (\text{Object}) \ (\text{Adjuncts})\} \ (\text{V-nonfin}) \ t_i
\end{equation}

\textsuperscript{10} See Brody (2006) for a system that also cannot distinguish copy chain structures and multidominance structures from non-movement structures.
XP can be of any syntactic category and does not receive a uniform interpretation, as shown in (16).

(16)  
a. \([_{DP} \text{Max}] \backslash\text{backt dienstags Kuchen.}\)  
\text{\‘Max bakes cakes on Tuesdays.’}\n
b. \([_{PP} \text{Im Garten}] \backslash\text{singt ein Vogel.}\)  
\text{\‘A bird is singing in the garden.’}\n
c. \([_{Adv} \text{Manchmal}] \backslash\text{backt Max Kuchen. (XP = no topic, no focus)}\)  
\text{\‘Sometimes Max bakes cake.’}\n
d. \([_{DP} \text{Diesen Kuchen}] \backslash\text{will nicht einmal Max essen. (XP = topic)}\)  
\text{\‘Not even Max wants to eat this cake.’}\n
e. \([_{DP} \text{Nur Max}] \backslash\text{will keinen Kuchen essen. (XP = focus)}\)  
\text{\‘Only Max doesn’t want to eat cake.’}\n
The initial XPs in (16a–e) are interpreted differently, as suggested in the parentheses. As a result, feature checking analyses of V2 risk circularity (cf. Haider 1993). Fanselow’s (2004a) analysis illustrates the problem: He proposes that in one type of V2 clause, XP fronts to check an operator feature (e.g. wh, focus, or topic); in a second type, XP fronts to check an EPP feature on C. While the theoretical status of a generalized operator feature may be debatable, there simply doesn’t seem to be an independent motivation for the EPP feature that is postulated to account for the remaining V2 clauses.

2.4.2 Analysis in CSS

2.4.2.1 Verb movement  
Ackema et al. (1993) propose that a moved verb can project an additional VP, as in (17a) below. The structures we assume for V2 verb movement resemble projection-free versions of their self-attachment structures, i.e. (17b).

(17)  
a. \([_{VP} a \ [_{V'} V_1 \ [_{VP} b \ [_{V'} t_i \ c ]]]]\)  
b.  
\[
\begin{array}{c}
a \overrightarrow{V} \\
V \\
b \downarrow \quad \downarrow c
\end{array}
\]

2.4.2.2 Analysis of V2 in CSS  
(18a–c) are the relevant tree structures for explaining V2 phenomena. The corresponding CSS are in (19a–c), where (19c) will generate a V2 sentence, with the finite verb in the higher T corresponding to the pronunciation position.
In CSS, a tree structure where T is “remerged” without a specifier, that is, (18b), is undistinguishable from (18a), as shown in (19a–b). Thus, without a filled specifier, it leads to the same PF order, Sub-T-<..>. If however a structure contains a remerged T with an additional specifier, that is, the tree structure in (18c), then its CSS, that is, (19c), is distinct from the CSS for (18a), that is, (19a).

This means that a moved verb can only be pronounced in the PF position of a “remerged” category (i.e. the PF position of the higher T in (18c)) if it has a filled specifier. This specifier’s category and interpretation are irrelevant, as long as it isn’t empty. The basic V₂ pattern is thus derived from structural principles, without the introduction of any features that lack an independent motivation.11

Verb initial sentences seem to pose a problem for this analysis, but note that in CSS, a distinct category can be added on top, as shown in (20).

With the structure in (20), we may pronounce the verb in the position of a distinct functional head C, which can lead to the verb initial PF order.12 The choice of this category C is arbitrary here and note that since C is a different item from T in Cat, we do not need a “filled specifier” for C in order to represent (20a) as a distinct CSS.

We suggest that the insertion of a distinct functional head C on top of T is justified since V₁ in German is interpreted in a way that is distinct from the interpretation of

11 If we treat head movement as a PF operation (cf. Brody 2000, Boeckx and Stjepanović 2001) and distinguish it from syntactic copying (cf. Bury 2003), this analysis also becomes compatible with Chomsky’s (1995) suggestion that V₂ is a phonological phenomenon (see also Hock 1991).

12 See section 2.6 for the PF linearization.
common declarative sentences (for discussion, cf. Weerman 1989, Roberts and Roussou 2002). Use of inversion in the subordinate clause of conditionals or counterfactuals can be dealt with in a similar manner as we treat V₁.\(^{13}\)

What might appear problematic for this analysis are verb-initial languages, in which the main verb may appear at the top of a canonical declarative sentence. However, such languages often have preverbal particles and it can be argued that these preverbal particles correspond to a functional head higher than T, such as C in (20), and that the verb is pronounced in the position of this head (cf. Bury 2005).\(^{14}\)

In addition to the proposals about V₁/V₂ discussed here, the notion that verb movement can involve reprojecion structures has had a central role in the analysis of other empirical domains (cf. Neeleman and Weerman 1999, Janke and Neeleman 2009 on VP shells/double object constructions, and Koeneman 2010 on V-to-I movement). While these proposals differ in various ways from the one described here, we think that their results can be interpreted in terms of our model by shifting some of the work done in the syntax to a richer semantic component. A full discussion of how this could be achieved is beyond the scope of this chapter.

To summarize our treatment of V₂ phenomena, a self-attachment structure cannot be represented without a filled specifier (cf. Koeneman 1995). Since this is a purely structural requirement, this specifier position is compatible with any syntactic category and imposes no constraints on the interpretation of the category that fills it. This derives the correlation between verb movement and an obligatory XP that is characteristic of the V₂ pattern.

Whereas our CSS can express the duplication of the same category onto the mother node as long as the “specifier” position is filled, CSS cannot represent the copying of a category onto a terminal node under any circumstances. If such a syntactic copying is absolutely necessary in a syntactic theory, CSS are not usable. The basic test cases will be A-bar and A movement phenomena. In this chapter, we only provide a rough sketch of our analysis of A-bar movement in CSS.

\subsection*{2.5 Wh-extraction}

While CSS can represent self-attachment structures by having one item in the numeration set head more than one treelet, we cannot copy a category onto a terminal node. Consider (21).

\(^{13}\) This analysis will associate more than one category, say, C and T in (20), with one overtly realized verb. This requires a more relaxed relation between the numeration set Cat and the set of overt phonological items.

\(^{14}\) In our analysis, this is analogous to the situation in which a verbal head attaches itself to a higher functional head such as C in a common phrase structure tree.
The wh-copy-chain structure in (21a), where one copy of \textit{what} occurs in the specifier position of \textit{did} and one copy occurs as the complement of \textit{eat}, collapses to the wh-in-situ structure in (21b) when we represent them in CSS, as (22a) and (22b) show.\footnote{Based on a similar property of unordered sets, Brody (2006) also argues that phrasal movement should not be definable in syntax. Epstein and Seely (2006) argue, also on the basis of unordered sets, that chains are not syntactic objects.} In order to represent a wh-extraction structure as distinct, we have to either multiply the number of wh-expressions in the numeration set, say, \textit{what}_1 and \textit{what}_2 for one phonological word \textit{what}, or insert the wh-item only in the landing site, as in (21c), which is represented as a distinct CSS in (22c).

Having rejected telescope structures in order to avoid multiplying copies in the numeration set (cf. section 2.1), we adopt the structure in (22c). The question then is how to interpret this CSS in the semantics. In fact, interpreting (22c) does not pose a particular problem. Type Logical Grammar (TLG) treats wh-extraction in a similar way, that is, an extracted wh-expression is merged only in the landing site (cf. Moortgat 1997, Vermaat 2006). Since there is no semantic term in the in situ position, the corresponding argument selection of the local verbal functor is...
percolated until the wh-expression is merged.\textsuperscript{16} The percolated argument selection cannot be saturated in the same way as an in situ argument selection is and thus the only way of saturating the percolated argument slot is to merge an operator of the right type in an adequate operator position. We adopt basically the same mechanism, though we do not use a categorial calculus to implement the mechanism.

To see how this works, look at the CSS in (22c) again. In the semantics, we replace each item in this structure with the corresponding simply-typed lambda expression (cf. Carpenter 1997) via lexical assignment. We then successively compose the typed lambda terms via type compositionality. We start this interpretation process with the minimal treelets and then incorporate the outputs into the treelets that immediately contain them in a successive manner.\textsuperscript{17} The semantic output of the maximal treelet is the interpretation of that CSS. Here, we skip all the intermediate stages and consider the interpretation of the maximal treelet in (22c). Its semantics is \{\textit{Wh}'((\lambda x_0 . e_0 . do' ((\textit{eat}'x)e_{\textit{eva}'}) \textit{Wh}'))\}. The head of the treelet contains the output of this final stage, i.e. \(\textit{Wh}'((\lambda x_0 . e_0 . do' ((\textit{eat}'x)e_{\textit{eva}'})\textit{Wh}')\). The dominance set contains the lambda terms that are merged to produce this final output. Now, the semantic type of the wh-expression \(\textit{Wh}'\) is \(<<\Diamond e_0 , q_0 , w>>\) and the type of \(\lambda x_0 . e_0 . do' ((\textit{eat}'x)e_{\textit{eva}'})\) for \textit{did Eva eat} is \(<<\Diamond e_0 , q_0 >>\), where “\(q\)” is mnemonic for “questions” and “\(w\)” is mnemonic for “wh-questions.” First, the auxiliary term \(do'\) is of type \(<t_0 , q_0 \rangle\), which maps a type \(t_0\) expression to a Yes-No question expression (thus, \textit{Do you eat pizza} counts as a Yes-No question).\textsuperscript{18} With the treelet in (22c), what is crucial is the percolated internal argument selection of the verbal functor \textit{eat} whose lexical type is \(<e_0 , <t_0 , t>>\). We assume that a type \(e_0\) argument selection must be saturated in situ either by an argument of the corresponding type (i.e. type \(e_0\)) or by a “hypothetical” argument of type \(\Diamond e_0\).\textsuperscript{19} If a type \(e_0\) argument is provided, the type \(e_0\) argument

\textsuperscript{16} TLG is unique in that it derives the effect of the percolation of an argument requirement by way of deductive inferences over categorial formulas. If we ignore this deductive element, though, the percolation process itself is shared by other grammar formalisms that do not use syntactic movement or copying, such as GPSG (Gazdar et al. 1985), HPSG (Pollard and Sag 1994), and the proposal in Neeleman and van de Koot (2002).

\textsuperscript{17} See Uchida and Bury (2008) for details. The minimal treelets are the identity treelets in the form of \([a_0, [a_0]]\), where we replace each \(a_0\) with its semantic entry according to the lexical information. As a reviewer notes, this process refers to IC in (5b) in section 2.2, which is derivable from the basic RC in (5a). Each occurrence of a functor category such as \(V\) for a lexical verb as the head of a treelet corresponds to the saturation of exactly one argument requirement of the semantic entry for this functor category. Since our CSS cannot represent phrasal copying into terminal nodes which would cause the same sort of compositionality problems that copy theories of movement have (cf. Cormack and Smith 2001), semantic compositionality is straightforwardly maintained with CSS.

\textsuperscript{18} \textit{Do}-inversion is not only used in questions but we ignore such extra complications here.

\textsuperscript{19} Though the reader can ignore the type logical details, “\(\Diamond\)” is a unary type forming connective that forms a “residuated pair” with “\(\Box\).” The reason why modal operators are used is that the resultant type logical system is a modal logic (“\(\Diamond\)” and “\(\Box\)” are interpreted in the opposite directions in the accessibility relation though we omitted the directionality markers). The box operator is used to create islands for A-bar movement in a type logical inference. See Morrill (1994) and Vermaat (2006) for a TLG analysis of wh-islands. Percolation of the hypothetical argument “\(\Diamond e_0\)” can continue up to the point where the required
selection is saturated once and for all and there is no abstraction from that position. However, if a hypothetical argument saturates an argument slot, then it must be discharged at a later stage of the derivation. It can be discharged at any stage of the derivation after it saturates the argument selection in question. For example, if we let the meta-variable \( t \) represent the hypothetical argument of type \( \Diamond e \), then we can first derive the term \( do'(\text{eat}'t)\text{eva}' \) of type \( q \), and then discharge this hypothetical argument \( t \) via lambda abstraction, deriving the term \( \lambda x_{\Diamond e}.do'(\text{eat}'x)\text{eva}' \). However, unlike the initial internal argument selection of the functor \( \text{eat}' \), this percolated argument slot is type \( \Diamond e \), rather than the in situ type \( e \). We assume that a type \( \Diamond e \) argument selection cannot be saturated by a type \( e \) argument, and thus, we cannot saturate the percolated type \( \Diamond e \) argument requirement by inserting a normal type \( e \) argument in the extracted position. This explains the ungrammaticality of *Tennis did Meg play. We also assume that every hypothetical argument \( t \) of type \( \Diamond e \) and every type \( \Diamond e \) argument selection must disappear by the time the interpretation of the CSS becomes complete. Thus, the usual way of eliminating this \( \Diamond \) operator in the semantics is to abstract away from the type \( \Diamond e \) argument slot, producing an expression of type \( \langle \Diamond e,a \rangle \) and then merge the result with an operator of type \( \langle \langle \Diamond e,a \rangle,b \rangle \), where \( a \) and \( b \) are meta-variables for semantic types. In (22c), the wh-expression of type \( \langle \langle \Diamond e,q \rangle,w \rangle \) requires a type \( \langle \Diamond e,q \rangle \) argument, which we can get only after merging the auxiliary did. Thus, we abstract away from the type \( \Diamond e \) argument only after merging did. As we implied above, we can discharge a hypothetical argument \( t \) itself at any stage after its insertion. Thus, if the language in question is equipped with a wh-expression of the right type, then the language can have an in situ wh-expression, as in Japanese, though such an in situ operator needs to be licensed by a question particle at a later stage in the case of Japanese. The discussion of more complex wh-constructions is beyond the scope of this chapter.

We have shown that we can deal with the basic wh-extraction data without inserting any item in the in situ argument position. Some might argue that such an
analysis will have a problem explaining reconstruction with regard to reflexive binding. However, note that the in situ argument slot is not the only reconstruction site for reflexive binding.

(23) [Which aspect of himself,₁/₂, does Bob₁ think that Jack₂ likes t₁?]

Given this, we merge the lambda term for which aspect of himself in the position as it appears in the overt phonological string and then define the semantics of the operator which in such a way that it conjoins its nominal restriction term for aspect of himself of type <e,t> with any type <◊ e,t> sub-expression of the logical term for does Bob think that Jack likes via set intersection.²³ The basic idea is that which takes its nominal restriction of type <e,t> as its first argument and then the expression of type <◊ e,q> for does Bob think that Jack likes as its second argument (thus, the type of the logical operator which’ is <<e,t>,<<◊ e,q>,w>>, as shown in (24)).²⁴

(24) [[Which [picture of himself]<e,t,>] [ does Bob think that Jack likes]<◊ e,q>]]

However, when we merge the term for which picture of himself of type <<◊ e,q>>,w> with the term for does Bob think that Jack likes, we can conjoin the nominal restriction picture of himself of type <e,t> with one of several type <◊ e,t> sub-expressions of the logical term for does Bob think that Jack likes. Since the hypothetical argument t of type ◊ e can be abstracted away from at any stage of the semantic term composition, there are many locations in the semantic term for conjoining the nominal restriction functor.²⁵ The reflexive pronoun can then be bound by a different binder depending on where the nominal restriction term is inserted in the semantic term. Whether this semantic reconstruction analysis over-generates reconstruction sites is left for future research. For an analysis of (24) using a similar idea but in a different framework, see Uchida (2010).

This section has provided a sketch of our analysis of wh-extraction in CSS. The analysis depends on the percolation of the argument selection of the (verbal) functor during the composition of the lexically assigned typed lambda terms.

²³ In this section, we let phonological strings represent their logical terms. The semantic type of aspect of himself is actually more complex since we interpret reflexives as functors from (hypothetical) individuals to individuals. But we stick to type <e,t> in this chapter. See Uchida (2008) for a similar analysis of reflexives in TLG.

²⁴ In (24), we used phonological expressions to represent their semantic terms for convenience. “[A]ₐ” means that the semantic type of the expression A is a.

²⁵ Crucially, we can saturate both the internal and external argument selections of like with (different) hypothetical arguments and then abstract away from the internal argument slot. This temporarily creates an appropriate type <◊ e,t> expression with which we can conjoin aspect of himself. We can then bind himself with Jack. Note that the term for Jack can directly saturate the external type e argument slot of like or it can merge as type <<◊ e,t>,t> functor after abstracting away from the external hypothetical argument ◊ e of like. Both options lead to the same logical form via normalization.
2.6 PF linearization

This section sketches the PF linearization of our syntactic structures by way of an example. For a more systematic exposition of the PF linearization, see Uchida and Bury (2008).

When we linearize a CSS, we linearize each treelet, starting with the identity treelets of the form \( \{a,\{a\}\} \) and incorporating the output of each treelet into the dominance set of the treelet that immediately contains it. Let us show the process with an example. Consider (25).

(25) a. \( \text{Cat}_7 = \{\text{T}(\text{can}), \text{V}(\text{play}), \text{D}_1(\text{Meg}), \text{D}_2(\text{tennis})\} \)

b. \( \text{CSS}_7 = \{\{\text{T},\{\text{T},\text{V},\text{D}_1,\text{D}_2\}\}\}; \{\text{V},\{\text{V},\text{D}_2\}\}\}; \{\text{D}_1,\{\text{D}_1\}\}\}; \{\text{D}_2,\{\text{D}_2\}\}\} \)

First, we replace each category in the smallest treelets in \( \text{CSS}_7 \) with the corresponding phonological items. The PF output of an identity treelet \( \{a,\{a\}\} \) is the phonological word for \( a \). For example, the PF output for the treelet \( \{\text{D}_2,\{\text{D}_2\}\} \) is tennis. This output is incorporated into the dominance set of the treelet that immediately contains it, that is, \( \{\text{V},\{\text{V},\text{D}_2\}\} \). This is the smallest treelet in which V appears, so the PF word for V is inserted into the dominance set of this treelet as it is. Thus, the dominance set \( \{\text{V},\text{D}_2\} \) becomes \( \{\text{play}, \text{tennis}\} \). When the dominance set contains two units as in this case, we can potentially linearize them either as \( \text{(play-tennis)} \) or \( \text{(tennis-play)} \). Further constraints are assigned at the level of post-syntactic PF which choose the former PF for English (see Uchida and Bury 2008). This output unit \( \text{(play-tennis)} \) is incorporated into the treelet that immediately contains \( \{\text{V},\{\text{V}, \text{D}_2\}\} \), that is, the maximal treelet. Crucially, once \( \text{(play-tennis)} \) is generated, it counts as one PF unit in the next stage. Together with \( \text{Meg} \), which is the output of an identity treelet \( \{\text{D}_1,\{\text{D}_1\}\} \), the dominance set of the maximal treelet will then be \( \{\text{can}, \text{Meg}, \text{(play-tennis)}\} \). When the dominance set contains three PF units as in this case (this happens when the treelet immediately contains two treelets), then we can merge the three PF units \( a, b, c \) either as \( \text{(a-b-c)} \) or as \( \text{(c-b-a)} \), where \( b \) is the PF item for the head of the treelet in question. In the above case, this can generate either \( \text{(Meg-can-(play-tennis))} \) or \( \text{((play-tennis)-can-Meg)} \). Again, we can postulate some syntax external constraint to exclude the latter (cf. Uchida and Bury 2008).

The underlying PF generation strategy behind the above procedure is as in (26) (cf. Marantz 1989, Bury 2003).

(26) Immediate containment in the syntax is represented as linear adjacency in the phonological structure.

26 “·” represents a merge of phonological items. It is used for both binary and ternary merges.
Thus, if a treelet \( x \) immediately contains another treelet \( y \) and if the PF word for the head of \( x \) is \( a \) and the PF output for the treelet \( y \) is \( b \) (which can be complex, as we have seen above), then the PF output for \( x \) is either \( (a \cdot b) \) or \( (b \cdot a) \). If a treelet \( x \) immediately contains two treelets \( y \) and \( z \), and if the PF word for the head of \( x \) is \( a \) and the PF outputs for \( y \) and \( z \) are \( b \) and \( c \) (again, \( b \) and \( c \) can be complex), then the PF output for \( x \) is either \( (b \cdot a \cdot c) \) or \( (c \cdot a \cdot b) \). Some of these theoretical linearization possibilities are filtered out at the level of PF, or because of some PF-LF correspondence requirement which is outside the scope of the constraints on CSS and their formal interpretation rules. However, we believe that providing some flexibility of linearization might be useful for generating various argument orders in the PF without modifying the structural hierarchies among the arguments in the syntax (cf. Neeleman and Weerman 1999, Bury 2010).

2.7 CSS in comparison

Before concluding the chapter, we compare CSS to some other structure representation systems. We also consider our proposal in the background of Chomskyan linguistics.

2.7.1 Comparison to other structure representation systems

This subsection briefly discusses the differences of our proposal from other proposals made in theoretical linguistics. Firstly, it has been suggested that our proposal is similar to the dependency graphs of Dependency Grammar.\textsuperscript{27} In some sense, we could categorize such dependency graphs in the same group as the “telescope structure” that we defined in section 2.1. That is, just like telescope structures, common dependency graphs do not use the notion of syntactic nodes in an essential manner since word-level items are directly linked by dependency relations that are commonly represented as arrows. However, remember that in this chapter we have focused on the expressive powers of various structure representation systems, not the expressive powers of the grammatical theories that use such representation systems. Dependency graphs as a representation system are equipped with the notion of vertices and directed edges (i.e. “arrows” as commonly used in dependency graphs) that link the vertices (see Bröker 1998 and Dikovsky 2004 for formal definitions of dependency graphs). Thus, if we allow one label to decorate more than one vertex, dependency graphs can be at least as expressive as the labeled trees that we defined in section 2.1. As suggested above, we can follow common practice in Dependency Grammar and

\textsuperscript{27} Many thanks to Natasha Kurtonina, Richard Hudson, and an anonymous reviewer for suggesting this comparison (see also Carnie 2008). Dependency Grammar itself was introduced to modern linguistics in Tesnière (1959) and many versions of Dependency Grammar have been proposed since. See Kruijff (2006) for an overview.
identify the set of vertices in each dependency graph with the set of word level items. However, even then, there is no essential reason to assume that the vertices in a dependency graph must be partially ordered with a unique maximal vertex. In other words, it is possible to draw a dependency graph in which the vertices are not partially ordered with the unique maximal vertex. In contrast, it is not possible for the power set of the numeration set to stop being a partially ordered set with the unique maximal item by way of the set containment relation.

The richer syntactic relations that dependency graphs can represent are not only a representational possibility. Many dependency graphs used in linguistic analyses do not maintain a partial order with a unique maximal element. For example, transitivity of the dependency relation is not uniformly maintained. Moreover, though this is also theoretically optional, each dependency arrow is often annotated with a relational index so that each arrow can represent a different kind of dependency relation. For some linguists, two words may even be linked by two arrows with the opposite dependency directions, so that one word may depend on the other in one way while the latter word may depend on the former in another way.

From a slightly different perspective, note that linear adjacency in the phonological string is not necessarily respected in a dependency graph, whereas, as we saw in section 2.6, the reflexive containment relation in each CSS corresponds to linear adjacency in the phonological string. Some theorists derive a partial order with a unique root from each dependency graph in order to derive a phonological string (cf. Bröker 1998), but such derived tree structures are postulated separately from dependency graphs, and, also, derived tree structures are normally as expressive as labeled trees.

The greater expressive power of dependency graphs as a representation system and the manipulations of such expressivity in linguistic (and other) analyses are not surprising. The main utility of the dependency graph as a structure representation system is not its representational restrictiveness. It is quite the opposite; the main merit is that dependency graphs can potentially represent many kinds of dependency.

---

28 The unique maximal vertex must correspond to the unique maximal word-level item with the above-mentioned one-to-one correspondence between the vertices and the word-level items.

29 Remember that each CSS corresponds to a subset of the power set of the numeration set, where our partial order $RC$ is a set containment relation.

30 For example, Meľčuk (2003) defines dependency as an “anti-transitive” relation. See also Hudson (2007) and Bröker (1998), among others. In computer science and electrical engineering, the dependency relation is often assumed to be transitive, but some other properties of the partial order that we adopted in sections 2.1–2 are often thrown away instead, such as upward non-branching or the presence of a unique maximal vertex. See Balamas (2004), for example.

31 For example, Hudson (2007: 142) suggests a mutual dependency relation between wh-pronouns and verbs, instead of a one-way dependency.

32 Moreover, we can derive phonological strings directly from dependency graphs without going via such tree structures, as in Richard Hudson’s theory.
relations between many kinds of items within one dependency graph. It is possible to restrict the great expressive power of dependency graphs through the introduction of additional theoretical constraints, a possibility that we also discussed with regard to labeled trees in section 2.1. However, as we emphasized above, our proposal goes in the opposite direction. We propose a more restrictive structure representation system that automatically excludes many of the theoretical options that are available with alternative structure representation systems.

Next, we briefly compare CSS to another representation system. Within generative linguistics, there have been proposals to represent syntactic structures using sets (see e.g. Chomsky 1995a: 397, 1995b, 2004, 2005, Langendoen 2003, Brody 2006). However, as far as we can see, none of these proposals has defined each syntactic structure as a partially ordered set of subsets of the numeration set as we have done in this chapter. Note that we implemented both labeled trees and telescope structures as partially ordered sets, as well as CSS. However, the three systems are still provably different from one another in their expressive powers. Though a particular property of sets is crucial for us, what distinguishes our system is not the use of sets on its own. Our representation system is different from others in terms of what elements enter into the basic syntactic relation and what kind of relational properties the basic syntactic relation has.

2.7.2 CSS in the context of Chomskyan linguistics

In this subsection, we focus on one major issue in the development of Chomskyan syntactic theory, namely the role of phrasal categories in syntactic structures. Comparing our system with others in this regard is not meaningful from the viewpoint that we have adopted in this chapter since our proposal is based on the algebraic properties of the relational structure that we have proposed and we have compared our system to alternative structure representation systems, such as labeled trees and telescope structures, with regard to the implementations of those alternatives as relational structures, not with regard to the original theories associated with those alternative structures, such as Government and Binding theory, Minimalism, or Brody’s mirror theory. On the other hand, discussing our proposal with a view to the role of phrasal categories in syntactic structures may help place our proposal in the context of Chomskyan syntactic theories.

The role of phrasal categories is of interest because it has undergone a major shift in the history of phrase structure-based syntax, which, as Chametzky (2000: 6) and Hudson (2007), among others, argue, makes modern versions of phrase

---

33 “One of the attractions of dependency theory as a model of semantic representation is that it seems to offer the possibility of semantic representations in terms of cognitive networks, which are more flexible than the strictly hierarchical organisations implied by constituent structure” (Hudson 1980: 196).

34 We discussed this property at (12) in section 2.2.
structure-based grammars more similar to Dependency Grammar. In particular, since Chomsky’s (1957) original formalization of PSGs, the heads of phrases, i.e. word-level categories, have become increasingly important as the Chomskyan generative paradigm evolved (see Bresnan 1970 for an important early paper; see Bury 2003 for discussion). As a result of this development, phrase-level categories have gradually lost their significance, until they eventually disappeared altogether in the Bare Phrase Structure system where they were replaced by a new system of labels that encodes the derivational history of a constituent (Chomsky 1995b).

One proposal that arguably goes further in eliminating a dedicated labeling system for complex constituents is that of Hornstein (2009) (see also Hornstein and Nunes 2008). Hornstein proposes that phrase structures are built by two operations: concatenation, which concatenates two atomic items, and label, which applies to the output of concatenation and “yields a new complex atomic unit [...] whose content is no longer available for further concatenation” (Hornstein and Nunes 2008: 66). This model differs both from Chomsky’s (1995b) and from older PS models in that “the label of a derived structure is one of the atoms that concatenated” (Hornstein 2009: 59). Thus, unlike in Chomsky’s system, here the label is always an atomic, simple category. This means that in this model, unlike in traditional PS systems but similar to the system of Brody (2000), structures can contain lexical items that dominate other lexical items. However, unlike Brody’s telescope structures (and unlike CSS), Hornstein’s model appears to make essential use of the concept of tree nodes. This can be seen in the treatment of adjuncts. While typically nodes are labeled, labeling is not always forced for structures involving adjuncts. That is, adjunction can take place, i.e. an adjunct can be concatenated with some existing object, and the output of adjunction need not be labeled.

Collins (2002) develops the Bare Phrase Structure system in a different way. Like Chomsky, Collins assumes an operation Merge that combines two linguistic elements and forms a constituent. Unlike Chomsky (and Hornstein and Nunes), Collins assumes that the output of this operation is not labeled. Thus Collins’ model does not make use of labeled nodes, and it does not refer to phrasal categories. However, the structures that his Merge operation generates have more or less the same shape as conventional minimalist (or PS) trees. Moreover, the theory makes crucial use of these structures, for example, to calculate minimality. Therefore, it seems to us that Collins’ theory still makes essential use of tree nodes.

This brief discussion then suggests that the work of Collins and of Hornstein and Nunes continues the shift away from phrasal category labels that has occurred in Chomskyan syntax, a trend that arguably reached its logical endpoint with Brody’s (2000) abandonment of the concept of tree nodes. While the proposals of Collins and Hornstein (and Nunes) are similar in this respect to Dependency Grammar, Brody’s model, and the implementation of CSS in Bury (2003), they differ from the latter in that they make essential use of tree nodes. We hope that by now it is clear to the
reader that both telescope structures and dependency graphs are algebraically quite
distinct from our system (where telescope structures are closer to ours as the reader
can see from the discussion in section 2.1).

As we indicated above, whenever we distinguish our system from alternatives, we
implement alternatives as relational structures. Some might argue that this is not
fair since few of these alternative structures were proposed as algebraic structures.
But this criticism misses the main point of our proposal, which we emphasize again
here. Generative syntacticians do not always take seriously the expressive power of
the representation system that they use (say, labeled trees). Such representation
systems are often simply employed as convenient tools for theoretical descriptions
and there is an assumption that the syntactic theory as a whole (such as a generative
grammar using features) will generate the adequate subset of the structures that their
chosen system can potentially represent. As noted in the introduction, such an
approach generally increases the computational complexity of the theory.

2.8 Conclusion

In this chapter, we have proposed a new structure representation system with weak
expressive power. Adopting this system reduces the expressive power of the grammar
at the foundational level. In the proposed structure representation system, we order
subsets of the numeration set via the basic syntactic relation, rather than directly
ordering the members of the numeration set as in “telescope structures” (cf. Brody
2000) or ordering tree nodes decorated by the members of the numeration set, as in
phrase structure trees. Our system maintains the basic merit of telescope structures.
That is, if the numeration set is finite, then each structure that we can generate is
automatically finite and we can generate only a finite number of distinct structures
without further stipulation. On the other hand, unlike telescope structures, our
system can still represent syntactic copying with a provable maximal limit, which
we argue is linguistically useful. Whether we can cover various crucial data using this
representation system without compromising its main merits is left for future
investigation. However, we have shown that German V2 data can be captured in
an elegant manner with this system. We have also shown that inability of our system
to represent a copy chain or a multidominance structure for phrasal movement does
not provide a serious problem for our analysis of wh-extraction. Finally, we consid-
ered the relation of our system to other approaches, notably the structures assumed
in Dependency Grammar and recent proposals in Chomskyan syntax.
A Parallel Merge Solution to the Merchant/Johnson Paradox*

BARBARA CITKO

3.1 Parallel Merge: a new way of structure building

Since Chomsky’s (2004) Beyond Explanatory Adequacy, it is standardly assumed that the basic operation responsible for structure building in minimalist syntax, Merge, comes in two guises: External Merge and Internal Merge. External Merge, illustrated in (1), takes two objects and combines them into one bigger object.

\[ \alpha \beta \rightarrow \gamma \quad \alpha \beta \]

Internal Merge also takes two objects and forms a new object, but this time, one of these two objects is part of the other. The basic mechanism is the same; the only difference is that External Merge applies to two disjoint objects and Internal Merge applies to two objects that stand in a subset relationship. In (2), for example, \( \beta \) is a part of \( \alpha \) even before Internal Merge has applied.1

* I would like to thank Myriam Uribe-Etxebarria and Vidal Valmala, two anonymous reviewers, and the audience at the Ways of Structure Building conference for many useful suggestions and comments, which led to many improvements in the chapter. I alone am responsible for any remaining errors and omissions.

1 The structure in (2) represents Internal Merge as a Copy and Merge operation. This is the convention I will stick to throughout this chapter, mostly for the ease of exposition, without assigning any theoretical significance to it. Representing movement in terms of multidominance (see, for example, de Vries (this volume) and Gračanin-Yuksek (this volume), among others, for such representations) is a more accurate reflection of the term Internal Merge.
In Citko (2005), I argued that the existence of External Merge and Internal Merge predicts the existence of a third kind of Merge, which I dubbed Parallel Merge.2 Parallel Merge, which could thus be thought of as a new way of structure building, is the mechanism responsible for creating multidominant structures. It combines the properties of External and Internal Merge; it takes two distinct objects (which makes it like External Merge) and combines one with a part of the other (which makes it like Internal Merge), as shown in (3).3,4


---

2 See also Citko (2000) and (2003) for an early discussion of Parallel Merge, where I introduced the mechanism but did not provide much conceptual motivation for it (nor derived its properties from independently motivated principles).

3 Parallel Merge is by no means the only mechanism that can generate multidominant structures. Others include (but are not limited to) grafting of van Riemsdijk (1998, 2000, 2006a, and 2006b) or node contraction in Tree Adjoining Grammar (see Chen-Main 2006). I refer the interested reader to Citko (2011a) and Citko (2011b) for a more complete introduction to (and an overview of) multidominance in generative grammar. Furthermore, if movement is thought of as Internal Merge/Remerge, sideward movement and Parallel Merge arguably become indistinguishable. I thank an anonymous reviewer for bringing these possibilities to my attention.

4 The idea that the mechanism responsible for generating multidominant structures combines the properties of more standard structure building operations is also behind the proposals of van Riemsdijk (2006b) and de Vries (this volume). And the very idea that multidominant representations exist goes back at least to Sampson (1975) and Williams (1978).
with conjoined wh-pronouns (Gračanin-Yuksek 2007, this volume, Citko (to appear), Citko and Gračanin-Yuksek (in progress), Ratiu (to appear)), serial verb constructions (Hiraiwa and Bodomo 2008a and 2008b), free relatives (Haider 1988, Citko 1998, 2000, 2011, van Riemsdijk 1998, 2000, 2006a, 2006b), parasitic gaps (Kasai 2007), amalgams (de Vries 2009b), parentheticals (McCawley 1982, de Vries 2007b, this volume), nominalizations (Fábregas, this volume). My humble goal in this chapter is to present one argument in favor of analyzing gapping in a multidominant way. The evidence comes from a paradox arising from a combination of Merchant’s (2008) account of the ungrammaticality of voice mismatches in pseudogapping (as opposed to VP ellipsis) with Johnson’s (2004, 2009) account of gapping as ATB movement (rather than ellipsis). Merchant (2008) uses the ungrammaticality of voice mismatches in pseudogapping to argue that pseudogapping is vP (rather than VP) ellipsis. Since gapping patterns with pseudogapping in that it also disallows voice mismatches, we would expect it to involve vP ellipsis as well. Such an analysis, however, is incompatible with Johnson’s evidence against treating gapping as ellipsis. My goal in this chapter is to show that a Parallel Merge approach to gapping, in which the gapped string is shared between two conjuncts, can solve this paradox. I proceed as follows. In section 3.2, I present the main facts to be analyzed in subsequent sections. In section 3.3, I review the so-called small conjunct analysis of gapping proposed by Johnson (2000, 2004, 2009) and show that without extra assumptions, it cannot account for the ungrammaticality of voice mismatches in gapping. In section 3.4, I show how a multidominant account can handle it, and in section 3.5, I conclude the chapter.

3.2 Voice mismatches in VP ellipsis, pseudogapping, and gapping constructions

The three constructions that are at the core of this chapter are gapping, pseudogapping, and VP ellipsis (VPE), illustrated in (4a–c), respectively. Throughout the paper, I indicate elided elements with a strikethrough.

---

5 This list of multidominant constructions (and references) is not meant to be exhaustive; its purpose is just to show that multidomiance is quite ubiquitous in the grammar. In Citko (2011b), I provide both empirical and theoretical arguments in favor of analyzing ATB wh-questions, Right-Node-Raising, free relatives (standard and transparent alike), gapping, and serial verbs in terms of Parallel Merge. Here, I focus on one argument in favor of a multidominant structure for gapping.

6 The idea that gapping involves multidomiance is not new; see Citko (2011b), Goodall (1987), Kasai (2007), Moltmann (1992), and Muadz (1991), among others. In Citko (2006a), which deals with the interaction of ATB and left branch extraction, I briefly suggest that gapping might involve a Parallel Merge structure, without going into any details. The multidomiant analysis of gapping is also implicit in Citko (2006b), which deals with determiner sharing. Due to space considerations I will not be able to compare the account developed here with other multidomiant accounts. Nor will I be able to offer an adequate comparison of multidomiant approaches to gapping with non-multidomiant approaches, which tend to involve ellipsis, ATB movement, or copying (see Repp (2008) for a recent overview).
Merchant (2008) notes that pseudogapping differs from VP ellipsis in that it does not tolerate voice mismatches, as shown by the contrasts given in (5) and (6) below. The examples in (5a–b) involve VP ellipsis; in (5a) the elided verb is active and its antecedent passive, whereas in (5b) the elided verb is passive and its antecedent active. Both are grammatical. The ungrammatical examples in (6a–b), on the other hand, involve pseudogapping; in (6a) the elided verb is passive and its antecedent active, and in (6b) the opposite is the case.

(5) a. **VP ellipsis: passive antecedent, active ellipsis**

This problem was to have been looked into, but obviously nobody did **look into this problem**.

b. **VP ellipsis: active antecedent, passive ellipsis**

The janitor must remove the trash whenever it is apparent that it should be **removed**.

(Merchant 2008: 169)

(6) a. **Pseudogapping: active antecedent, passive ellipsis**

* Some brought roses, and lilies were **brought** by others.

b. **Pseudogapping: passive antecedent, active ellipsis**

* Roses were brought by some, and others did **bring** lilies.

(Merchant 2008: 170)

Merchant accounts for this contrast by proposing that the ellipsis sites are different in VP ellipsis and pseudogapping. In both cases, ellipsis targets some projection of the verb, is licensed by the E feature (ellipsis feature) on the head whose complement is to be deleted, and obeys some form of syntactic identity. Not surprisingly, in VP ellipsis cases, what is elided is a VP. Voice information, however, is part of a higher head, v head (or voice head of Kratzer (1994)). Since this voice head is external to VP, as shown in (7) below, it is going to end up outside the ellipsis site and

---

7 In this respect, Merchant differs from Coppock (2001) and Johnson (2004), who deem voice mismatches in pseudogapping to be marginally acceptable (see also Tanaka 2011):

(i) ?That should be explained to individual students by the TA, but the professor will to the class in general. (Coppock 2001, ex. (4c))

(ii) ?The budget cuts might be defended publicly by the chancellor, but surely she wouldn’t her labor policies. (Johnson 2004: 31)

8 Merchant attributes example (5a) to Kehler (2002). The differences in voice matching between VP ellipsis, pseudogapping, and gapping are also discussed in Johnson (2004).
thus escape the identity requirement. This explains why voice mismatches are allowed in VP ellipsis.\(^9\)

\[ (7) \quad \text{TP (=structure of (5a))} \]

\[
\begin{array}{c}
\text{TP} \\
\text{but} \\
\text{TP} \\
\text{T'} \quad \text{T'} \\
\text{T} \quad \text{vP} \quad \text{T} \quad \text{vP} \\
\text{v[voi:pass]} \text{VP} \quad \text{nobody} \quad \text{v'} \\
\text{look_into DP} \quad \text{v[E] [voi:act] <VP_E>} \\
\text{this problem} \quad \text{look_into DP} \\
\text{this problem} \\
\end{array}
\]

Pseudogapping, by contrast, involves deletion of a \( vP \), which means that the two \( v \) heads have to be identical. Otherwise, the identity requirement on ellipsis would be violated. This is what explains why voice mismatches are not allowed in pseudogapping. In (8), which is the structure of the ungrammatical example in (6b) above, the deleted \( v \) is active whereas its antecedent is passive.

\[ (8) \quad \text{TP (=structure of (6b))} \]

\[
\begin{array}{c}
\text{TP} \\
\text{and} \\
\text{TP} \\
\text{rooses}_1 \quad \text{T'} \quad \text{others}_2 \quad \text{T'} \\
\text{were} \quad \text{vP} \quad \text{did} \quad \text{X_{[foc]P}} \\
\text{vP} \quad \text{PP} \quad \text{lilies}_3 \quad \text{X'} \\
\text{v[voi:pass]} \text{VP} \quad \text{by some} \\
\text{brought} \quad t_1 \quad \text{X_{[foc][E]} <vP_E>} \\
\text{t}_2 \quad \text{v'} \\
\text{v[E] [voi:act]} \text{VP} \quad \text{t}_3 \\
\text{bring} \\
\end{array}
\]

\(^9\) The ternary branching structure for coordination is used here solely for the ease of exposition.
With Merchant’s analysis of voice mismatches in VP ellipsis and pseudogapping as background, let us turn to gapping. The ungrammaticality of the examples given in (9a–b) below shows that gapping patterns with pseudogapping in that it also disallows voice mismatches:

(9)  a. **Gapping: passive antecedent, active ellipsis**
    * Roses were brought by some, and others brought lilies.

    b. **Gapping: active antecedent, passive ellipsis**
    * Some brought roses, and lilies were brought by others.

The fact that both gapping and pseudogapping disallow voice mismatches would receive a straightforward account if gapping were also analyzed as involving vP (rather than VP) ellipsis, along the lines schematized in (10).

(10) TP (=structure of (9a))

However, this is precisely the kind of analysis of gapping that Johnson (2000, 2004, 2009) argues against. We thus encounter a paradox if we try to reconcile Merchant’s analysis of the ungrammaticality of voice mismatches in pseudogapping with Johnson’s arguments against treating gapping as ellipsis. In the next section, I discuss Johnson’s proposal in more detail, before turning to a multidominant alternative.

---

10 Merchant (2008:170) notes that gapping (as well sluicing, fragment answers, and stripping) patterns with pseudogapping with respect to voice mismatches, but does not provide an analysis of these constructions, or discuss the implications of his analysis for them.
in section 3.4, where I show that there is a way to analyze gapping in a way that captures Johnson’s structural insights without sacrificing Merchant’s account of voice mismatches.

3.3 Small conjunct account

3.3.1 Gapping as ellipsis?

Johnson questions two core aspects of the ellipsis analysis of gapping schematized in (10) above: the level of coordination and the actual process of deletion. Many of his arguments against treating gapping as ellipsis involve the differences between gapping on the one hand, and pseudogapping and VP ellipsis on the other. Superficially, all three look alike, and one might be tempted to analyze all three as involving coordination of two clauses, followed by deletion of a VP in one of them. In such an account, pseudogapping and gapping differ from VP ellipsis only with respect to minor details, such as movement of the object out of the VP inside the second conjunct (or the absence thereof), and deletion of the lower versus the higher VP, as shown in (11–13).11,12

11 Only the structures in (12b) and (13b) are from Johnson (2009).
12 Coppock (2001) gives three arguments in favor of a deletion account, involving the parallelism between VP ellipsis and gapping with respect to scope, the distribution of strict and sloppy readings, and the availability of split antecedents. However, Johnson (2009) shows that Coppock’s data are not a fatal blow to non-deletion accounts of gapping.
3.3 Small conjunct account

(12) a. Some have served mussels to Sue and others have served swordfish to Sue.

b. Some have served mussels to Sue and others have served swordfish to Sue.

deleted VP

(13) a. Some have served mussels to Sue and others have served swordfish to Sue.

b. Some have served mussels to Sue and others have served swordfish to Sue.

deleted VP

(cf. Johnson 2009:290)

However, such an analysis leaves unaccounted for a number of differences between VP ellipsis, gapping, and pseudogapping. These differences have been discussed quite extensively in the relevant literature (see Coppock (2001), Hankamer and Sag
(1976), Lobeck (1995), Jackendoff (1971), Johnson (2000, 2004, 2009), Sag (1976), among many others) and are summarized in the table given in (14) below.\footnote{The table (and the discussion below it) draws primarily on Coppock (2001) and Johnson (2004). I focus here on the differences between VP ellipsis and gapping (and refer the reader to these works for a discussion of the differences between gapping and pseudogapping).}

<table>
<thead>
<tr>
<th></th>
<th>Gapping</th>
<th>VP Ellipsis</th>
<th>Pseudogapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compatibility with non-linguistic antecedents</td>
<td>−</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Active/Passive voice mismatches</td>
<td>−</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Backward deletion</td>
<td>−</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Grammaticality in subordinate clauses</td>
<td>−</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Compatibility with subordinating conjunctions</td>
<td>−</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Grammaticality of left branch extraction</td>
<td>+</td>
<td>NA</td>
<td>−</td>
</tr>
</tbody>
</table>

First, gapping differs from VP ellipsis in that it is incompatible with non-linguistic antecedents. In the contexts given in square brackets, VP ellipsis is possible, as shown in (15a), but gapping is not, as shown in (15b).\footnote{A related difference, brought to my attention by one of the reviewers, concerns the fact that the elided string can be a combination of two different VPs in VP ellipsis cases but not in gapping cases.}

(15) a. [Hankamer brandishes a cleaver, advances on Sag] 
    Sag: Don’t! My god, don’t [VP__]! (Hankamer and Sag 1976: 409)

b. [Hankamer produces an orange, proceeds to peel it, and just as Sag produces an apple, says:] 
    # And Ivan peels an apple. (Hankamer and Sag 1976: 410)

Second, gapping, unlike VP ellipsis, disallows mismatches in voice, as we have seen above (see the ungrammatical gapping examples in (9a–b) above versus the grammatical VP ellipsis examples in (5a–b) above). Third, VP ellipsis can operate in a backwards fashion but gapping cannot:

(16) a. Because Sue didn’t \textit{eat} meat, John ate meat. 

b. * Sue \textit{ate} meat and John ate fish. (Lobeck 1995: 22)
Fourth, VP ellipsis is possible in subordinate clauses introduced by subordinators such as although or because, but gapping is not.¹⁵,¹⁶

(17) a. John will have caviar, although others won’t have caviar.
   b. * John will have caviar, although others had beans.

(18) a. Some had eaten mussels because others had eaten mussels, too.
   b. * Some had eaten mussels because others had eaten shrimp.
   (Johnson 2009: 293)

And fifth, VP ellipsis is possible in embedded clauses, but gapping is not:

(19) a. Some had eaten mussels and she claims that others had eaten mussels, too.
   b. * Some had eaten mussels and she claims that others had eaten shrimp.
   (Johnson 2009: 293)

Another host of Johnson’s arguments against treating gapping as ellipsis involves the level of coordination. The structure for gapping given in (12b) above involves coordination of full clauses. The evidence against such a structure, due to Siegel (1984), comes from scope considerations. Consider the gapped example in (20a). Siegel observes that it differs from its non-gapped counterpart in (20b) in that it allows wide scope interpretation for the modal can’t. This interpretation is paraphrased in (20c). This difference in interpretation is surprising on the ellipsis account, in which (20b) is the source for ellipsis in (20a).¹⁷

(20) a. Ward can’t eat caviar and Sue, beans.
   b. Ward can’t eat caviar, and Sue can’t eat beans.
   c. It is not possible (or desirable) for Ward to eat caviar and for Sue (simultaneously) to eat (merely) beans. (Siegel 1984: 524)

¹⁵ Only the b examples in (17) and (18) are from Johnson 2009, who contrasts gapping and pseudogapping (rather than VP ellipsis) with respect to the same properties.

¹⁶ One of the anonymous reviewers brings to my attention the fact that for many speakers gapping is also possible in except phrases:

(i) Nobody here reads any newspaper, except Mary reads the New York Times.
Comparatives are another non-coordinate construction that allows gapping:

(ii) Mary read more books than Bill read articles.

Jackendoff (1971), however, notes that deletion in comparatives is freer than deletion found in gapping.

¹⁷ There is some variation regarding the acceptability of negation in gapping, discussed recently by Repp (2008). This variation, while interesting in itself, does not have any direct bearing on the issues pursued in this chapter.
The interaction of negation with the disjunction or points toward the same conclusion. Example (21a) is interpreted as a conjunction of two negative statements (paraphrased in (21b)) rather than negation of a disjunction of two positive statements (paraphrased in (21c)). This is due to the equivalence given in (22), known as De Morgan’s Law. This logical equivalence can only hold if negation has wide scope over both conjuncts in a gapped structure.

(21) a. Bob can’t play checkers or Mary play chess.
    b. Bob can’t play checkers and Mary can’t play chess.
    c. Bob can’t play checkers or Mary can’t play chess. (Lin 2000: 277)

(22) \( \text{NOT (A OR B)} = (\text{NOT A}) \& (\text{NOT B}) \)

A related argument, according to McCawley (1993), comes from variable binding. As shown by the ungrammaticality of (23a) and (24a), variable binding across sentential conjuncts is impossible. Interestingly, it becomes possible if the verb inside the second conjunct is gapped, as shown in (24a) and (24b).

(23) a. * No one’s duck was moist enough or/and his mussels were tender enough.
    b. No one’s duck was moist enough or his mussels were tender enough. (McCawley 1993: 248)

(24) a. * No boy joined the navy and his mother headed the army.
    b. No boy joined the navy and his mother join the army. (Johnson 2000: 60)

In this respect, gapping also differs from pseudogapping, which disallows variable binding across conjuncts:

(25) * No woman can join the army and but her girlfriend will join the navy.

This is expected given the analysis of pseudogapping that assimilates it to VP ellipsis, such as the one given in (13b) above.

3.3.2 Small conjunct account of gapping

Johnson’s alternative, referred to as a small conjunct account, is illustrated in (26a–b). It differs from its predecessors in two respects. First, coordination involves vPs not TPs; hence the term “small conjunct” account. Second, the gapped element undergoes ATB movement out of the coordinate vP complex to a functional head above the coordination level, which Johnson takes to be a Pred head. The pre-movement structure is given in (26b); the verbs undergo ATB movement to Pred, and the
subject of the first conjunct moves to [Spec,TP], as shown in (26c). What is interesting about Johnson’s analysis of gapping is that there is no ellipsis whatsoever in it; the illusion of ellipsis is created by the interplay of two factors: low level of coordination and ATB movement.

(26) a. Some brought roses and others brought lilies.
   b. \[
   \begin{array}{c}
   \text{TP} \\
   \overbrace{\text{T}'} \quad \text{PredP} \\
   \overbrace{\text{T} \quad \text{Pred} \quad \text{vP}} \\
   \overbrace{\text{vP} \quad \text{and} \quad \text{vP}} \\
   \overbrace{\text{some} \quad \text{v'} \quad \text{others} \quad \text{v'}} \\
   \overbrace{\text{v \quad VP} \quad \text{v \quad VP}} \\
   \overbrace{\text{brought \quad roses \quad brought \quad lilies}} \\
   \end{array}
   \]
   c. \[
   \begin{array}{c}
   \text{TP} \\
   \overbrace{\text{some}_j \quad \text{T}'} \quad \text{PredP} \\
   \overbrace{\text{brought}_i \quad \text{vP}} \\
   \overbrace{\text{vP} \quad \text{and} \quad \text{vP}} \\
   \overbrace{\text{t}_j \quad \text{v'} \quad \text{others} \quad \text{v'}} \\
   \overbrace{\text{v \quad VP} \quad \text{v \quad VP}} \\
   \overbrace{\text{t}_i \quad \text{roses} \quad \text{t}_i \quad \text{lilies}} \\
   \end{array}
   \]
The structure in (26c) is somewhat of an oversimplification; Johnson argues that instead of the verbs undergoing ATB movement to Pred, the verb phrases undergo remnant ATB movement to the specifier of PredP. In order for this to be possible, the non-gapped material inside each conjunct has to vacate the verb phrase first. For Johnson (2009), this involves rightward adjunction of the object to VP, and remnant ATB movement of the two VPs. In (27b), I assume that vPs (rather than VPs) undergo ATB movement to the specifier of PredP. Furthermore, each conjunct contains an extra projection that attracts non-gapped elements to its specifier. This projection “houses” discourse features such as [Focus], [Contrast], [Topic] or [Ground].18,19 What this means is that in the case of (27a), the object of the first conjunct, and both the subject and the object of the second conjunct move to the specifiers of FP and the entire vP undergoes remnant movement to the specifier of PredP, as shown in (27b).

(27) a. Some brought roses and others brought lilies.

b. TP
   some j
   T’
   T  PredP
      vP i
      t j v brought t k
      Pred’
      FP
      and
      FP
      roses k
      F’ others j
      FP
      F
      t i
      lilies k
      t i

The small conjunct account has two obvious advantages over the ellipsis account.20 First, it does not treat gapping as analogous to VP ellipsis or

---

18 This is closer in spirit to Johnson (2004), where the remnants also move to the specifiers of conjunct internal heads. I do not think anything substantial hinges on the choice between these two possibilities.

19 For lack of a better term, I dub this projection FP, which is not to be equated with a focus phrase. The FP can house elements with various discourse features (not just foci), in which respect it resembles ΔP proposed by Uriagereka (1995), for whom Δ refers to discourse relatedness. Lambova (2002) adopts his proposal to explain the distribution of topics and foci in Bulgarian, and Galperina-Radu (2009) uses it to account for the properties of discontinuous noun phrases in Russian.

20 One question that such a structure might raise involves the ability of the second conjunct subject to get its case features checked (or valued, to use current terms). This could be a problem in a system that...
pseudogapping; thus, it does not predict that the three constructions should behave alike. Second, it can explain scope and variable binding facts discussed above. The modal in (20a) and (21a) has scope over both conjuncts because it is merged above both of them (as opposed to being present in both and deleted in one). Similarly, in (23b) and (24b) the subject inside the first conjunct can bind a variable inside the second conjunct because the subject moves to [Spec,TP], which is the position from which it c-commands material in both conjuncts.

With Johnson’s account as background, I turn to the main issue under consideration here, which is the issue of whether the small conjunct + ATB movement account can explain the ungrammaticality of voice mismatches in gapping constructions.

3.3.3 Ungrammaticality of voice mismatches

The two versions of the small conjunct analysis discussed in the previous section have slightly different implications for the analysis of voice mismatches in gapping. Let us first consider the derivation in which the verb undergoes head movement to Pred, such as the one given in (26b–c) above. Note that there are two $v$ Ps in this structure. Thus, since there are two distinct $v$ heads (and no deletion, thus no identity requirement), the presence of distinct voice specifications on these two heads should not be a problem. In other words, it is not clear what would exclude the structure in (28b) for the ungrammatical example in (28a).21

(28) a. *Roses were brought by some and others brought lilies.

requires movement to case checking positions but is not a problem on current minimalist assumptions, where case can be valued in situ via an Agree operation, and Agree can happen between a single probe (single T in (27b) and two goals (two subjects), as argued by Hiraiwa (2005) on independent grounds). See also Lin (2000, 2001, and 2002) for relevant discussion.

21 Johnson (2004) does note that gapping parallels ATB movement in that both disallow voice mismatches (as shown in (i–ii)), but given the discussion in this section, it is not clear how this parallelism follows from the ATB movement analysis.

(i) *The budget cuts might be defended publicly by the chancellor, and the president might defend publicly her labor policies.

(ii) *It’s [VP defended publicly], that the budget cuts might be $t$, but the president won’t $t$.

(Johnson 2004: 74)
Coordination at a lower level is not an option. Coordinating VPs instead of vPs would entail identity of voice. However, as shown in (29) below, such a structure would leave no room for the subject of the second conjunct.

(29) \[ TP \]
\[ T \]
\[ PredP \]
\[ Pred \]
\[ SUBJECT \]
\[ v' \]
\[ v \]
\[ VP \]
\[ VP \]
\[ and \]
\[ VP \]

We could try to exclude the derivation in (28b) by appealing to the fact that movement of the verb has to proceed through \( v \) heads on the way to Pred, as expected given standard locality considerations. This alternative is given in (30); the verbs first move to \( v \) heads inside the two conjuncts and next to Pred (in an ATB fashion).
The derivation of the ungrammatical examples would then violate the identity condition on ATB movement, since in such cases one of the $v$ heads would be active and the other one passive, as shown below.

(31)  

a. * Roses were brought by some and others brought lilies.

b. 

It is not clear, however, whether this explanation would extend to cases involving complex gaps, such as the ones given in (32a–c) below, which on Johnson’s
account involve remnant movement of a larger phrasal constituent to the specifier of PredP.

(32) a. Some eat poi for breakfast and others eat poi for lunch.
    b. Some gave albums to their spouses and others gave tapes to their spouses.
    c. Some went out to buy beer and others went out to buy fried chicken.

(Johnson 2004: 1)

Such complex gaps also require voice matching:

(33) a. * Some eat poi for breakfast and natto was eaten by some for lunch.22
    b. * Some gave albums to their spouses and roses were given to their spouses by others.

On the remnant movement approach, the same issue arises for examples involving simple gaps. In the derivation of the ungrammatical example (34a), the non-identical elements move out first, and the vPs undergo ATB movement, as shown in (34b).

(34) a. * Roses were brought by some and others brought lilies.
    b. 

\[
\begin{array}{c}
\text{TP} \\
\text{Roses}_j \\
\text{T'} \\
\text{were} \\
\text{PredP} \\
\text{vP}_i \\
\text{Pred'} \\
\text{t}_j \text{v brought } t_k \text{ Pred} \\
\text{FP} \\
\text{FP and FP} \\
\text{by some}_k \text{ F' others}_j \text{ FP} \\
\text{F} \\
\text{lilies}_k \text{ F'} \\
\text{F} \\
\text{t}_i \text{ and } t_i
\end{array}
\]

22 This example is grammatical on the irrelevant reading, in which only the verb is gapped, as shown in (i). On this reading it does not involve any voice mismatches.

(i) Some eat poi for breakfast and eat natto for lunch.
Whether (34b) obeys the identity restriction on ATB movement depends on one’s theoretical assumptions about the nature of traces. If we think of them as copies, as is standard on minimalist assumptions, the two vPs undergoing ATB movement are not identical. As shown in (35) below, one of the vPs contains the copies of by some and roses, whereas the other one contains copies of others and lilies. If we relax the identity requirement on ATB movement and allow copies (and empty heads in general) to be distinct, it is not clear what would prevent v heads from being distinct, and, consequently, voice mismatches to be grammatical. Simply put, it is not clear what could block the following structure.

Furthermore, if non-identical elements vacate the vP, nothing in principle would prevent v heads from moving out of their respective vPs as well, as shown in (36).

23 There is also the issue of what kind of identity ATB movement is sensitive to. Is it featural identity or string identity? In (34), the two tokens of the verb brought are formally distinct; one is a past tense verb and the other one a participle. The example in (i) below also suggests that the identity requirement cannot be taken too literally as the gapped pronoun in the second conjunct allows sloppy reading.

(i) Every girl showed her project to the teacher and every boy showed her or his project to the principal.

(Johnson 2004: 32)
The conclusion that emerges from the discussion in this section is that the ATB movement account does not provide us with a straightforward way to account for the ungrammaticality of voice mismatches in gapping, without modifying either the identity requirement on ATB movement or the copy theory of movement. In the next section, I suggest how a Parallel Merge approach can solve this issue, without sacrificing any of the insights of the small conjunct approach.

3.4 A Parallel Merge approach to gapping

The question that I turn to in this section is whether the Parallel Merge approach to gapping can solve the issues raised above for the ATB movement approach. The most straightforward way to derive the obligatory voice matching in gapping is to posit a structure containing a single v head. Furthermore, it has to be a structure in which v is shared between two conjuncts but the subjects are not. This is precisely the kind of structure that Parallel Merge is best suited for. More specifically, Parallel Merge can generate the structure in (37b), in which a single v head, containing a single voice specification, is shared between two VPs and two subjects. Also, the verb itself is shared between two objects.24,25

---

24 Such a structure, in which more than one node is shared, and the shared nodes do not form a constituent is what Gračanin-Yuksek (2007) refers to as non-bulk sharing (see also Gračanin-Yuksek, this volume).

25 A natural question to ask at this point is whether the account of voice mismatches in gapping suggested here can be extended to pseudogapping (which, as we saw above, also does not tolerate voice
At first glance, Johnson’s ATB movement account and the Parallel Merge account seem to be notational variants of each other. However, I believe that the Parallel Merge approach advocated here can explain the ungrammaticality of voice mismatches in gapping constructions in a more natural way, as it allows sharing a v head without sharing the subject. I maintain the insight that there is a discourse-related F head inside each conjunct, and that the remnants move to the specifier positions of this head.26 The verb undergoes head movement to the head of PredP, the subject of the first conjunct moves to [Spec,TP] and the object to [Spec,FP]. In the second conjunct, both the subject and the object move to [Spec,FP].27 The end result is given in (38); note that there is no remnant movement in it.

mismatches). Given all the differences between gapping and pseudogapping (as well as VP ellipsis) discussed in section 3.1, I do not think it would be wise to extend the Parallel Merge account to pseudogapping. I thus maintain Merchant’s (2008) insight that voice mismatches in pseudogapping are due to vP ellipsis, and that both VP ellipsis and pseudogapping involve deletion (rather than movement).

26 The elements in the specifier of FP do not necessarily have to be interpreted as foci. For example, in the following context, the subjects are naturally interpreted as topics and the objects as foci.

(i) What did Mary and Susan bring?
(ii) Mary brought roses and Susan brought lilies.

The idea that FP is a generalized discourse-related projection can capture such interpretations. Thanks to Myriam Uribe-Etxebarria and Vidal Valmala for suggesting this scenario.

27 I assume that F is not shared between the two conjuncts. Nothing in the analysis, however, hinges on this assumption.
A natural question that arises here is what excludes the alternative given in (39), in which only the verb (but not the v head) is shared between the two conjuncts. If (39) were possible, voice mismatches in gapping constructions should be grammatical, contrary to fact.

To rule it out, I appeal to the assumption, typically associated with the distributed morphology view of the lexicon, that lexical categories are category-less roots,
which acquire their categorial status (of verbs, nouns, etc.) by merging with appropriate functional projections (such as \( v \) or \( D \) heads) (see Halle and Marantz (1993, 1994) for an overview of the framework). Since there is one verbal root in (39) (one copy of \( \textit{brought} \)), only one \( v \) is necessary. Economy considerations thus rule (39) out.\(^{28}\)

Due to space considerations, I will not discuss here how this analysis can extend to complex gaps (which can involve either constituents or non-constituents) and refer the interested reader to Citko (2011b) for a discussion of such cases (based on data from Polish), and to Johnson (2004) and section 3.3.3 above for a slightly different take. Since in such cases, voice mismatches are also ungrammatical, they have to involve sharing of \( v \) heads (in addition to whatever other material is shared). Such cases are a simple extension of the account presented above for simple gaps, as on this account even simple gaps involve non-constituent sharing, i.e. the sharing of \( V \) and \( v \) heads. Thus, in all three cases of gapping (i.e. gapping involving simple gaps, complex constituent gaps, and complex non-constituent gaps), identity of voice follows from a structure in which there is a single \( v \) head shared between the two conjuncts.

### 3.5 Conclusion

To conclude briefly, I provided an account for the ungrammaticality of voice mismatches in gapping constructions. I showed that extending Merchant’s (2008) analysis of voice mismatches in pseudogapping to gapping would imply \( vP \) ellipsis in gapping, which is incompatible with Johnson’s (2004, 2000, 2009) arguments against treating gapping as ellipsis. I suggested a way of reconciling these two analyses, relying on the mechanism of Parallel Merge of Citko (2005). The advantage of the Parallel Merge account is that it allows for sharing of non-constituents; \( v \) and \( V \) could be shared in a gapped structure without the objects (and subjects) being shared as well. The fact that \( v \) has to be shared between two conjuncts is what accounts for obligatory voice matching in gapping constructions.

\(^{28}\) Alternatively, we could appeal to feature mismatch to rule out the structure in (39). If one of the \( v \) heads is passive and the other one active and \( V \) has to raise to \( v \), the verbal complex is going to end up with contradictory voice specifications (both active and passive). I thank Vidal Valmala for bringing this possibility to my attention.
Evidence for Multidominance in Spanish Agentive Nominalizations*

ANTONIO FÁBREGAS

4.1 The empirical problem and its relevance

This chapter addresses the analysis of agentive nominalizations in Spanish, which—in contrast with that of event nominalizations—has been somewhat side-stepped in the syntactic approaches to word formation (with some recent exceptions, among them Marvin 2007, Alexiadou and Schäffer 2008 and Baker and Vinokurova 2009). The main empirical challenge that the syntactic analysis of this kind of nominalizations has to face is that the suffix acts as an argument of the verbal base, while at the same time changing its grammatical category. This dual role seems to require conflicting syntactic configurations. The argumental properties suggest that the affix should be introduced in the derivation as a constituent inside the verbal projection that assigns the agent theta role to it; the category-changing property seems to imply that the affix must be introduced as an independent head which dominates the verbal projection, imposing a new category label, N[oun]. Different previous analyses, which I will discuss in this chapter, have concentrated on only one of its roles, trying to derive the other from independent processes inside the lexicon or syntax. In this chapter, I will argue for a purely syntactic analysis which accounts for the dual nature of this affix within a multidominance approach to syntactic structure.

The multidominance approach to syntactic structure, developed in recent work (Gärtner 1997, 2002, Abels 2004, Citko 2005, this volume, Svenonius 2005,*

* I am grateful to Sergio Scalise, Soledad Varela, Antonietta Bisetto, Joseph Emonds, Gillian Ramchand, two anonymous reviewers, and the attendants of WOSB in Vitoria-Gasteiz (November 2008) for remarks to previous versions of this chapter. Special thanks go to Myriam Uribe-Etxebarria and Vidal Valmala for their detailed comments and suggestions, which have considerably improved the chapter in its final version. The research that underlies this work has been financed by projects 199749/V11 (Aurora Program) and 199852 (DAAD program) from the Norwegian Research Council. All disclaimers apply.
among others), takes to its ultimate consequences the view that trees represent
hierarchical organizations of sets (Chomsky 2004) such that each node, terminal,
or non-terminal stands for a set that contains its constituents as members. In the
multidominance approach, movement—understood as remerge—equals a situation
in which a particular constituent belongs to more than one set (in such a way that
displaced constituents are the intersection of two or more sets). This situation can
be illustrated by (1), where the constituent Y belongs to two sets: XP (\{\{Y\}, \{X\}\}) and
YP (\{Y\}, \{XP\}). As can be seen in (1), this approach to movement allows a constituent
of XP to reproject and head its own projection, imposing its category to XP
(cf. also Starke 2004).

(1)

\[
\begin{array}{c}
\text{YP} \\
\text{XP} \\
\text{Y} \\
\text{X} \\
\text{X} \ldots
\end{array}
\]

In the previous literature, several empirical phenomena have been argued to provide
evidence for remerge, among them antecedent-free relatives and ATB extractions
(Abels 2004, Citko 2005). This chapter argues that empirical evidence for multi-
dominance also comes from the domain of word formation processes, in the phe-
nomenon of affixation processes that at the same time require and cancel (part of)
the argument structure of the base, as is the case with agent suffixes in Spanish. To
the extent that my analysis of Spanish agent nominalizations is satisfactory, I provide
evidence in favor of the multidominance approach. The secondary goal of this
chapter is, profiting from the detailed work on word formation processes which
already exists in the literature, to establish some of the conditions under which
remerge is allowed or disallowed.

The chapter is structured as follows: in section 4.2, the empirical phenomena
are described and it is shown that some category-changing affixes require
thematic positions of the base which, after affixation, get canceled. In section 4.3,
a multidominance analysis of these cases is proposed, with particular emphasis in
delimiting the situation under which reprojection is necessary. Section 4.4 shows
that analyses which do not use reprojection are not able to predict the empirical
patterns of the phenomenon under consideration and explores other advantages
of the approach adopted in this chapter. General conclusions close the chapter in
section 4.5.
4.2 Empirical data: affixes that cancel argument structure

Several nominalizations involve affixation of a particular affix to a verbal base which, as a result of it, loses part of its argument structure. One well-studied instance of this situation is the case of suffixes that construct agentive nominalizations in the major European languages. (2) illustrates the pattern with the Spanish suffix –dor, roughly equivalent to English –er.

(2) a. María limpia suelos de madera.
   ‘María cleans wooden floors.’

b. La limpiadora de suelos de madera
   ‘The cleaner of wooden floors’

c. #La limpiadora de María
   ‘The cleaner of María’

The suffix –dor is sensitive to the theta grid of the verbal base in a very specific way: it requires the verb to contain a position for the agent (in a sense that will be clarified below). There is general consensus that the Spanish nominalizations in (3) are ungrammatical because the verbal stems are not consistent with an agent theta role.

(3) a. *nace-dor (vs. naciente, “raising”)
   born-er

b. *mori-dor (vs. muriente, “agonising”)
   die-er

c. *crece-dor (vs. creciente, “growing”)
   grow-er

d. *aparece-dor
   appear-er

This situation raises many questions, which have been extensively discussed in the morphological literature. The main empirical problem is that the suffix requires the verbal stem to allow for a particular thematic position which, nonetheless, turns out to be unavailable in the resulting nominalization. From a surface perspective, then, the suffix seems to be doing two jobs at the same time: changing the grammatical category of the word (from verb to noun) and canceling a theta position which would have been otherwise available. Some solutions, which will be discussed in section 4.3, have been proposed to cope with this phenomenon.
4.2.1 Against the semantic redundancy approach: refining the meaning of agentive nominalization

One first approach to cope with the phenomenon has been to propose that –dor is incompatible with an independent agent because they both express the same semantic notion and, thus, one of them is redundant. To argue against this approach I need to be more specific about the meaning of “agentive” in “agentive nominalization.” The nominalizations in –dor can be used to denote agents—defined as volitional and conscious initiators of an action, as in the most salient reading of James Moriarty infected London—but also to denote causers—non-volitional initiators of an action, such as the subject in The Bubonic Plague infected Romania—and instruments (entities used by an entity to trigger an action). These three notions have in common that they represent the entity involved in the initiation of a particular event. I will use, therefore, the term “initiator” as a cover term to refer to agents, causers, and instruments (following Ramchand 2008). In (4) the three possible initiator-related interpretations of –dor nominalizations are represented.

(4)  I. Agent.
   a. María atraca bancos.
      Maria robs banks.
   a’. #la atracadora de María / la atracadora de bancos
      the robber of Mary / the robber of banks

   II. Instrument.
   b. Esta crema tonifica los músculos.
      this cream tones up the muscles
   b’. #El tonificador de la crema / El tonificador de los músculos
      the toner up of the cream / the toner up of the muscles

   III. Causer.
   c. El plomo contamina el río.
      the lead contaminates the river
   c’. #El contaminador del plomo / El contaminador del río
      the contaminator of the lead / the contaminator of the river

---

1 These are the only three notions available in an agentive nominalization, but do not exhaust the possible interpretation of –dor nominalizations in Spanish. Locative meanings—X denoting the place where an event takes place—are also possible, and increasingly frequent in modern corpora: comedor (lit. “eater”) can be both the initiator of an eating event and the designated room for eating; the same with intercambiador—“exchang-er”—and some others. These locative nominalizations with –dor will be disregarded in this analysis, with some minor observations in section 4.4.4, but notice that they also comply with the need that the verbal base contains an initiator position (from dormir, “to sleep,” *dormi-dor* is ungrammatical, vs. *dormi-torio,* “sleeping room”).
The fact shown in (4) is that –dor nominalizations can refer to a variety of initiator-related notions. This situation casts doubt on the semantically based explanation that –dor is not compatible with another initiator because its presence makes the meaning contribution of the second initiator redundant. Notice in (5) that the grammar—if we ignore for the moment case marking—does not disallow several semantic initiators in the same sentence, indicating either that this is not considered redundant or that redundancy is not enough to make the structure ungrammatical.\(^2\) (6) shows that, even in the context established by the sentence in (5), –dor is incompatible with any of the three initiators. In other words, it is not possible that –dor takes one of the initiator-related semantic roles and the prepositional argument takes another.

\[\text{(5)}\]
\begin{align*}
\text{Moriarty}_{\text{Ag}} & \text{ contaminó Londres con plomo}_{\text{Caus}} \text{ con unas bombas camufladas}_{\text{Instr.}} \\
& \text{Moriarty contaminated London with lead with some camouflagged bombs.}'
\end{align*}

\[\text{(6)}\]
\begin{enumerate}
\item #el contaminador de Moriarty \\
the contaminator of Moriarty
\item #el contaminador del plomo \\
the contaminator of the lead
\item #el contaminador de las bombas camufladas \\
the contaminator of the bombs camouflaged
\end{enumerate}

The data in (6) suggest that the suffix –dor occupies the only initiator-related argument position available inside the nominalization, blocking the presence of any other argument of this kind. Presumably, in (5), different initiators are possible because the verbal projection is dominated by a number of functional heads (related to aspect, voice, mood, and tense) which can host the additional initiator-related modifiers. Observe that, in (5), only one of the initiators behaves as an argument with respect to case marking, while the other two need to project as adjuncts. It has been noticed that adjuncts are generally impossible inside nominalizations (Fu, Roeper, and Borer 2001, Alexiadou 2001 for event nominalizations; Baker and Vinokurova 2009 for agent nominalizations). These authors have proposed that the impossibility is due to the absence of the relevant projections that license adjuncts (Cinque 1999) inside the nominalization, where the functional structure is, at best, impoverished. Although this is a standard assumption in the study of nominalizations, I will discuss

\(^2\) The following symbols are used: \text{Ag=Agent, Ins=Instrument, Caus=Causer.}
4.2 Empirical data: affixes that cancel argument structure

direct evidence for the absence of higher functional projections inside agentive nominalizations in section 4.3.3.

4.2.2 The relation between –dor and the verbal base: comparison with –ción

An analysis of –dor nominalizations should also take into account additional properties that have been noticed in the empirical descriptions, although they may in principle be unrelated to the interaction between the suffix and the argument structure of the verbal base. One property that has been noticed is that –dor nominalizations require a verbal base. This property is so strong that, whenever a noun is taken as its base, it is turned into a verbal stem by the addition of a theme vowel (by default, -a-, exponent of the unmarked conjugation class in Spanish, cf. Oltra-Massuet and Arregi 2005); in these cases, the corresponding verbal forms do not exist, but the verbal stem is required in the –dor nominalization (see (7) for these cases in Spanish and Italian).

(7) a. Sp. leñ-a-dor, *leñ-a-r
   wood-ThV-er,  wood-ThV-infinitive
   “woodsman”, “to wood”

   b. It. cesoi-a-tore, *cesoi-a-re
   shears-ThV-er, shears-ThV-infinitive
   “man that works with a clipper”, “to shear”

This is not the case with the suffix –ción, which also differs from –dor in that it does not cancel any argument structure of the base (see footnote 1 for illustration). Among many other cases, the nominalizations in (8) are attested, where the base is not verbal and does not show the morphological shape of a verbal stem; in other words, there is no theme vowel.

(8) fun-ción, “function, performance”
   aten-ción, “attention”
   emo-ción, “emotion”

3 In this chapter we assume that allomorphs are taken as exponents of syntactic and semantic differences (see, among others, Jablonska 2007, Lundqvist 2008), and it is not considered to be merely the result of a PF operation of lexical insertion. For this reason, we do not treat the suffix –dor and its allomorph –tor as phonological variants of the same morphosyntactic bundle of features, but assume that they are two lexical items corresponding to two different syntactic representations. Notice that all agentive nouns that appear without a theme vowel use the lexical item –tor (or, alternatively, –(s)or): lec-tor, “reader”; profes-or, “teacher”; ac-tor, “actor”; etc. The fact that –dor requires the theme vowel is not due to phonological reasons, as the segment /d/ can be preceded by some consonants in Spanish (for example, mal-dad, “evil”); alternatively, –tor can be preceded by a vowel (audi-tor, “tax official”). I must immediately admit, though, that I have not worked out an analysis of –tor, some of whose properties are different from –dor.
Another property of the nominalizations with –dor is that the base must denote a dynamic event and the nominalization denotes an entity defined by its relation to (the initiation of) an event. Thus, the interpretation of (9), where –dor is combined with the verb conocer, “to know”, is not just “someone who knows”. Although the verb conocer is normally stative, (9) denotes “someone who has acquired some particular knowledge through practice or study.” As expected, this particular reading of the verb is restricted to knowledge which implies a process of studying. This predicts that (9b) is not acceptable, as here conocer has the sense of “to be aware of the existence of someone” and thus does not require a previous process of study. (9c) is, as expected, acceptable only on the reading that someone studied Napoleon’s life and times, and not that she is only aware of the fact that he existed.

(9) a. conoe-dor de vinos
    connoisseur of wines
b. #conoose-dor de Juan
    know-er of Juan
c. conoe-dor de Napoleón
    know-er of Napoleon

This restriction does not affect –ción, which can give rise to names of states, properties or entities which do not entail the existence of any event (10).

(10) a. na-ción, “nation”
    b. inten-ción, “intention”
    c. inhibi-ción, “quality of being inhibited”

In the analysis that I propose in the next section, these additional properties of –dor—as opposed to suffixes like –ción—follow from the same principles that explain its interaction with argument structure.

4.2.3 Other affixes like –dor

Although my analysis in this chapter is restricted to –dor, this suffix is not the only case of suffixation with this behaviour in Spanish. Although not generally noticed in the literature, for some speakers, this situation is also illustrated by the Spanish suffix –miento, which forms event nominalizations from verbs. Fábregas (2010) argues that –miento selects verbal bases that have an available position for an incremental theme or, in general, a measurable path (cf. Ramchand 2008 for the relationship between these notions). The incremental theme measures the temporal extension of the event, as in to eat the cake, where different parts of the (bound) entity “the cake” are mapped into different time intervals corresponding to the (bound) event “to eat the cake”. The same theta role can be assigned, inside the
spatial domain, to some path arguments, such as “to the house” in “run to the house”, where different parts of the (bound) path which goes to the house correspond to different stages of the (bound) event of running.

The case of –miento is similar to –dor in the sense that many native speakers tend to reject the realization of these incremental themes as a prepositional phrase in the nominalization with –miento, and therefore have a contrast between the two sentences in (11).4

(11) a. La persecución del sospechoso (por parte de los agentes) hasta el aeropuerto

‘The pursuit of the suspect to the airport by the agents’

b. *El seguimiento del sospechoso hasta su casa

‘The tracking of the suspect to his house’

The prepositional phrases with hasta, “until, to” are ambiguous in Spanish between a spatial path reading (hasta su casa, “to his house”) and a temporal boundary reading (hasta su fuga, “until his escape”). The same speakers find a contrast between (12a), with the spatial path, and (12b), with the temporal boundary use.


the tracking of the suspect [to his place]

b. El seguimiento del sospechoso [hasta su fuga]Tem

The tracking of the suspect [until his escape]

The data in (11) and (12) suggest that –miento occupies the argument position that is otherwise used for incremental themes, in such a way that, if one appears, the other is blocked. Notice that –miento, like –dor and unlike –ción, also requires the presence of the theme vowel; the minimal pairs in (13), each one derived from the same verb, illustrate this property.

4 For some speakers, the use of the path PP interacts with the kind of preposition that introduces the other internal argument of the verb, and like this, (11b) improves if instead of de, “of”, the preposition a, which generally marks accusative and dative, is used before el sospechoso, “the suspect”. This suggests that, at least for these speakers, the choice of the preposition interacts with the positions occupied by the arguments. It seems that for them, in (11b), “the suspect” occupies the undergoer position and –miento must occupy the path object position; if the same argument is introduced with a, the position of path object is available, maybe because “the suspect” does not project as an undergoer and –miento can occupy its position.
This suggests that the analysis that accounts for the properties of -dor should also be extended to cases such as -miento. For reasons of space, I will not elaborate on the analysis of the second affix here (cf. Fábregas 2010).

4.3 How a multidominance account can explain these cases

I will argue that the behavior of the affix -dor—and, by extension, of other affixes which at the same time change the grammatical category of the word and cancel part of the argument structure of the base—can be explained if the affix is merged in the relevant argument position inside the verbal domain and, in a further step, it remerges, projecting a phrase that changes the grammatical category of the structure that introduced it. I assume that nominalizer affixes contain a category feature N (for “noun”). This N feature, not being dominated by any nominal functional structure in its base position, is still syntactically active and needs to project as soon as possible. The fact that the feature is not licensed in the base position triggers remerge of the affix. This remerge projects the affix as the head of an NP which takes the verbal structure as its complement, giving as a result a deverbal noun. The difference with an affix such as -ción is that this class of affixes is never introduced as an argument of a verb; thus, remerge is not necessary and the affix does not require a theme vowel.

4.3.1 How it happens

Throughout this chapter I adopt the hypothesis of verb decomposition put forth by Ramchand (2008). The reason for this is that, as has been shown before, the range of meanings associated with -dor are those that have been identified with the InitiatorP by Ramchand.5 A further advantage of adopting this analysis is that it also opens the
way to account for the historical relation between agentive nominalizations and result nominalizations. Finally, it will also let us explain some denominal formations discussed in section 4.4.4. It should be pointed out, however, that nothing hinges on this choice and that the multidominance analysis I defend in this chapter is independent from Ramchand’s approach and could be easily restated within alternative approaches to verb structure with minor adjustments (like Larson 1988 or Chomsky 1995a).

The proposal is that the affix –dor (as a lexical item) spells out an N constituent which is placed in the specifier position of InitP, the projection which denotes the causing subevent. The head Init denotes a stative relation between an entity that causes a process and the process caused by this entity. Thus, in the specifier position, the N spelled out as –dor is interpreted as the initiator of the process denoted by the complement of Init, which is Proc(ess)P (14).

\[
\text{InitP} -\text{dor} \{\text{N}\} \text{Init ProcP}
\]

The configuration in (14) straightforwardly explains why no other initiator is available with –dor nominalizations: the position is already occupied by the affix, so it is not available for any constituent. A second constituent with an initiator-related interpretation could only be introduced if there was another position available, which seems to be the case in full sentences (cf. 5), but not in nominalizations. (14) also explains why –dor compulsorily selects verbs that have an initiator position, explaining data such as those in (3), where the suffix is unavailable with unaccusatives. (14) also explains why the interpretation of a stative verb cannot be canonical with –dor, accounting for (9) and similar data: InitP requires the presence of ProcP—the projection that denotes a dynamic activity—in the complement position of Init.

verbs that differentiate the initiator and the undergoer arguments, the –dor nominalization unambiguously refers to the initiator: this is generally the case in caused-motion verbs (arrastrar, “to drag,” empujar, “to push,” . . . ), change of state verbs that do not create and do not destroy the object (restaurar, “to restore,” amueblar, “to furnish,” . . . ), and the causative reading of degree achievements (engordar, “to fatten up,” blanquear, “to whiten,” . . . ). Psychological verbs clearly differentiate the undergoer (which experiences the psychological state) and the initiator (which might not be sentient and causes the state); in this case, to the extent that they allow for –dor nominalizations (something possible sometimes only in American dialects), they refer only to the causer of the state: preocupar, “to worry,” interesar, “to interest,” or calmar, “to calm down.” Notice, also, that in verbs that contain a ProcP but do not have an InitP, such as crecer, “to grow” in Spanish, which cannot be causativized, –dor nominalizations are out. However, with some verbs, due to their meaning, –dor seems to occupy both positions, spec, ProcP and spec, InitP. We will go back to these in footnote 6.
in order to be able to have a specifier, given Bare Phrase Structure (Chomsky 1995a). ProcP, furthermore, is also necessary to satisfy the semantics of InitP.

The structure in (14), however, is not enough to explain how –dor is able to turn the whole structure into a noun. This is performed in the following step of the derivation.

The features that correspond to –dor contain N, that is, a nominal category feature. Notice that in (14) this feature N is still syntactically active, because it is not dominated by the nominal functional projections which nouns require in Spanish—arguably, universally—to be licensed as referential arguments: Gender, Number, and Determiner (see 4.2). These features do not dominate the structure in (14), and will not be able to dominate it as it stands, as the head of the structure, InitP, is a verbal projection, not a nominal projection. In order for the derivation not to crash, the feature N motivates that –dor remerges, projecting NP, the phrase which corresponds to its unlicensed feature. Notice that in (14), –dor does not have any complement and, therefore, it is still ambiguous between a maximal and a minimal projection (through Bare Phrase Structure, again); this allows it to project as maximal projection, a possibility which is now unavailable for InitP, which has already taken a complement and a specifier (15).

\[
\text{(15)}
\]

\[
\text{NP} \\
\text{-dor} \\
\text{InitP} \\
\text{Init} \\
\text{Init} \\
\text{ProcP}
\]

In (15), the requisites of –dor can be satisfied, as its feature N can be dominated by the relevant set of nominal functional projections. This straightforwardly explains why the whole structure is nominalized by –dor, without any need to propose zero affixes or to claim that nominalization is obtained through merge of functional heads like number or determiner (see below on why this is a good result).

In contrast, a suffix such as –ción is not introduced in a thematic position inside the verbal phrase, but as an external head which takes the verbal phrase as its complement (16). From here it follows that –ción does not cancel argument structure, and also that, not requiring to be introduced in an argument position, it does not require its complement to be a verb. Not being introduced in an argument position, when –ción combines with a verbal phrase, it does not denote any thematic role of the base, but the event itself.
To summarize, my specific proposal is the following. Three properties of –dor nominalizations are explained in (14): –dor nominalizations denote initiators; –dor nominalizations require bases that contain an initiator position; and –dor nominalizations are incompatible with an additional DP interpreted as an initiator. The remaining property is explained by (15): words with –dor are nouns and, as such, combine with gender, number, quantifiers, and determiners. The relationship between (14) and (15) is explained by the need of –dor to satisfy its features: N is not licensed in the base position, so it is still active and remerges to get licensed. The difference with an affix such as –ción is that this is introduced as a head that takes the verbal phrase as a complement, and never occupies an argument position.

4.3.2 Why it happens

Let us be more precise about how (15) is obtained: why exactly does –dor project? As presented before, my proposal is that this is due to the presence of an N feature in the set of features to which the affix corresponds. In this section I will present why the presence of N in that particular configuration forces the affix to remerge.

Longobardi (1994), based on previous suggestions by Stowell (1989), proposes that noun phrases in an argument position need to be dominated by determiners. This is required for semantic reasons, related to the need to be a constant in order to saturate a variable, and also for formal reasons related to the case filter. The principle can be formulated as (17).

(17) N in an argument position is licensed if and only if it is dominated by DP.

Thus, in (14), repeated here as (18), N is not satisfied; it occupies an argument position (the specifier of InitP), where it gets an initiator theta role, and it is not dominated by DP.

(18) \[
\begin{array}{c}
\text{InitP} \\
\hline
-\text{dor} \\
[N] \\
\hline
\text{Init} \\
\text{ProcP...}
\end{array}
\]
N is not licensed here because it is dominated by InitP, which does not contain a D feature. This configuration does not allow DP to be introduced in a position that dominates –dor. DP belongs to the functional sequence of nominal categories, and the head of the structure in (18) is InitP, a verbal category. At this point, N needs to project in order to get licensed. When N projects NP over InitP, it becomes the head of the structure in (15); now the head of the whole projection is a nominal category, and the structure can be dominated by DP. Without the remerge, however, DP cannot be introduced in the structure and, therefore, N is not licensed. The result is ungrammaticality. (19) shows the remerge process.

The activity condition (Chomsky 2004) guarantees that N is able to remerge: as it is not licensed in its base position, it can take part in syntactic processes. This automatically prevents a DP in an argument position from remerging, predicting that verbal phrases which contain DP arguments will not become nominalizations of any class (see 3.3). Bare Phrase Structure also guarantees that –dor will be able to project, as in its base position it does not take any complement and is thus ambiguous between a head and a maximal projection. Let us move to the next subsection, where I will show how my analysis uses these two independent principles to avoid overgeneration.

4.3.2.1 Avoiding overgeneration We will show now how the analysis presented here solves a problem of overgeneration that previous syntactic analyses of agentive nominalizations have run into. Consider the following structure, taken from Marvin’s (2007: 206) analysis of Slovenian nominalizations.

(20) a. plava-l-ec
swim-Part-er, “swimmer”

b. NumberP
   Number
   T2P
   T2
   -l- n
   ec vP v
   ∅ v
   √ plav-a
The affix -ec- is in this account a manifestation of little n, that is, a functional projection that turns a root into a noun (Marantz 2001); the verbal structure contained inside the noun goes up to a temporal head which materializes as the past participle and the whole structure is nominalized by means of the higher functional projections, in this case, number.

Notice that in Marvin’s structure there is no direct connection between the lexical item merged in the agent position (-ec-) and the nominalization process; they are performed by different heads.

The question that arises in this analysis, because of the absence of a direct connection between the noun getting the agent theta role and the nominalization process, is why another noun merged in the argument position cannot be used to create a nominalization. For example, why is the noun corresponding to “person” unable to occupy the place that -ec- occupies in (20)? What exactly prevents the nominalization from containing a full DP, or a pronoun, instead of the nP contained in (20)? Let us see how my proposal blocks agitative nominalizations being obtained by pronouns or whole noun phrases, something that Marvin’s approach cannot attain without additional assumptions.

One important difference between an affix such as –dor and a noun such as “man” is that the first is the spell-out of a functional projection, while the second is the spell-out of a functional projection which contains a root (Marantz 1997). This difference has consequences for the semantic content of each one of these lexical items: while the noun phrase is associated to conceptual meaning, and therefore has a full entry in the speaker’s mental dictionary which corresponds to a particular definite description, the affix –dor does not have a specific entry in the mental dictionary and thus does not correspond to any definite description. As a consequence of this, nominalizations with –dor, or the English equivalent –er, can denote animate entities as well as non-animate entities (instrument nouns being an example of the second class). The affix is defined by a particular theta role, assigned in a specific syntactic configuration, but not by a dictionary entry in the speaker’s mental lexicon.

The absence of a root in the case of the affix not only has these semantic consequences, but also structural consequences. It is standardly assumed that roots are merged as complements of the functional head (21).

(21) \[ \begin{array}{c}
\text{NP} \\
\text{N} \\
\sqrt
\end{array} \]

(21) corresponds to the structure of any noun in Spanish, English, or Slovenian; the nominalizer affix, in contrast, corresponds to a N without any root as a complement.
Imagine that we merge the structure of (21) in the position of the specifier of InitP, as represented in (22).

(22) \[ \text{InitP} \]
    \[ \text{NP} \]
    \[ \text{N} \sqrt \text{Init} \ldots \]

In (22), N requires to remerge in order to be satisfied. However, it cannot project as a whole phrase, because in the base position it has already projected as NP (when it took the root as a complement). By Bare Phrase Structure, NP will not be able to take InitP as a complement. However, if NP does not take InitP as a complement, it will not be able to head the whole structure. This, in turn, prevents any nominal functional projection from dominating NP, and the derivation will crash.\(^6\)

A similar situation takes place if, instead of a noun phrase, we have a pronoun in the relevant argument position. If pronouns are not merely DPs, but also contain NPs inside them (Panagoitidis 2002), the structure corresponding to them will be (23).

(23) \[ \text{DP} \]
    \[ \text{D} \ldots \text{NP} \]

This blocks remerge of DP in the same way as the presence of the root blocks remerge of NP in the previous case. Moreover, it is not clear that there are syntactic reasons for the DP to project as a head in (23) in any case. Pronouns contain the feature D (cf., for example, Longobardi 2001). This means that in this case we have a D element in an argument position; the feature is licensed in situ, and does not require to perform further syntactic operations to guarantee formal licensing. Even though DP is presumably active in the specifier of InitP, because it still needs to get its case checked, projecting as a head would not have the effect of satisfying this

---

\(^6\) As noticed before, in some verbs the initiator argument and the undergoer argument are fulfilled by the same constituent (for example, in verbs such as \textit{run}), where the subject both controls the initiation of the event and suffers some change (of location in this case) by performing the process. In these cases, \textit{-dor} occupies both the specifier of ProcP (where it is interpreted as undergoer) and that of InitP (where it is interpreted as initiator, which is the compulsory part for \textit{-dor}). Thus, following my reasoning, in this class of verbs \textit{-dor} must remerge twice. In the first remerge operation, it is remerged from the position of specifier of ProcP into that of specifier of InitP. The question arises whether this first remerge should not block further remerge operations. Notice, however, that the first remerge of \textit{-dor} from ProcP to InitP, does not project \textit{-dor} as a phrase, as the suffix does not take any complement when it is merged with InitP. Assuming, as Ramchand (2008) does, that in the specifier of InitP the constituent must get a particular theta role, it is crucial for this remerge operation that \textit{-dor} does not take InitP as its complement; if that was the case, InitP, not dominating it, would not be able to assign a theta role to it. Thus, the only convergent derivation would be that \textit{-dor} remerges as a specifier of InitP, letting InitP dominate it. As \textit{-dor} does not have any complement at this point, it is still a head and further projection is not blocked.
requirement. If the DP heads the verbal structure, it will turn the whole configuration into a nominal constituent, and, therefore, it will block the higher verbal functional projections, among them TP, to be introduced in the derivation. Thus, for a pronoun, becoming the head of the configuration is impossible (by Bare Phrase Structure) and, moreover, would have catastrophic results with respect to case checking.

Our analysis thus guarantees that no NP and no DP in the agent theta position will be able to project and nominalize the verbal structure. This result is, I believe, an advantage with respect to Marvin’s analysis, to the extent that her proposal, in not connecting the argument position with the nominalization process, cannot block other constituents different from the affix to occur in the agent place.

4.3.3 When it happens

In this section I will explore the interaction between remerge and phases. If remerge is due to the need to satisfy a formal feature, then it follows that this remerge has to take place before a syntactic phase (Chomsky 2001) is completed. If the head containing the unlicensed feature remains in the domain of the phase when the structure is transferred, the derivation will crash. This means that verbal heads which define phases that transfer the whole verbal domain, such as Tense and Complementizer, must be missing from inside the nominalization. Their presence would give rise to the configuration in (24), impossible given the phase impenetrability condition (Chomsky 2001).

(24) * NP
    \  /  \\
   TP  N
   /  \
T   ...InitP

We can empirically show that, if these projections are inside a nominalization, remerge is unavailable. Tense and Complementizer have been identified in some types of nominalizations, more in particular in English gerund nominalizations (van Hout and Roeper 1998). The argument put forward by these authors is that the presence of tense dominating the verb binds the event variable, resulting in a specific reading of the event (in other words, the event expressed by the verbal domain is factive, presupposed to have taken place).

This conclusion allows us also to treat factive nominal infinitives in Spanish as nominalizations that contain Tense (and, possibly, also C). The prediction of my analysis of agent nominalizations is that in English gerunds or in Spanish infinitives
the nominalization cannot be obtained by remerge of an N from an argument position. By implication, this means that no argument structure can be canceled inside infinitives or gerunds. This prediction is borne out; the factive infinitives in (25) keep all the argument structure of the verbal phrase, including agents and locative paths.

(25) a. Este misterioso haber perdido las llaves de Juan
    this mysterious to.have lost the keys of Juan
    va a causar problemas.
    is.going to cause troubles
    ‘John’s having lost the keys will cause trouble.’

b. Este frenético seguir a los sospechosos
    this frantic to.follow ACC the suspects
    hasta su guarida de los policías
    to their hideout of the policemen
    ‘The policemen’s frantic following the suspects to their hideout’

In contrast, –dor nominalizations cannot be used to refer to single occurrences of events, showing that no T or C head are present in their internal structure. Agent nouns cannot be used to refer to specific instances of an event. Contrast, for example, the grammatical phrase el fumador de puros, “the smoker of cigars,” where the event of smoking is taken as a general habit, with the ungrammatical *el fumador de este puro, “the smoker of this cigar” where the speaker intends to describe a particular single occurrence of the event of smoking—although pragmatic factors can help improve this second reading in some contexts. Denoting specific events, though, is possible in gerund and factive infinitive constructions, as shown in (25).

Thus, in an infinitive, an affix cannot occupy any argument position, because, as T is present, the affix would not be able to escape from the phase to get licensed. The only alternative in the infinitives is to introduce DPs or pronouns in the argument positions and nominalize the structure by having the nominal functional projections take it as a complement.

Notice that Marvin’s analysis cannot account for this empirical generalization. These infinitives are nominalized either by a zero morpheme or by the nominal functional projections, Determiner, Number, etc., but this does not make the verb lose its verbal properties, such as accusative case assignment. Marvin’s analysis of agent nominalizations already uses the extended functional projections of the noun to nominalize and allows Tense layers inside the noun, so she cannot predict any difference between gerunds or infinitives, on one side, and nominalizations on the other side.
4.4 Why a multidominance account is necessary

In the previous section I have argued that a multidominance analysis of –dor nominalizations is possible and has additional empirical advantages with respect to some of the previous analysis. In this section I will argue that it is necessary, for the other possible analyses are unable to give account of the data.

To the best of my understanding, there are three other possible solutions to explain the fact that the argument position gets canceled inside the nominalization: the argumental position is bound—not occupied—by the suffix (4.1), the argumental position is occupied by something other than the suffix (4.2), and the argumental position is not available inside the nominalization (4.3). I will discuss these three possibilities and present the reasons why they do not account for the data; after this, I will provide further empirical evidence in favor of my analysis with multidominance and the decomposition of verbs assumed here (4.4).

4.4.1 What if the argumental position is bound by the affix?

Assume that –dor is never merged in an argument position, but introduces an operator that binds the argumental variable contained in the specifier of InitP.

(26) \[
\text{NP} \\
\text{N} \quad \text{InitP} \\
\text{–dor} \quad \text{Init} \\
\text{Op}_i \quad \text{vi} \\
\text{Init} \quad \ldots
\]

The question here would be why this variable cannot be syntactically realized as an indefinite DP which then would satisfy the case filter by getting genitive case which is independently available inside the noun projection headed by the nominalization. This theory does not explain why (27) is ungrammatical (when the indefinite is interpreted as agent):

(27) *el corredor de un chico
the runner of a boy

The situation does not improve if instead of treating –dor as containing an operator that links a variable, I claim that it introduces a semantic function that takes a verb as input and denotes an agent. The same problem of why the bound position cannot be syntactically expressed would arise in all these approaches.

Precisely because of this problem, several authors have proposed that the operation whereby a suffix saturates an open argument position of the base does not take place in syntax, where the structure would still be available for projection, but takes
place in the lexicon (see in particular Reinhart and Siloni 2005 for this difference between lexical and syntactic operations). This is the kind of analysis that Williams (1981) and Di Sciullo and Williams (1987) suggest for suffixes that are associated to thematic positions of their bases. These authors argue that the suffix binds the argument position before syntactic projection in such a way that, when the base is introduced in the syntax, the position is not available. In (28), coindexation represents this lexical argument satisfaction operation.7

(28)  

\[-dor_i + V \quad [\text{Agent}_i, \text{Patient}]\]

Booij (1986) proposes an analysis of agent nominalizations in Dutch along the same lines; Scalise (1984) for Italians, and Varela (1990) for Spanish, provide analyses where the affix itself satisfies the argument position in the lexicon, blocking its syntactic projection. These accounts assume the lexical integrity hypothesis and, in general, that morphological and syntactic operations are substantially different, so it will have to face the number of arguments that have been proposed against this division of labor (among many others, Marantz 1997, Embick 2000, Emonds 2000, Arad 2005). Assume for the time being, though, that the distinction is basically right. The problem is that, even in this case, Di Sciullo and Williams (1987) cannot explain the data in (7), repeated here as (29).

(29)  

b. It. cesoiatore, “scissors-Agentive,” “a particular type of gardener”

These forms are problematic for their theory, and in fact have been treated as exceptions by most lexicalist accounts, because (a) they are constructed over nouns that as such do not have an argument structure, and (b) the base taken by the suffix cannot show up as a verb (*leñar, *cesoiare). If the lexical selection of –dor performs an operation on argument binding, this would not be satisfied here; –dor would look for a particular thematic position which the lexicon cannot provide, for the base is a noun in the lexicon; if it is turned into a verb by addition of the theme vowel, though, the problem for the proposal is to explain why this verb is not stored in the lexicon and used normally.8

---

7 The specific way in which derivational affixes satisfy argument positions of their bases’ theta grids is, however, unclear in Di Sciullo and Williams’ analysis (1987: 34–5), as the theta role is not assigned to the affix, but rather “used up” by the whole nominalization, which denotes an agent. Derivational affixes contrast with members of compounds (Lieber 1983), which occupy argument positions.

8 The problems that these formations present for lexicalist analyses of agent nominalizations become even clearer when we consider Samek-Lodovici’s (2003) analysis of light verb constructions, performed from a non-syntactic approach to word formation. This author provides evidence that the argument structure of a predicate is a grid of “anonymous variables” which is responsible for the addicity of the predicate; an independent lexical conceptual structure (LCS) provides each one of the variables with its interpretation. The two representations are linked to each other. In a light verb construction, the light verb provides the argument variables and the adjective or noun which combines with it, the LCS representation...
Lieber (2004)—see also Booij and Lieber (2004) for the particular case of agentive nominalizations in English—has a proposal that tries to give account of the cases in which the suffix does not bind an agent. Her proposal is that the position that –dor binds is not fixed and this affix is able to bind other theta positions. Similar in spirit to Lieber’s analysis is Heyvaerts (2003), who proposes from a cognitive perspective that the agent suffix does not target an agent theta position, but rather a “subject” position, and is therefore not restricted to bases that allow for agent thematic positions. These proposals have, in my opinion, two dangerous consequences. The first one is that it gives up the relationship between semantics and structure, which is crucial in a morphological approach where affixes are assumed to be units (rather than processes). The second one is that, by giving up the correlation between the affix and a thematic position, it overgenerates: specifically, it cannot explain why words such as those mentioned in (3) do not exist in Spanish: they all have a theta role that can be bound by the affix, only that it is not an initiator role (*nace-dor, *mori-dor). Therefore, I believe that it is fair to conclude that this first alternative does not give full account of the data.

4.4.2 What if something else occupies the position of the agent?

Consider now what happens if the initiator position is unavailable because another constituent occupies this position.

Let us consider for one moment the highly unlikely possibility that it is the verb that moves to the agentive position. Imagine, in order to make the proposal have some plausibility, that the nominalizer suffix is –or (as in other deverbal nouns in Spanish, such as tem-or, “fear,” from tem-er, “to fear”) and that a morpheme -d-, which could be identical to the Spanish past participle, lexicalizes Init.

gives semantic content to the variables. In (i), the verb dare, “to give,” provides the predicate strizzare, “wringing,” with argument variables that otherwise the noun itself lacks; this explains that the structure allows for three arguments, corresponding to the variables of a ditransitive verb such as “to give.” In (ii), in contrast, the light verb is fare, “to make,” a transitive verb, and therefore there are only two argument variables available.

(i) I ragazzi hanno dato una strizzare *(ai panni).
   the boys have given a wringing to.the clothes
   ‘The boys wrung the clothes.’

(ii) L’architetto ha fatto una passeggiata *(al cane).
   the architect has made a walking to.the dog
   ‘The architect walked (*the dog).’

Given this situation, it is clear that the nouns do not contain argument variables. They provide the structure with an LCS, but the argument variables are provided by the light verb. In consequence, we do not expect the nominal bases in the examples of (28) to contain argument variables that the affix –dor can satisfy.
This proposal encounters several problems. Even if we assume that movement of the verb to the specifier of InitP does not violate antilocality (Abels 2003)—because, in a fine-grained decomposition such as Ramchand’s, intermediate projections could be argued for—and, also, that the category selection requisites of InitP do not hold, the problem that remains is that the position of specifier of InitP requires that the constituent that occupies it gets interpreted as the causer of the event. However, the interpretation of *corredor*, “runner,” is not that there is an event of *correr*, “run,” that has caused some state or action (even though this can be conceptualized, as in *The running made all the athletes tired*). This noun is interpreted as the “someone” (or “something”) that initiates and controls an event of running. I conclude, then, that this position cannot be occupied by the verb because the semantic entailments expected from the configuration simply do not follow.

Assume, now, that the position is occupied by a nominal category, say a silent pronoun which acts as a variable; the suffix binds this position somehow (see Alexiadou and Schäffer 2008 for an analysis along these lines of English –*er*).

The question is why this pronominal category cannot be expressed by a phonologically materialized DP. Once the NP is dominated by its full set of functional projections, the DP could receive genitive case, so, following standard assumptions about the distribution of empty categories, a DP could occupy the position of pro in
the syntax. However, this never happens. I believe that it is also fair to conclude, for these reasons, that this alternative is not empirically accurate.

4.4.3 What if the position of agent is not available inside the nominalization?

To begin with, nothing prevents nominalizations from having agent-denoting complements, so agent theta roles cannot be generally excluded from all nominalizations. Assume, though, for the particular case of Spanish agent nominalizations that InitP inside a nominalization is unable to project a specifier position. This is plausible if InitP is considered as a notational variant of vP—against Ramchand (2008). If we substitute InitP for vP, then an argument could be constructed that the only vP that can occur inside nominalizations is the weak vP, not the strong vP able to assign an agent theta role. A sound reason to assume that weak vPs are the only option inside a nominalization is that in this configuration vP does not assign accusative case, as expected from the weak version of little v. The structure that would result from this analysis, with weak vP, is represented in (32).

\[
\begin{array}{c}
\text{NP} \\
\text{N} \\
vP \\
v \\
\end{array}
\]

(32)

In this account, the agent is not projected because there is no position for it inside the tree; in other words, nothing needs to occupy this position because the position is not available in any case, due to the weakness of vP in the configuration.

The problem of this approach is that it predicts that –dor can take verbs which, by their lexical properties, always combine with a weak vP phrase and, therefore, do not take an agent at all. This prediction is contradicted by the data (remember the examples in (3)). Notice also that if a verb allows two readings, one without agent and one with it, –dor allows only the agentive reading, as in (33).

(33) hervi-dor
boil-er
‘thing that makes something boil’, not ‘thing that boils’

The data actually point to more radical counterexamples. The following cases show that, not only is –dor incompatible with the non-agentive readings of possible non-agentive verbs, but also that it forces agentive readings in normally non-agentive verbs:
(34) a. entra-dor, go.in-er, “soccer player that often tries to steal the ball or person that flirts with girls,” not #“person that comes into some place.”

b. El delantero entra al defensa. Es un entrador.
   The forward tackles Acc-the defender. He.is a go.in-er.

c. Juan entra a muchas chicas en la discoteca. Es un entrador.
   Juan chats.up Acc many girls in the disco. He is a go.in-er

   Juan comes into many bars. *He is a go.in-er

As seen in the gloss in (34a), the –dor nominalization is associated with the reading of the verb in (34b) and (34c), where the initiator controls the event, but not to (34d), where the verb is unaccusative. Another example is given in (35), where the –dor nominalization takes the agentive meaning in (35b), not the inherently directional one in (35c).

(35) a. sali-dor, go.out-er, ’person that parties a lot during the night,’ not #person that goes out.’

b. Juan no puede salir hasta que no termine las clases
   Juan cannot go partying until he doesn’t finish his lessons

c. Juan sale de su cuarto.
   Juan gets out from his room.

This shows that, if the analysis treats –dor as a head external to the verbal domain, it needs to give account of the fact that it selects a piece of structure that contains the specifier interpreted as agent; weak vPs are not good enough for this suffix.9

9 The contrasts noticed in (34) and (35) have potential relevance for the analysis of unaccusative verbs in the general framework that we have assumed here. Ramchand (2008) analyzes them as containing an InitP which, in the temporal dimension, must be simultaneous with the result state. However, in light of these examples, other properties must be at play in this class of verbs, as the initiator would be present in the two different readings presented. Assuming that Ramchand’s analysis is correct, it seems that the suffix –dor is only compatible with these verbs to the extent that the initiator can conciously control the process. It is possible, then, that the unmarked reading of the unaccusative verbs contains an InitP only in a defective version, with an impoverished initiator, something which would not be too far from the traditional analysis of these verbs as not involving true causers; the rejection of –dor would come from the fact that this impoverished causer is not compatible with the suffix’s semantic entailments. I leave this problem for further research, although I believe that the contrasts warn us against positing full-fledged InitP projections in unaccusative verbs. A second interesting consequence of these examples is that the grammatical examples show that there is no problem with –dor occupying three argument positions in the verb: in the reading of entrador as a soccer player, it refers to the entity that decides to tackle the other player (initiator), the one that moves in that process (undergoer), and the one that ends up in the area occupied by the other player as a result of the process (resultee).
Baker and Vinokurova (2009) have a recent analysis which is a variation of the one presented here. The authors assume Kratzer’s (1996) proposal that the external argument is introduced by a voice projection, VoiceP. The agent suffix is a nominal version of VoiceP, semantically identical to this projection but with different category features (nominal as opposed to verbal). The agent position is not available in the structure because the affix, as VoiceP does, selects for VP, where only the internal arguments are introduced. The affix cannot select for VoiceP for the same reason that VoiceP does not select for a second VoiceP. Being nominal in nature, the affix does not introduce any argument structure, so, in a sense, the nominal materialization of VoiceP is a defective version where no agents are introduced. The conclusion is that the agent DP cannot be introduced by the verbal projections or by the nominal projection.

To my mind, Baker and Vinokurova’s analysis has to face two problems. The first one is that identifying the affix with a nominal version of VoiceP predicts that the bases with which the agent suffix can combine are those which allow passive voice (Baker and Vinokurova 2009: 530). This prediction is, however, not borne out. In (36), the suffix –dor can combine with the verb conocer, “to know” in a dynamic reading, but the predicate rejects the passive voice.

(36)  

a. conocedor de vinos  
know-er of wines, “wine connoisseur”

b. *Los vinos son (bien) conocidos por Juan.  
The wines are (well-) known by Juan.

The second problem is that the authors’ proposal that the suffix is VoiceP cannot account for the cases mentioned in (8), where the base taken by the affix is a noun, on the assumption that VoiceP always selects a verbal complement, explicitly made by this author. I will elaborate on this point in section 4.4.4.1.

Therefore, I conclude that the agent position must be present inside the structure required by –dor.

4.4.4 Extensions of the analysis: other aspects of agent nouns

In this section I will see how my analysis faces other properties of –dor that I have presented in the previous paragraphs, but not yet analyzed.

4.4.4.1 Explaining the relationship between –dor and the base: nouns and states

Our analysis explains why –dor, a suffix that cancels argument structure, requires verbal bases, as opposed to –ción, which does not cancel argument structure and can combine with bare roots without a theme vowel (remember the cases presented in (8), (10), and (12)). This is explained by the fact that –dor requires being introduced in a particular thematic position, which bare roots or noun stems do not have by themselves.

This can also explain the cases in (7), where a noun is combined with a theme vowel in order to be the base for –dor.
The configuration in (37) explains the interpretation of leñador, “woodsman,” as “someone that initiates a relation with wood;” the theme vowel -a- provides the verbal skeleton needed for the theta roles to be available; -dor expresses the Initiator and leñ-, “wood,” expresses the theme. The structure, where the initiator and the base are arguments of the same head, reflects that the two elements establish a semantic relationship, whose nature is left to encyclopedic information provided that the initiator is the causer of the relation. In this noun it seems to be “X produces Y,” but in other cases, the base represents the entity that is transferred by the initiator (agua-dor, “water-er, water-boy”), the entity that the initiator travels through (via-dor, “way-er, pedestrian, traveller”), the entity that is used as an instrument to perform a task (avia-dor, “airplane-er, pilot”), or the entity which gets affected by some potential actions (viña-dor, “vineyard-er, wine grower, vineyard owner”), among others.10

As opposed to what a theory such as Di Sciullo and Williams’ needs to assume, the structure gets assembled in the syntax, so in the lexicon the verbal stem leñ-a- does not exist. It is obtained configurationally, but if the base with which the theme vowel combines does not contain the formal properties required by a verb—among others, the possibility to get tense operators—it will never be used as a normal verb. These cases are not substantially different from other syntactic cases where a unit (here, the noun) requires another unit (here, the theme vowel) to be used in a particular construction, without the union of the two units having any special status in the lexicon (think, for example, of light verb constructions or auxiliary verbs).

The question remains of how to explain the impossibility of the verbal counterpart in a principled way. An anonymous reviewer suggests that, in order to project as a verb, a base needs to combine with projections such as VoiceP and AspectP. If a base such as leñ- is independently unable to combine with these heads (because its semantic type is not the kind that these functional projections need), then we explain why (37) is a legitimate configuration in syntax, but not its fully fledged verbal counterpart.

10 It is tempting to relate the structure of leñador with a light verb construction such as hacer leña, “to make wood,” similar in meaning. However, the equivalence is not possible in all the cases; words such as viñador, aguador, aviador, or viador do not have equivalent light verb constructions. We suggest that this absence is related to the fact that these agent nouns lack a ProcP, as shown in (37), which seems to be necessary to build the light verb construction version. In a case such as leñador, the equivalence seems possible because the encyclopedic interpretation of the base is one of result object, which conceptually allows for an interpretation with a specific action. We claim, however, that this correspondence is only conceptual, subject to the interpretation of the relation, and not structural, as shown by the majority of cases which lack an equivalence.
counterpart: the contrast comes from the independent fact that nouns do not combine with VoiceP. This situation requires further elaboration, but notice that, if it is on the right track, it constitutes a strong counterexample against Baker and Vinokurova’s (2009) analysis. If, as these authors propose, the agent suffix is a nominal version of VoiceP, formations such as leñador imply that the base leñ- is combinable with VoiceP, and then we would expect the verb *leñar to be possible in Spanish, counterfactually. On the other hand, if the presence of the affix does not imply the presence of any higher functional projection, as I propose, the absence of the verb is explained by the incompatibility between the base and this functional projection.

4.4.4.2 Agentive nominalizations and participles Part of the initial appeal of Marvin’s analysis of Slovenian nominalizations is that it attempts to explain a salient property of agentive nominalizations cross-linguistically: they seem to be morphologically built on top of the (past) participle (see also Benveniste 1948). My analysis also captures this relationship, given the nature of InitP in Ramchand’s system. In Ramchand’s system, InitP and the result phrase are substantively identical: they both denote a stative relationship between two entities. The causative state and the resultative state are differentiated configurationally by their relative position with respect to ProcessP, the head that introduces the dynamicity of the event. This has important consequences for the treatment of data. First of all, if InitP denotes a state, the presence of the participial morphology is plausibly due to the historical origins of the agent nominalization. Participles express states (Embick 2004), and, as such, they are candidates for being associated with InitP in the absence of tense heads (38).

This may be a factor in the historical motivation of the choice of participial morphology to make agent nominalizations, but notice that it runs into empirical problems in a synchronic perspective, at least in Romance languages, because the base of –dor is not always identical to the participle: the participle of correr, “run,” is corrido, but the agentive nominalization is corredor, not *corridor. Be that as it may, the association between InitP and states grounds the cross-linguistic similarity between these two morphological objects in semantic and syntactic principles.

Treating InitP as a state phrase also gives the first steps in the direction of explaining why some languages use the agentive suffix to express a variety of relationships. This is the case of English, where –er is used not only in agent-denoting nouns, but also in Londoner, two-decker, six-packer, nine-grader, etc. (see Ryder 1999, Heyvaerts 2003, Alexiadou and Schäffer 2008). These words are explained—although
its productivity is not properly restricted by this—if—er in English is merged in the specifier of a state-denoting phrase, independently of its relationship with ProcessP (39).

\[ \text{StateP} \]
\[ \begin{array}{c}
-er \\
\hline \\
\text{State} \\
\end{array} \]
\[ \begin{array}{c}
\text{State} \\
\text{Ø} \\
\text{six-pack} \\
\end{array} \]

The contemporary Spanish suffix –dor also seems to be sensitive to the relationship of the state head with process, for the suffix does not produce this type of nominalization. However, in Spanish historical morphology it has been noticed that –dor formations in Spanish gave rise to –dero nominalizations which are interpreted as passive deverbal nouns—in the sense that they refer to the entity which becomes something as a result of something—(casa-dera, “marriable” = that may end up being married) (Pascual and Sánchez 1992). This relationship between causer and result becomes understandable if in a previous stage of Spanish the lexical item –dor behaved in a way similar to modern English –er, being able to combine with both kinds of states.

4.5 Summary and conclusions

In this chapter I have provided evidence for the multidominance approach to structure building, and, in particular, for the possibility that a constituent projects its own phrase after movement. I have argued that this situation takes place in the case of nominalizer affixes that play the double role of changing the category of a structure and canceling part of the argument structure of the base. The operation is restricted to units that do not take a complement and is only available when there are no phase-defining projections inside the structure of the nominalization.

Given the empirical facts in the case of –dor nominalizations, I have shown that analyses which do not place the nominalizer affix in the agent thematic position either overgenerate or fail to account for the attested cases. I have also argued for a purely syntactic analysis of these nominalizations, to the extent that accounts which perform the operation in a generative lexicon do not predict that nominal bases can be used to derive agent nouns.

I would like to emphasize that the particular verb decomposition assumed in this chapter is independent from the syntactic operation proposed to account for agent nominalizations. My evidence in favor of a treatment which uses the InitP constituent is based on the relation between agent nouns and state-denoting morphemes (discussed in 4.4.2), but—if we decide to ignore these data—the analysis would be equally possible in a framework where the verbal domain has other ingredients. The interaction between these affixes and the functional heads contained inside the verbal domain are, thus, left for further research.
5

Clitic Placement and Multidominance

MARTINA GRAČANIN-YUKSEK

5.1 Introduction

Ever since Kayne’s (1994) Linear Correspondence Axiom (LCA), it has been a fairly common assumption in the literature that linear order of terminals in a syntactic structure is determined based on asymmetric c-command relations that hold among non-terminal nodes in the structure. The LCA, however, in its original form cannot linearize multidominance (MD) or sharing structures. This has led to a number of attempts to make the LCA compatible with MD (Citko 2005, Gračanin-Yuksek 2007, Wilder 1999, 2008). All these proposals make the claim that all and only MD structures that are linearizable are well formed. Thus, linearization emerges as a crucial factor that constrains MD. In this chapter, I argue against this view.

The argument I present proceeds as follows: first, I present evidence that in some non-MD structures, an element is pronounced so that it follows rather than precedes the material that it c-commands. The relevant examples come from the behavior of the Croatian third person singular auxiliary clitic je. Unlike other auxiliary clitics, je follows pronominal clitics in a clitic cluster, but it can be shown that its syntactic position is higher than that of pronominal clitics (Bošković 2001, Stjepanović 1998). Given this, it seems that LCA alone cannot account for the linear order of sentences containing je, which leads to a conclusion that the linear order of elements in the terminal string is to an extent independent of the structure. Rather, if we are to retain a general view that linearization is computed based on asymmetric c-command, then the cases such as the ordering of je must be handled in some post-syntactic component.

Next, I give examples which show that this problem arises in MD structures as well. I discuss two such cases: Croatian multiple wh-questions where wh-phrases seem to be coordinated at the left periphery of the clause, which I refer to as Q&Qs,
and German *Subjektlücke in finiten Sätzen* (subject lacking in finite clauses [SLF]). In relevant Q&Qs, there is an unshared element, namely the clitic *je*, which is linearized so that it follows some shared material, even though it c-commands this material in the syntactic structure. In SLF constructions, there is a shared element, the subject, linearized so that it follows some unshared material, even though it c-commands it. This again leads to a conclusion that linear order is, at least to a point, independent of the structure. If this conclusion is on the right track, then linearization cannot be the factor that determines *syntactic* well-formedness of MD structures.

If this reasoning is correct, we are left with the question: “What *does* constrain MD?” I propose that MD is constrained by a constraint which I refer to as the *Constraint on Sharing* (COSH), originally proposed in Gračanin-Yuksek (2007).

(1) **Constraint on Sharing**

If a node α has more than one mother node, but does not have a unique highest mother (a single mother of α not dominated by any of its other mothers), all the mother nodes of α must completely dominate the same set of terminal nodes.

We will see that all of the examples that are not linearizable under the asymmetric c-command approach to linearization, but are nevertheless grammatical, obey COSH. However, as it is stated, COSH is a condition that is specific to MD. We would like to derive it from principles independent of MD. Towards the end of the chapter, I present an attempt to do so.

### 5.2 Clitic *je* in non-MD structures

Croatian clitics fall into two classes: pronominal clitics and auxiliary clitics. Clitics in Croatian are second-position elements; they follow the first prosodic word or the first maximal projection in their own clause (Franks and Progovac 1994, Halpern 1995, Progovac 1996, among others). If a clause contains more than one clitic, the whole clitic cluster appears in the second position in the clause. Within the cluster, clitics appear in the order in (2), illustrated in (3).

(2) **AUX < DAT < ACC**

(3) Mi **SMO** VAM GA pokazali.
we AUX.IPL you.PL.DAT he.ACC shown
‘We showed him to you.’

Crucially for our purposes, an auxiliary clitic cannot follow a pronominal clitic.

(4) *Mi GA **SMO** vidjeli.
we he.ACC AUX.IPL seen
‘We saw him.’
The only exception to this is the third person singular auxiliary clitic *je*. Unlike all the other auxiliaries, *je* always appears following all the pronominal clitics in the cluster. This is shown in (5).

(5)  
\[ a. \] Petar \text{GA} \text{JE} \text{vido}.  
\begin{align*} 
\text{Petar} & \quad \text{he.ACC} \quad \text{AUX.3SG} \quad \text{seen} \\
\text{’Petar saw him.’} 
\end{align*}  

\[ b. \] *Petar \text{JE} \text{GA} \text{vido}.  
\begin{align*} 
\text{Petar} & \quad \text{AUX.3SG} \quad \text{he.ACC} \quad \text{seen} 
\end{align*}  

One possible explanation for the positioning of *je* in a clitic cluster is that in the syntax, *je* occupies a different (lower) position than other auxiliary clitics (Franks and King 2000, Franks and Progovac 1994, Tomić 1996). However, based on data from VP ellipsis, Stjepanović (1998) shows that the syntactic position of *je* is the same as the syntactic position of other auxiliary clitics.\(^{1}\) Assuming that in VP ellipsis the elided structure is syntactically lower than the pronounced remnant, the fact that in the environment of VP ellipsis *je* behaves the same as other auxiliary clitics indicates that it occupies an equally high syntactic position. This is shown in (6) and (7).\(^{2}\)

(6)  
\[ a. \] Mi smo mu ga dali, a i vi ste \text{weAUX.1PL} \text{he.DAT} \text{he.ACC} \text{given and also you} \text{AUX.2PL} \text{mu ga dali} \quad (također).  
\begin{align*} 
\text{he.DAT} \quad \text{he.ACC} \quad \text{given too} \\
\text{’We gave it to him, and you did too.’} 
\end{align*}  

\[ b. \] *Mi smo mu ga dali a i vi \text{weAUX.1PL} \text{he.DAT} \text{he.ACC} \text{given and also you} \text{mu ga ste dali} \quad (također).  
\begin{align*} 
\text{he.DAT} \quad \text{he.ACC} \quad \text{AUX.2PL} \quad \text{given too} 
\end{align*}  

(7)  
\[ a. \] On mi ga je dao, a i he me.DAT \text{he.ACC} \text{AUX.3SG} \text{given and also} \text{ona je mi ga dala} \quad (također).  
\begin{align*} 
\text{she} \quad \text{AUX.3SG} \quad \text{me.DAT} \quad \text{he.ACC} \quad \text{given too} \\
\text{’He gave it to me, and she did too.’} 
\end{align*}  

\[ b. \] *On mi ga je dao, a i ona \text{he me.DAT} \text{he.ACC} \text{AUX.3SG} \text{given and also she} \text{mi ga je dala} \quad (također).  
\begin{align*} 
\text{me.DAT} \quad \text{he.ACC} \quad \text{AUX.3SG} \quad \text{given too} 
\end{align*}  

\(^{1}\) Bošković (2001) presents evidence to the same effect from VP fronting, parenthetical placement, and placement of subject-oriented adverbs.  

\(^{2}\) Examples in a) are from from Stjepanović (1998), while those in b) are mine.
Assuming a high syntactic position for *je*, there are at least two ways in which we can explain its exceptional placement with respect to pronominal clitics. One is to say that all auxiliary clitics in Croatian, including *je*, are merged in the position higher than the pronominal clitics in a clitic cluster, and *je* is then placed into the position where it surfaces by some purely PF mechanism that operates on the surface string and is independent of the underlying structure. Under this view, the sentence in (5a), repeated here as (8), has the structure in (9).³

(8) Petar _GA_ _JE_ vidio.
    Petar  he.ACC  AUX,3SG  seen
       ‘Petar saw him.’

(9)

```
TP
  Petar ??
    jeAUX ??
       gaHIM  VP
          videoSEEN
```

Another possibility is to propose that all auxiliary clitics in Croatian are generated below pronominal clitics and subsequently move to a higher position. What is special about *je* is that it is pronounced in the tail rather than in the head of the chain. This solution is argued for in Bošković (2001). Bošković proposes that the placement of *je* is an instance of a more general strategy employed by languages to spell-out a lower copy in a chain whenever spelling-out the highest one leads to a PF violation.⁴ On this view, (5a)/(8) has the structure in (10). In the rest of the chapter, I will assume that this structure is correct, but the arguments presented apply equally to the structure in (9).

³ The structures in (9) and (10) are simplified reflecting the fact that I abstract away from the question of how the entire clitic cluster ends up in the second position. It may well be that clitics are adjoined to one another within the cluster. The relevant thing for us is the relative ordering of *je* and the pronominal clitics, which is problematic for the LCA.

⁴ Bošković proposes that the PF violation in the case of *je* arises due to the fact that *je* is in the process of losing its clitichood. It is sufficiently a non-clitic to block cliticization across it, but is not yet non-clitic enough to be able to provide a host for other clitics. If the lower copy of *je* is spelled out, the pronominal clitics do not have to cliticize across *je* and the problem is avoided.
Importantly, regardless of which of these explanations we adopt for the placement of *je*, we still face a problem of how to linearize the structure relying solely on the LCA: the correspondence between the asymmetric c-command and precedence is lost. This is taken as evidence that the surface order of terminals in a string is to a certain extent independent of the structure.

In the following sections, I discuss consequences of this conclusion for MD structures. As noted in 5.1, Introduction, the LCA in its original form is incompatible with MD. However, attempts have been made to reconcile the asymmetric c-command view of linearization with MD (Citko 2005, Gračanin-Yuksek 2007, to appear Wilder 1999, 2008). We will see, however, that the problem of the lack of correspondence between the asymmetric c-command and precedence discussed above re-emerges even under the modified, MD-compatible version of the LCA. The conclusion we will be forced to reach is that, even in MD environments, the linear order of terminals does not entirely depend on the structure. This will in turn be taken as evidence that linearization is not a constraining factor on MD.

The relevant structures that I will discuss are Q&Qs in Croatian and SLF in German.

### 5.3 Q&Qs in Croatian

I use the term Q&Q to refer to multiple wh-questions in which wh-phrases seem to be coordinated at the front of the clause. A simple example of a Q&Q is given in (11).

(11) Što i kada Ivan jede?
    what and when Ivan eats
    ‘What and when is Ivan eating?’

A Q&Q in Croatian can also contain clitics, which may appear after each wh-phrase, as in (12).
Following Gračanin-Yuksek (2007), I assume that in Croatian, Q&Qs like the one in (12), in which clitics follow each wh-phrase, are necessarily derived from the bi-clausal underlying structure in (13).

(13) \[ \text{[\&P [CP$_1$ WH$_1$\ldots t$_{WH_1}$] and [CP$_2$ WH$_2$\ldots t$_{WH_2}$]]} \]

A bi-clausal analysis of such Q&Qs offers a natural explanation for why they contain two (sets of) clitics: each (set) is part of its own clause, and each (set) appears in the second position in that clause, as shown in (14).\(^5\)

(14) Što će Ivan jesti i kada će Ivan jesti?
what will$_{3SG}$ Ivan eat and when will$_{3SG}$ Ivan eat

What and when will Ivan eat?

The analysis receives additional support from the fact, noted in Gračanin-Yuksek (2007), that Q&Qs with two (sets of) clitics, in which one of the wh-phrases is a direct object cannot contain an obligatorily transitive verb, such as kupiti “buy.” This is because on this view, the conjunct introduced by a wh-adjunct kada “when” does not contain a direct object, which is required by the verb.\(^6\) Thus, (12) and (14) contrast with (15) below.\(^7\)

\(^5\) Here, I use the strikethrough to indicate a non-pronunciation of material, without committing myself to an ellipsis analysis.

\(^6\) For further arguments in favor of a bi-clausal analysis of Q&Qs with repeated clitics, see Gračanin-Yuksek (2007).

\(^7\) A corresponding Q&Q that does not contain two (sets of) clitics is well formed with the verb kupiti “buy.”
(15) *Što će i kada će Ivan kupiti?
   what will.3SG and when will.3SG Ivan buy
   *What and when will Ivan buy?

If Q&Qs with repeated clitics are bi-clausal, a question arises as to how the surface form is derived from the larger underlying structure. I assume without discussion the following MD representation for bi-clausal Q&Qs, proposed in Gračanin-Yuksek (2007). In (16), the Q&Q contains two CPs which share everything except the wh-phrases (and clitics). Wh-phrases and clitics (unshared material) are pronounced within the respective conjuncts where they are merged, while the subject and the verb (shared material) are pronounced only once, following all the unshared material.

(16) &P &P
    CP1 iAND CP2
    ŠtoWHAT C_1 CP2
    kadaWHEN C_2 CP2
    ćeWILL TP1 TP2
    kadaWHEN TP2
    Ivan VP1 VP2
    jestiEAT VP2
    štoWHAT

The shared string Ivan jesti “Ivan eat” does not form a constituent to the exclusion of the lower copies of wh-phrases. Consequently, the two terminals may not be shared in bulk (i.e. at the TP level). Instead, each must be shared individually. I call this kind of sharing non-bulk sharing.

The structure can be linearized by the linearization algorithm proposed by Gračanin-Yuksek (to appear), which preserves the general antisymmetric approach to linearization and builds on proposals by Wilder (1999, 2008) in proposing modifications to the LCA which make it compatible with MD. The algorithm is summarized as follows:

---

8 For reasons of space, I do not argue here for the MD representation of Q&Qs in (16). For arguments in favor of such a structure, and against alternative analyses that might be responsible for deriving the surface string of a Q&Q from the underlying structure, see Gračanin-Yuksek (2007).
a. Linearization

If \( \alpha \) asymmetrically c-commands \( \beta \), every node completely dominated by \( \alpha \) precedes every node completely dominated by \( \beta \).

b. C-command

\( \alpha \) c-commands \( \beta \) iff \( \alpha \neq \beta \), \( \alpha \) does not dominate \( \beta \), and every highest mother of \( \alpha \) dominates \( \beta \) (where a highest mother of \( \alpha \) is a mother of \( \alpha \) not dominated by any other mother of \( \alpha \)).

c. Complete dominance (from Fox and Pesetsky, in preparation)

\( \alpha \) completely dominates \( \beta \) iff every path from \( \beta \) upwards to the root goes through \( \alpha \).

This algorithm yields the following order within CP1:

(18) CP1: što < ě < Ivan < jesti

Similarly, the algorithm computes the following order of terminals in CP2:

(19) CP2: kada < ě < Ivan < jesti

Since the conjunction \&'0 asymmetrically c-commands everything contained in CP2, we obtain the following:

(20) \&': i < kada < ě < Ivan < jesti

Next, CP1 asymmetrically c-commands \&'0, yielding (21).

(21) što ě < i

CP1 also asymmetrically c-commands CP2 and everything it dominates. Both CP1 and CP2 each completely dominate only the wh-phrase (što “what” and kada “when” respectively) and the auxiliary clitic ě “will”. This yields (22):

(22) što ě < kada ě

The ordering statements in (18) through (22) result in the unique and non-contradictory order of terminals given in (23):

(23) što < ě < i < kada < ě < Ivan < jesti

Given this result, it seems that structures along the lines of (16) are in principle linearizable. The assumed linearization algorithm allows for the shared material to remain in situ. However, it linearizes all the shared terminals so that they follow all the unshared terminals. This is a welcome result.

We will next look at Q&Qs that contain clusters of clitics.

---

9 CP does not c-command the subject and the verb, since it dominates them.
5.3 Clitic clusters in Q&Qs

In Croatian Q&Qs, each wh-phrase may be followed by a cluster of clitics, as in (24).

\[(24) \text{Što si mu i zašto si mu pjevao?} \]
\[\text{what AUX.2SG he.DAT and why AUX.2SG he.DAT sung} \]
\[‘\text{What did you sing to him and why did you sing to him?}’ \]

However, it is not necessary that both conjuncts contain both clitics. It is possible for one conjunct to contain both clitics and the other only one. The patterns of clitic distribution are illustrated in (25) and (26).

\[(25) \text{Što si mu i kada si pjevao?} \]
\[\text{what AUX.2SG he.DAT and when AUX. 2SG sung} \]
\[‘\text{What did you sing to him and when did you sing?}’ \]

\[(26) \text{Što si i kada si mu pjevao?} \]
\[\text{what AUX.2SG and when AUX.2SG he.DAT sung} \]

Reading 1: ‘What did you sing and when did you sing to him?’

Reading 2: ‘What did you sing to him and when did you sing to him?’

In (25), the dative pronominal clitic mu “him” is present only in the first conjunct and it is interpreted only in the first conjunct. This indicates that this clitic is syntactically not present in the second conjunct. The Q&Q in (25) thus has a structure as in (27).\(^{10}\)

\[(27) \]
\[\text{The only shared node in the structure is the verb } pjevao \text{ “sung,” which is pronounced following all the unshared material (wh-phrases and clitics). The auxiliary clitic, as discussed above, moves from the site where it is externally merged (labeled the auxiliary phrase) to its derived position in the clitic cluster (labeled the clitic phrase). If there is a pronominal clitic in the structure, the auxiliary “jumps over” it as it moves upwards.} \]

\(^{10}\text{I omit the null pro subjects from the representations.}\]
The situation in (26) is somewhat more complicated. The pronominal clitic surfaces only in the second conjunct, and on the first reading it is interpreted only in the second conjunct. Thus, the structure of (26) with reading one, given in (28), is in a sense a mirror image of (27). Again, the only shared node is the verb \( pjevao \) “sung,” and it follows all the unshared material in the structure.

(28)

\[
\begin{array}{c}
\text{\&P} \\
\text{CP}_1 \quad \text{i}\text{AND} \quad \text{CP}_2 \\
\text{Što}_{\text{WHAT}} \quad \text{ClP}_1 \\
\text{si}_{\text{AUX}} \\
\text{si}_{\text{AUX}} \\
\text{AuxP}_1 \\
\text{si}_{\text{AUX}} \\
\text{si}_{\text{AUX}} \\
\text{VP}_1 \\
\text{pjevao}_{\text{SUNG}} \\
\text{što}_{\text{WHEN}} \\
\text{ClP}_2 \\
\text{kada}_{\text{WHEN}} \\
\text{si}_{\text{AUX}} \\
\text{si}_{\text{AUX}} \\
\text{AuxP}_2 \\
\text{si}_{\text{AUX}} \\
\text{si}_{\text{AUX}} \\
\text{VP}_2 \\
\end{array}
\]

The interesting case is the second reading in (26). Here, the pronominal clitic is interpreted in both conjuncts, although it is only pronounced in the second one. This indicates that the clitic is shared between the conjuncts. Example (26) with reading two thus presumably has the structure in (29).

(29)

\[
\begin{array}{c}
\text{\&P} \\
\text{CP}_1 \quad \text{i}\text{AND} \quad \text{CP}_2 \\
\text{Što}_{\text{WHAT}} \quad \text{ClP}_1 \\
\text{si}_{\text{AUX}} \\
\text{si}_{\text{AUX}} \\
\text{Cl'}_1 \\
\text{mu}_{\text{HIM}} \\
\text{AuxP}_1 \\
\text{si}_{\text{AUX}} \\
\text{si}_{\text{AUX}} \\
\text{VP}_1 \\
\text{pjevao}_{\text{SUNG}} \\
\text{što}_{\text{WHEN}} \\
\text{ClP}_2 \\
\text{kada}_{\text{WHEN}} \\
\text{si}_{\text{AUX}} \\
\text{si}_{\text{AUX}} \\
\text{AuxP}_2 \\
\text{si}_{\text{AUX}} \\
\text{si}_{\text{AUX}} \\
\text{VP}_2 \\
\end{array}
\]
5.3.2 Q&Qs with clitic clusters containing je

What happens when the auxiliary clitic in the clitic cluster is je? First, note that the counterpart of (26) with the third person singular clitic je shows the same ambiguity, as shown in (30).

\[ (30) \quad \text{Što je i kada mu je pjevao?} \]

Reading 1: ‘What did he sing and when did he sing to him?’

Reading 2: ‘What did he sing to him and when did he sing to him?’

The availability of reading two in (30) is again an indication that the pronominal clitic mu “him” may be shared between the conjuncts. Thus, on this reading, the Q&Q receives the representation in (31).

\[ (31) \quad \text{Što je i kada mu je pjevao?} \]

In line with Bošković’s proposal about the movement of the auxiliaries in (Serbo-) Croatian, we posit the syntactic movement of je across the shared pronominal clitic in each conjunct, but phonology is instructed to spell-out the lower rather than the higher copy, as indicated in (31) by the strikethrough.\(^\text{11}\) In particular, in the second conjunct, je must follow the pronominal clitic mu “him.”

This order, however, is non-derivable by our assumed algorithm. In the second conjunct in particular, the clitic je\(_2\) c-commands the pronominal clitic mu, since it is true that every highest mother of je\(_2\), and there is only one (ClP\(_2\)), dominates mu “him.” Thus, the algorithm predicts that je\(_2\) should precede mu.

\(^{11}\) Alternatively, as mentioned above, je might be merged only in the higher position, and not undergo movement. It would then be placed in the position following mu “him” by some PF rule. This would still be problematic for the LCA approach to linearization.
On the other hand, the pronominal clitic \textit{mu} “him” c-commands neither of the auxiliary clitics (\textit{je}_1, \textit{je}_2), either in their base positions, or in their derived positions. In order for \textit{mu} “him” to c-command \textit{je}_2, it would have to be the case that every highest mother of \textit{mu} “him” dominates \textit{je}_2. This is clearly not the case for the derived position of \textit{je}_2. It is also not the case for the base position of \textit{je}_2, given that \textit{mu} “him” has two highest mothers, \textit{Cl'}_1 and \textit{Cl'}_2. While \textit{Cl'}_2 does dominate \textit{je}_2, \textit{Cl'}_1 does not. We can conclude more generally that a shared node with more than one highest mother never c-commands an unshared node. Consequently, a shared node should never precede an unshared node. And yet, in the second conjunct of (30), \textit{mu} “him” precedes \textit{je}_2.

This discrepancy between asymmetric c-command relations that hold in the structure and the linear order of terminals in the final string again point to the conclusion that the structural relations among the non-terminal nodes in the structure are not all that is responsible for the linear order of terminals. This is the conclusion that we have reached in section 5.3, where we looked (in less detail) at a non-MD structure containing \textit{je}. Some mechanism (partially) independent of syntax must be involved.

Crucially, it is not the case that the structure in (31) is not linearizable at all. The linearization algorithm operating on (31) yields a unique, total, and non-contradictory order in (32). It is just that this order happens not to be attested.

\begin{equation}
Što < je_1 < i < kada < je_2 < mu < pjevao
\end{equation}

We seem to be left with the situation where, if we posit the syntactic movement of \textit{je} in (31) on reading two, the structure can only be incorrectly linearized as (32).\footnote{The same result obtains if we assume that \textit{je} occupies only the position higher than the pronominal clitic(s).} So, perhaps positing this movement is wrong after all. Obviously, if \textit{je} did not move over \textit{mu} (or were not merged above \textit{mu}), it would not c-command \textit{mu}, and consequently would not have to precede it. This possibility is illustrated in (33).

\begin{equation}
Što_{\text{WHAT}} & \text{P} \\
\text{CP}_1 & \text{i}_{\text{AND}} & \text{CP}_2 \\
Što_{\text{WHAT}} & \text{CIP}_1 & \text{CIP}_2 \\
\text{mu}_{\text{HIM}} & \text{AuxP}_1 & \text{AuxP}_2 \\
\text{je}_1_{\text{AUX}} & \text{je}_2_{\text{AUX}} & \text{VP}_1 \\
\text{pjevao}_{\text{SUNG}} & \text{Što}_{\text{WHAT}} & \text{VP}_2
\end{equation}
However, despite the appeal of this possibility, it would not solve the problems associated with the linearization of (31).\(^{13}\) In fact, such a structure cannot be mapped onto any linear order at all. In (33), the unshared auxiliary clitics \(je_1\) and \(je_2\) do not (asymmetrically) c-command the shared clitic \(mu\). However, as we saw above, neither does \(mu\) c-command either \(je_1\) or \(je_2\). Recall from above that a shared node with more than one highest mother never c-commands an unshared node, regardless of the structural position of either of them. Consequently, \(mu\) does not c-command \(je\). Since the order of the two clitics cannot be deduced from any other asymmetric c-command relation in the structure, the whole representation is non-linearizable.

We thus have the following: the Q&Q in (30) is grammatical with reading two. We examined two plausible structures that might underlie this reading. One is linearizable, (31), but the computed order is unattested, and the other, (33), cannot be linearized at all. Given that (33) also runs into problems in accounting for the VP-ellipsis facts discussed in section 5.2, it seems warranted to dismiss it as a possible syntactic representation of (30). We are thus left with (31).

The fact that a legitimate output of the linearization algorithm yields an ill-formed sentence indicates that there is no absolute correlation between syntactic well-formedness and linearizability. Consequently, linearization cannot be what constrains the possible range of phrase markers in human language. In other words, it seems not to be the case that only those phrase markers in which all terminals can be linearized based on asymmetric c-command are legitimate outputs of a syntactic computation.

This is equally true of both non-MD and MD representations. However, it is particularly worrying in consideration of MD structures, since these are the ones where the possibilities of over-generation are literally countless. We thus need a condition which will constrain the possible range of MD representations.

Before going on to propose a possible condition of this sort, I would like to discuss another example of a structure in which the shared node is spelled-out in the position not predicted by the linearization procedure. The case in point is a phenomenon from German, which has been called SLF (\textit{Subjektlücke in finiten Sätzen} – “subject lacking in finite clauses”).

5.4 German \textit{Subjektlücke in finiten Sätzen}\(^{14}\)

The term SLF refers to “coordinations of V2 clauses that contain only one subject” (Mayr and Schmitt 2008). An example of the SLF is given in (34).

\(^{13}\) We would also have a hard time explaining the VP ellipsis possibilities pointed out by Stjepanović (1998), discussed in section 5.2.

\(^{14}\) The data and the analysis presented come from Mayr and Schmitt (2008).
Hans hat die Katze gestreichelt und wird jetzt den Hund füttern.

Hans has the cat stroked and will now the dog feed

‘Hans stroked the cat and will now feed the dog.’

Interestingly, SLF examples allow for an asymmetric extraction of material from one conjunct only, in the apparent violation of Coordinate Structure Constraint, as shown in (35).

Die Katze hat Hans t_i gestreichelt und wird jetzt den Hund füttern.

the cat has Hans stroked and will now the dog feed

‘The cat, Hans stroked and will now feed the dog.’

Mayr and Schmitt (2008) propose that in (35), the subject Hans is shared between the conjuncts and undergoes covert QR to a position above the coordination. If we assume single output syntax, this is equivalent to saying that the subject moves to a position above the coordination, and that its lower copy is spelled-out. This is illustrated in (36).

One piece of evidence that Mayr and Schmitt show in favor of the movement of the subject comes from quantificational subjects such as no one. They show that in (37) the existential must scope below both negation and the modal. This indicates that niemand “no one” may not undergo QR from its base position.

---

15 See Mayr and Schmitt (2008) for arguments that conjuncts are Cs.
Die Katze darf niemand schlagen. (¬ ◊ < ∃), *(¬ ∃ < ◊), #(◊ < ¬∃)
the cat may no-one hit
‘No one may hit the cat.’

Since the QR of the subject is prohibited for independent reasons, the SLF in (38) is ill formed. This indicates that in grammatical SLF constructions, the subject indeed undergoes QR, i.e. that the representation in (36) is on the right track.

Die Katze darf niemand schlagen und muss sich danach hinlegen.
the cat may no-one hit and must refl after lie-down
‘No one may hit the cat and must afterwards lie down.’

The situation in (36) is in a sense a mirror image of the situation in (31). The subject Hans is pronounced only once, but is interpreted in both conjuncts—an indication that it is shared. The structure is non-linearizable if the shared subject remains in situ, since from this position it neither c-commands nor is c-commanded by the unshared material contained in either vP. Thus, no order can be established between the subject and the vP-internal material. Evidence from quantificational subjects indicates that the subject in fact occupies a high syntactic position which is outside of the coordination. From this position it c-commands all the material within both conjuncts.

However, if the movement of the subject is posited, we would expect it to be linearized so that it precedes the rest of the sentence, rather than to be sandwiched between the auxiliary and the vP in the first conjunct. This is not what we find. Thus, German SLF constructions are another case where the syntactic MD structure needed to capture the semantic properties of the sentence seems to be well formed, even though the result of the linearization procedure that operates on this structure, while in principle derivable, is unattested. Yet again, we see the absence of the correlation between the syntactic well-formedness and linearization. Given this observation, we can again conclude that linearization is not the crucial factor that constrains possible MD representations.

However, MD must be constrained by something, since it is not the case that any MD structure that may in principle be generated by syntax is well formed. In the next section, I propose and discuss a condition that derives this result.

5.5 Constraint on Sharing

One possible candidate for the constraining factor on MD is the Constraint on Sharing (COSH), proposed in Gračanin-Yuksek (2007). An informal definition of COSH is given in (39), repeated from (1).
(39) **Constraint on Sharing (COSH)**

If a node \( \alpha \) has more than one mother node, but does not have a unique highest mother (a single mother of \( \alpha \) not dominated by any of its other mothers), all the mother nodes of \( \alpha \) must completely dominate the same set of terminal nodes.

Recall the definition of complete dominance from (17c):

(40) **Complete dominance** (from Fox and Pesetsky, in preparation)

\( \alpha \) completely dominates \( \beta \) iff every path from \( \beta \) upwards to the root goes through \( \alpha \).

COSH predicts the well-formedness of any structure which is *in principle* linearizable by the antisymmetric approach to linearization, regardless of whether the derived word order is attested or not. To see how this obtains, we need to determine when the multiple highest mothers of a shared node completely dominate the same set of terminal nodes. In fact, this is true only when the relevant sets are empty. This in turn may come about in two situations. One is when all the terminal nodes dominated by the multiple highest mothers of a shared node are themselves shared. This is illustrated by the abstract representation in (41).

(41) In (41), Y and W are the relevant shared nodes. Multiple highest mothers of Y are Z and Q. Z dominates terminal nodes \( y \) and \( w \), but it completely dominates neither of them, since it is not the case that every path from either Y or W upwards to the root (A) contains Z (there is an alternative path that contains Q, but not Z). Similarly for Q, there is a path from both \( y \) and \( w \) to the root that contains Z, but not Q. Thus, the set of terminal nodes completely dominated by both Z and Q is empty. The same reasoning applies to the multiple highest mothers of W, namely R and H. Since all highest mothers of any shared node in (41) completely dominate the same set of terminal nodes, namely the empty set, the structure does not violate COSH.
The other way in which a structure that contains shared nodes which do not have a unique highest mother can satisfy COSH is when the nodes dominated by the multiple highest mothers of a shared node move to a position higher than the highest shared node. This is illustrated in (42).

In (42), we have the same shared nodes, Y and W. Let us consider W. It has two highest mothers, R and H. Both of these nodes dominate the terminal w, but not completely, as discussed above. However, R dominates the unpronounced copy of V (seemingly completely) and H dominates the unpronounced copy of K (also seemingly completely). Thus, the set of terminals completely dominated by R seems to be \{v\}, while the set of terminals completely dominated by H seems to be \{k\}. Since COSH does not make reference to overt terminal nodes, but to all terminal nodes, it seems that (42) violates COSH.\(^{16}\)

I would like to claim that this is, in fact, not the case. To this end, I assume, together with Engdahl (1986), Frampton (2004), Gartner (1999, 2002), Kracht (2001), Starke (2001), de Vries (2007a) among others that Internal Merge (or Move) does not involve creating a copy and (subsequently?) moving the original element, but rather re-merging the same element into a new position, creating multiple occurrences, rather than multiple copies of the “moved” element. Under this assumption, the structure in (42) is better represented as in (43).

\(^{16}\) Consequently, Q&Q representations in (27), (28), (29), and (31) would also seem to violate COSH, since in each of them there is at least one shared node whose multiple highest mothers do not dominate the same set of terminals. Namely, there is always a situation where one of the mothers dominates a copy of wh\(_1\), and the other a copy of wh\(_2\).
On this view, neither R nor H completely dominates anything. R no longer completely dominates v, since it is not the case that every path from V to the root includes R. There now exists a path from V to A that traces the dotted line, which does not include R. For the same reason, H no longer completely dominates k. Consequently, COSH is satisfied.

According to COSH, both structures we have discussed above, the Q&Q in (31) and the SLF in (36), are well formed. In (36), the only shared node, the subject Hans, has a unique highest mother (the highest CP), so the structure trivially satisfies COSH. In (31), on the other hand, there are two shared nodes: the pronominal clitic mu “him” and the verb pjevao “sung,” and neither has a unique highest mother. COSH requires that every mother of each shared node completely dominate the same set of terminal nodes. For highest mothers of the verb, this is satisfied given that they each dominate only the shared verb and an occurrence of the wh-phrase which has “moved” to a position higher than the highest shared node. For highest mothers of mu “him,” the condition is again satisfied, since they dominate the pronominal clitic and the verb (both shared) and an occurrence of the auxiliary clitic je, which has moved to a position higher than all shared material. Thus, neither of the multiple highest mothers of any shared node completely dominates anything.

A question now arises as to what forces the effects of COSH? In other words, can COSH be derived from a more basic set of principles? In particular, can it be divorced from MD per se? I believe that this is possible, and that the place to look for answers to these questions is the LCA itself. Below is the definition of the LCA from Uriagereka (1998):

(44) Linear Correspondence Axiom
A category $a$ precedes a category $\beta$ iff (a) $a$ asymmetrically [c-]commands $\beta$ or (b) $\gamma$ precedes $\beta$ and $\gamma$ dominates $a$. (p. 200)
By tying linear order to structural relations (c-command) that hold in a syntactic tree, LCA in effect constrains a possible range of phrase markers in human language. Namely, only those phrase markers in which all terminals can be linearized based on asymmetric c-command are legitimate outputs of a syntactic computation. I believe that this is both correct and incorrect. Let me explain what I mean by this.

We have seen that some MD representations seem to be well formed, and allowed by COSH, even though their word order is not predicted by the (modified) LCA. COSH is thus independent of the actual linearization of any particular structure. This indicates that the well-formedness requirement is not tied to the PF interface, as the LCA leads us to believe. Rather, the constraint seems to be syntactic in nature. On the other hand, COSH-compliant structures are those that are in principle linearizable by the LCA. This points to a conclusion that the LCA is correct in stating that nodes in a possible phrase marker must stand in certain structural relations to one another.

What is required, then, is what I call Transcendence, as defined in (45), which is derived from (44).

(45) Transcendence
A node \( a \) transcends a node \( \beta \) iff (a) \( a \) asymmetrically c-commands \( \beta \), or (b) \( \gamma \) asymmetrically c-commands \( \beta \) and \( \gamma \) completely dominates \( a \).

COSH rules out all representations in which transcendence does not obtain, without making any claims about the linear order of terminals onto which these representations map. Note also that transcendence is a requirement that holds equally of non-MD and MD representations. The MD-specific nature of COSH is thus dispensed with.

5.6 Conclusion
What acts as a constraining factor on MD is an important question in contemporary syntactic theory if MD is to be considered a legitimate part of grammar. In recent years, the fact that an increasing number of authors successfully adopt MD to account for various cross-linguistic phenomena seems to indicate that the question is worth exploring (Bachrach and Katzir 2009, Kasai 2007, van Riemsdijk 2006a, de Vries 2007a, Wilder 2008, to name but a few). It is clear that some constraints on MD must be in place, because otherwise MD would lead to massive generation of unattested sentences. Another consideration that is at the heart of the discussion of MD is how MD structures are linearized. It has been claimed in the literature (Citko 2005, Graćanin-Yuksek 2007, Wilder 1999, 2008) that the answer to the latter question provides the answer to the former, namely, that what constrains MD is linearization. In particular, according to these proposals, well-formed MD representations are those that can be linearized by the (modified) LCA. In this chapter I
argued against this claim by examining the placement of the third person singular auxiliary clitic je in Croatian Q&Qs, and the placement of the subject in German SLF constructions.

I first showed that the LCA runs into problems in linearizing Croatian non-MD structures that contain the third person singular clitic je. Unlike other auxiliary clitics in the language, je follows rather than precedes pronominal clitics in a clitic cluster, even though it can be shown to occupy a syntactic position which is higher than that of the pronominal clitics. Consequently, the LCA was shown to predict a wrong word order.

I then introduced Q&Qs, structures where two wh-phrases seem to be coordinated at the front of the clause, and showed that each wh-phrase may be followed by a second-position clitic. I adopted an MD structure for a Q&Q in which the two CP conjuncts share everything except the wh-phrases and repeated clitics. The structure was shown to be linearizable by an algorithm based on the LCA, but compatible with MD, which computes the linear order of terminals based on the asymmetric c-command relations among the non-terminals.

Next, I presented data from Q&Qs where the wh-phrase in the first conjunct is followed by an auxiliary clitic only, while the wh-phrase in the second conjunct is followed by both the auxiliary clitic and the pronominal clitic. Interestingly, the pronominal clitic, which surfaces only in the second conjunct, may be interpreted in both conjuncts, indicating that it is shared. Crucially, this reading was shown to be available even for the Q&Qs in which the auxiliary clitic is je. This was taken as evidence that je occupies the same syntactic position as other auxiliaries; a position which is structurally higher than that of the pronominal clitics. I adopted the analysis proposed in Bošković (2001), that je, like all other auxiliary clitics in (Serbo-)Croatian, originates in a position lower than the pronominal clitics and subsequently moves across them. It ends up following the pronominal clitics because it is spelled-out in the tail of the movement chain, rather than in the head.

However, we saw that the linearization algorithm I adopted cannot map the structure onto the correct linear string. In particular, it was impossible for the Q&Q to be linearized so that the pronominal clitic in the second conjunct precedes the auxiliary clitic je. Moreover, it was shown that if the “covert” movement of je were not posited, the structure would not be linearizable at all. Given the fact that a well-formed Q&Q structure could not be mapped onto the correct linear order, I concluded that linearization, and in particular the approach to linearization that builds on Kayne’s (1994) LCA, is not the constraining factor on MD.

Finally, I showed that the problem of the placement of je in Croatian Q&Qs is replicated in German SLF constructions. I adopted Mayr and Schmitt’s (2008) analysis of SLF, on which the subject is shared between the two conjuncts and undergoes a covert QR to a position higher than the coordination phrase. I showed that the actual word order of such constructions, in which the subject surfaces
between the auxiliary and the vP material in the first conjunct only, is not derivable. This again led to the conclusion that syntactic well-formedness of an MD structure is independent of linearization, arguing against the claim that linearization is what constrains MD.

I suggested that well-formedness of both structures may be accounted for if we adopt a constraint along the lines of COSH, proposed in Gračanin-Yuksek (2007). COSH is formulated as a syntactic constraint that requires all multiple highest mothers of a shared node to completely dominate the same set of terminal nodes. I proposed that this condition derives from a requirement, which I called transcendence, that any two nodes in a syntactic structure either stand in asymmetric c-command relation, or one of them be completely dominated by a node that asymmetrically c-commands the other.

If the reasoning presented in this chapter is correct, we may have ended up with more questions than answers. For example, the question remains how syntactic structures, and in particular MD structures, are linearized at all. If we keep the antisymmetric approach to linearization, it must be amended by a set of post-syntactic PF rules which should apply in strictly defined environments. An alternative is to propose an algorithm which is entirely independent of asymmetric c-command (perhaps along the lines of de Vries 2009b). Another question is what forces the requirement of transcendence, if it is not linearization, as I have argued. Is this an (LF) interface requirement, is it part of universal grammar, or is it tied to more general principles of efficient computation (Chomsky 2007)? While not providing answers to these questions, I hope that the arguments presented in the chapter might at least provide future research with a stepping stone in the right direction.
Sideward Movement: Triggers, Timing, and Outputs*

JAIRO NUNES

6.1 Introduction

Within the GB (Government and Binding) model, sideward movement—i.e. movement from a syntactic tree K to another syntactic tree L independent from K—is not a theoretical possibility. D-Structure provides the computational system with a single root tree and all the syntactic computations after D-Structure must operate within this root syntactic object. Thus, it is not at all surprising that sideward movement was not explored in GB. However, the theoretical framework that prevented sideward movement in GB becomes completely different within Minimalism and it is worth discussing whether it should still be prevented and at which cost.

Let us focus on two major differences between GB and Minimalism that bear on this issue. First, D-Structure is dispensed with for not being an interface level. Generalized transformations are then revived, allowing the computational system to operate with more than one root syntactic tree at a time. This is in fact trivially true for the first steps of any syntactic derivation. Take the first steps of derivation of (1) below, for instance. Given the (simplified) numeration in (2), the computational system independently selects saw and her and then merges them, as shown in (3). Crucially, saw and her in (3b) are both root syntactic objects.

* For valuable comments and suggestions, I would like to thank the audience of Ways of Structure Building (University of the Basque Country, 13-14/11/2008), Myriam Uribe-Etxebarria, Vidal Valmala, Anna Snarska, and two anonymous reviewers. I would also like to thank FAPESP (2006/00965-2), CNPq (302262/2008-3), and Ikerbasque (Research Group GIC07/144-IT-210-07, Research Nets in Humanities HM-2009-1-1) for their support.
(1) The boy saw her

(2) \( N = \{\text{the, boy, saw, her}\} \)

(3) a. \( N' = \{\text{the, boy, saw, her}\} \)
   K = saw

   b. \( N'' = \{\text{the, boy, saw, her}\} \)
   K = saw
   L = her

   c. \( N''' = \{\text{the, boy, saw, her}\} \)
   M = [saw her]

Another case where the computational system must deal with more than one root syntactic object at a time involves complex specifiers or complex adjuncts. Consider again the derivation of (1). Chomsky (1995b) has argued that the computational complexity of syntactic derivations can be substantially minimized if we assume the extension condition, which requires that projecting operations work at the root node. Given the derivational step in (3c), for instance, the extension condition excludes the continuation in (4), where boy merges with M and then the merges with the nonroot syntactic object boy (cf. (4c–d)).

(4) a. \( N'''' = \{\text{the, boy, saw, her}\} \)
   M = [saw her]
   O = boy

   b. \( N''' = \{\text{the, boy, saw, her}\} \)
   P = [boy [saw her]]

   c. \( N'''' = \{\text{the, boy, saw, her}\} \)
   P = [boy [saw her]]
   Q = the

   d. \( N''' = \{\text{the, boy, saw, her}\} \)
   R = [[the boy] [saw her]]

Instead, the extension condition enforces the continuation in (5) below, where boy and the are selected and merged (cf. (5a–c)) before the resulting object merges with M. Crucially, at the derivational step in (5b), there are three root syntactic objects in the derivational workspace.
The second difference between GB and Minimalism relevant for our purposes is the copy theory of movement, which reinterprets Move as the output of the interaction between the more basic operations Copy and Merge. The adoption of the copy theory is motivated by the attempt to eliminate non-interface levels (see Chomsky 1993), as well as the attempt to reduce the theoretical apparatus by only assuming syntactic primitives that can be understood in terms of (a rearrangement of features of) lexical items (Chomsky’s (1995b) inclusiveness condition). Under the copy theory, the derivation of a sentence such as (6) below, for instance, proceeds along the lines of (7), where the computational system creates a copy of John, merges it with the previously assembled TP, and deletes the lower copy in the phonological component.\(^1\) Again, notice that in a system that has Copy as a basic operation, it must be the case that the computational system must be able to handle more than one root syntactic object, namely, the copy newly created and the root syntactic object containing the replicated material (cf. (7b)).

(6) John was arrested.

(7) a. \(K = [\text{TP was arrested John}]\)

   b. Copy:
      \[K = [\text{TP was arrested John}^1]\]
      \[L = \text{John}^1\]

   c. Merge:
      \[M = [\text{TP John}^1 \text{ was arrested John}^1]\]

   d. Delete:
      \[P = [\text{TP John}^1 \text{ was arrested } \text{John}^1]\]

\(^1\) Henceforth, superscripted indices will annotate copies.
What is relevant for our discussion is that if the computational system can operate with more than one root syntactic object at a time, and if movement is understood as the interaction between the basic operations of Copy and Merge, sideward movement becomes a logical possibility within the system. That is, given two root syntactic objects K and L, the computational system may copy a from K and merge it with L, as illustrated in (8).

(8) a. \[ K = [\ldots a \ldots] \]
   \[ L = [\ldots] \]

b. Copy:
   \[ K = [\ldots a^i \ldots] \]
   \[ L = [\ldots] \]
   \[ M = a^i \]

c. Merge:
   \[ K = [\ldots a^i \ldots] \]
   \[ P = [a^i [L \ldots]] \]

Terminological metaphors aside, note that there is no intrinsic difference between the “upward” movement seen in (7), for instance, and the “sideward” movement sketched in (8) with respect to the computational tools employed. In both cases, we have trivial applications of movement, viewed as Copy plus Merge. Sideward movement is therefore not a novel operation or a new species of movement. This point is worth emphasizing, as it has been consistently misunderstood. The fact that \( \text{Æ} \) in (8) does not merge with the structure that contains the “source” of the copy, as opposed to \( \text{John} \) in (7), may have independent explanations. First, (7) differs from (8) in an obvious way: the copy of \( \text{John} \) in (7) has only one syntactic object to merge with, whereas the copy of a in (8) has two. But more importantly, it may be the case that Last Resort licenses merger of the copy of a in (8) with L but not with K. The derivation of V-to-T movement under the sideward movement analysis sketched in (9) illustrates this point.\(^2\)

(9) a. \[ [\text{VP} \ldots \text{V} \ldots] \]
   \[ \text{T} \]

b. Copy:
   \[ \text{VP} = [\ldots \text{V}^i \ldots] \]
   \[ \text{T} \]
   \[ \text{V}^i \]

c. Merge (by adjunction):
   \[ \text{VP} = [\ldots \text{V}^i \ldots] \]
   \[ K = [\tau 0 \text{V}^i [\tau 0 \text{T}]] \]

If T and VP had merged in (9a), yielding \([TP \ T \ VP]\), the extension condition should then prevent the verb from adjoining to T, as T would no longer be a root syntactic object (cf. (4c–d)). However, V-to-T adjunction can comply with the extension condition if it proceeds as in (9b–d), with copying of V preceding merger of the two-segment T\(^0\) with VP. Crucially, once the derivational step in (9b) is reached, the copied V must merge with T rather than VP, as V arguably has features to check with T, but not VP.\(^3\)

In other words, sideward movement looks outlandish only if we examine it wearing GB lenses. If we wear Minimalist lenses instead, we realize that it is a mere label for a specific sequence of Copy and Merge, which arises as a natural consequence of the interaction among core architectural features of the Minimalist Program, namely, the abandonment of D-Structure, the copy theory of movement, and the extension condition. It is worth noting that these architectural features are in turn conceptually grounded on the Minimalist attempt to eliminate non-interface levels and reduce the number of primitives and the computational complexity of syntactic derivations. Thus, from a minimalist approach adopting these architectural features, sideward movement comes for free and does not increase the grammatical apparatus. In fact, one would need to complicate the system in order to exclude it.


\(^3\) Given that in (9b) V is a copy of the head/label of VP, merger between them would require that selfchecking be allowed, yielding massive overgeneration.
section 6.6, I show that sideward movement may also apply in the morphological component. Finally, section 6.7 offers some concluding remarks.

### 6.2 Some examples of sideward movement

In this section I present two analyses employing sideward movement: Nunes’s (1995, 2001, 2004) analysis of parasitic gaps and Hornstein’s (1999, 2001) analysis of adjunct control. The purpose of this presentation is just to familiarize the reader with the general mechanics of sideward movement approaches. I leave the discussion of more technical details to sections 6.3 and 6.5 below.

Nunes (1995, 2001, 2004) argues that parasitic gap constructions constitute empirical instantiations of sideward movement. A parasitic gap construction such as (10) can be analyzed along the lines of (11).4

(10) [which paper], did you file t, after John read PG?

\begin{align*}
(11) & \text{a. } K = [\text{John read } [\text{which paper}]] \\
& \quad L = \text{file} \\
& \text{b. } K = [\text{John read } [\text{which paper}]] \\
& \quad L = \text{file} \\
& \quad M = [\text{which paper}]^i \\
& \text{c. } K = [\text{John read } [\text{which paper}]] \\
& \quad P = [\text{file } [\text{which paper}]] \\
& \text{d. } K = [\text{John read } [\text{which paper}]] \\
& \quad Q = [\text{you } [\text{file } [\text{which paper}]]] \\
& \text{e. } [\text{vP } [\text{vP you } [\text{file } [\text{which paper}]]] [\text{after John read } [\text{which paper}]]] \\
& \text{f. } [\text{did } [\text{TP you T } [\text{vP you file } [\text{which paper}]] [\text{PP after John read } [\text{which paper}]]]] \\
& \text{g. } [[\text{which paper}]^i \text{ did } [\text{TP you T } [\text{vP you file } [\text{which paper}]] [\text{PP after John read } [\text{which paper}]]]] \\
& \text{h. } [[\text{which paper}]^i \text{ did } [\text{TP you T } [\text{vP you file } [\text{which paper}]] [\text{PP after John read } [\text{which paper}]]]]
\end{align*}

After the derivational step in (11a) is reached, the computational system makes a copy of \textit{which paper} (cf. (11b)) and merges it with \textit{file} (cf. (11c))—an instance of sideward movement. Further computations involve the building of the matrix \textit{vP} (cf. (11d–e)) and the matrix \textit{CP} (cf. (11f)). After standard wh-movement takes place in (11g), the higher copy of \textit{which paper} forms a distinct chain with each of the lower copies,

---

4 Throughout the chapter irrelevant details will be omitted for the sake of presentation.
capturing the fact that the two object positions are interpreted as bound by the wh-
phrase in the matrix [Spec,CP]. Deletion of these lower copies in the phonological
component (cf. (11h)) finally yields the parasitic gap construction in (10).

A similar derivation is employed by Hornstein (1999, 2001) to account for adjunct
control. Based on the syntactic distribution and semantic interpretation of obligato-
riely controlled PRO, Hornstein argues that it is a trace/copy left by movement to a
themtic position (see Boeckx, Hornstein, and Nunes 2010 for detailed discussion).
Crucially, the controlled PRO of adjunct clauses is no exception. More specifically,
Hornstein proposes that adjunct control involves sideward movement. An adjunct
control construction such as (12), for instance, can be derived as in (13).

(12) \[ TP \text{John}_i \ [v_P \ [v_P \text{t}_i \text{greeted everybody}] \ [\text{before PRO}_i \text{leaving the room}]\] \]

(13) a. K = [\text{John} leaving the room]
L = [greeted everybody]
b. K = [\text{John}^i leaving the room]
L = [greeted everybody]
M = \text{John}^i
c. K = [\text{John}^i leaving the room]
P = [\text{John}^i greeted everybody]
d. \[v_P \ [v_P \text{John}^i \text{greeted everybody}] \ [\text{before John}^i \text{leaving the room}]\]
e. \[ TP \text{John}^i \ T \ [v_P \ [v_P \text{John}^i \text{greeted everybody}] \ [\text{before John}^i \text{leaving the room}]\]
f. \[ TP \text{John}^i \ T \ [v_P \ [v_P \text{John}^i \text{greeted everybody}] \ [\text{before John}^i \text{leaving the room}]\]

Given the syntactic objects K and L in (13a), the computational system makes a copy
of \text{John} from K (cf. (13b)) and merges it with L (cf. (13c)), an instance of sideward
movement that allows the external \(\theta\)-role of the matrix clause to be discharged. After
before merges with K and the resulting PP adjoins to \(v_P\) (cf. (13d)), the subject moves
to [Spec,TP] (cf. (13e)), the lower copies of \text{John} are deleted in the phonological
component (cf. (13f)), and the structure surfaces as (12).

In section 6.1 we saw that sideward movement makes it possible for head adjunc-
tion to comply with the extension condition. In this section we have seen that
sideward movement also provides a straightforward analysis for multiple gap con-
structions where, descriptively speaking, an expression appears to be simultaneously
moving from more than one position as more than one chain is formed. In the case of
the parasitic gap and adjunct control constructions discussed above, for instance, it
looks as if in (10), which paper has moved from the two object positions and in (12),
\text{John} in the matrix [Spec,TP] has moved from the matrix [Spec,\(v_P\)] and the embedded
subject position. From standard Move-based approaches, this is simply not a
possibility. However, once Move is reinterpreted as Copy plus Merge, the derivation of multiple gap constructions such as parasitic gap and adjunct control constructions is not different from standard instances of movement. That is, like the derivation of standard “upward” movement, the derivation of parasitic gap and adjunct control constructions also involves applications of Copy and Merge. The only (irrelevant) difference is that in instances of sideward movement, the copy created merges not with the syntactic object that contains the source of the copying, but with another root syntactic object that is available to the computational system. It is, therefore, an empirical virtue of approaches that do not enrich the computational apparatus by (explicitly or implicitly) excluding sideward movement that they can provide a uniform treatment for movement operations that result in single or multiple chains.

Of course, one must also show that applications of Copy and Merge yielding sideward movement do not overgenerate. But this is no different a task than what must be done with respect to “upward” movement. I show below that the same conditions that block unwanted instances of upward movement can be used to prevent overgeneration in the case of sideward movement.

6.3 Preventing overgeneration

Once Move is reinterpreted in terms of Copy and Merge, all the conditions that were taken to regulate Move should accordingly be understood as holding of Copy, Merge, or the (chain) relation established among the copies. Note that this should be so regardless of whether we are dealing with upward or sideward movement. For instance, if applications of Move were required to satisfy Last Resort and Minimality, so is the interaction between Copy and Merge. Below I show how conditions that were taken to constrain Move can be used to rule out unwanted instances of sideward movement.

6.3.1 Last Resort

Exploring general least effort guidelines, Chomsky (1995b) proposes that every syntactic operation must be motivated, that is, every syntactic operation must be subject to Last Resort. Moreover, in consonance with the general attempt to reduce the computational complexity of derivations, Last Resort must be computed in a local manner (see e.g. Collins 1997). With this in mind, let us consider the contrast in (14), which illustrates the well-known fact that parasitic gaps can be licensed by arguments but not by adjuncts (see e.g. Postal 1993).

(14) a. [which paper]i did you file t1 after John read PG?
   b. *howi did Deborah cook the pork t1 after Jane cooked the chicken PG?
From a sideward movement approach, the derivation of (14a) involves sideward movement of *which paper*, as shown in (15) below, and (14b), sideward movement of *how*, as shown in (16). The question is why the latter is not licensed.

(15)  
   a.  \( K = [\text{John read } [\text{which paper}]] \)  
       \( L = \text{file} \)  
   b.  \( K = [\text{John read } [\text{which paper}]^i] \)  
       \( M = [\text{file } [\text{which paper}]^i] \)

(16)  
   a.  \( K = [\text{Jane cooked the chicken } \text{how}] \)  
       \( L = [\text{Deborah cook the chicken}] \)  
   b.  \( K = [\text{Jane cooked the chicken } \text{how}^i] \)  
       \( M = [\text{Deborah cook the chicken } \text{how}^i] \)

As discussed by Hornstein and Nunes (2002) and Nunes (2004), sideward movement satisfies Last Resort in (15), but not in (16). More specifically, the copying of *which paper* in (15) is triggered by \( \theta \)-considerations: *file* must assign its \( \theta \)-role and this convergence requirement licenses the copying of *which paper*. By contrast, there is no comparable requirement in \( L \) in (16a) that could trigger the copying of *how*. Although it is quite reasonable to say that *file* in (15a) needs an argument, it makes no sense to say that \( L \) in (16a) needs an adjunct. Once copying of *how* in (16a) is not (locally) licensed, the parasitic gap construction in (14b) is correctly excluded.\(^5\)

In sum, the Copy operation underlying sideward movement is not different from the one underlying upward movement: both must (locally) comply with Last Resort.

### 6.3.2 Derivational timing and the directionality of sideward movement

Let us now consider the contrast in (17) below. (17a) involves the sideward movement depicted in (15). In turn, the derivation of (17b) requires the instance of sideward movement shown in (18). The derivational step in (18) cannot be excluded by Last Resort, for the \( \theta \)-requirements of *file* can license the copying of *which paper*, as we saw in (15). Thus, it must be the case that *which paper* in (18a) is not accessible to Copy at the derivational step where *file* could have its \( \theta \)-requirements satisfied.

(17)  
   a.  \([\text{which paper}]_i \text{ did you file } t_i \text{ after John read } \text{PG}_i] \)  
   b.  \(*[\text{which paper}]_i \text{ did you file } t_i \text{ after John left the room without reading } \text{PG}_i] \)

(18)  
   a.  \( K = [\text{John } [v_P [v_P \text{ left the room } ] [v_P \text{ without reading } [\text{which paper}]])] \)  
       \( L = \text{file} \)  
   b.  \( K = [\text{John } [v_P [v_P \text{ left the room } ] [v_P \text{ without reading } [\text{which paper}]^i])] \)  
       \( M = [\text{file } [\text{which paper}]^i] \)

\(^5\) This indicates that \( \theta \)-relations can trigger both Merge and Copy, whereas modification relations can only trigger Merge. That we have different conditions applying to different operations is not in itself surprising. However, it remains to be explained why modification cannot license copying.
The puzzling contrast in (17) finds a straightforward answer if the computation works in a bottom-up and phase-by-phase fashion, as currently assumed (see e.g. Chomsky 2000, 2001, 2004). Assuming phase-based computations, movement must proceed from more to less embedded domains. This is indeed the case in both (15) and (18). However, there is a crucial difference between these two (see Nunes and Uriagereka 2000, Hornstein 2001, Nunes 2001, 2004, and Hornstein and Nunes 2002). In (18), movement/copying targets an expression that is inside an adjunct. Regardless of how one implements adjunct islands in Minimalist terms, such movement should induce an island effect; hence the unacceptability of (17b). By contrast, in (15) no element containing which paper is an adjunct. Crucially, adjunct is not an absolute, but relational notion: a given expression is an adjunct of another. In (15a) K is just a root syntactic object. The fact that later on K will become an adjunct is irrelevant at the derivational step where movement takes place. In other words, there is no island configuration in (15) that would prevent copying. In fact, the copying seen in (15) is no different from the copying found in licit instances of upward movement (cf. (7)): in both circumstances, copying proceeds from a configuration that is not an island.

One could ask why copying of which paper in the derivation of (17b) cannot take place before the relevant PP becomes an adjunct, as illustrated in (19) below. In (19a) K is not an adjunct and therefore which paper is accessible for copying. If it is indeed copied, it may later merge with file and there would be no island violation, which would incorrectly rule (17b) in.

(19)  

a. K = [reading [which paper]]  
L = [John left the room]  
b. K = [reading [which paper]i]  
L = [John left the room]  
M = [which paper]i  
c. P = [vP [vP John left the room] [PP without reading [which paper]i]]  
M = [which paper]i  
d. P = [vP [vP John left the room] [PP without reading [which paper]i]]  
M = [which paper]i  
Q = file  
e. P = [vP [vP John left the room] [PP without reading [which paper]i]]  
R = [file [which paper]i]  

Assuming that Last Resort must be computed in a local fashion, the derivation depicted in (19) violates Last Resort. Notice that in (19a) there is no motivation for copying which paper; hence, movement in (19b) is ruled out by Last Resort at this step. To put it in more general terms, if the computational system cannot resort to look-ahead (across phases), it cannot idly create a copy and leave it hanging around until it can be used. Triggers for copying must be locally available. Again, this is no different from upward movement: standard adjunct island violations could be
incorrectly circumvented if copies could be freely created and kept in stock for later use.

Another question that arises regarding the derivation of (17b) is why *which paper* cannot move from the object position of *file* to the object position of *reading*, as sketched in (20) below. Notice that in (20a) *which paper* is not within an adjunct and its movement in (20b) would satisfy Last Resort as *reading* would assign its internal \( \theta \)-role.

(20) a. \( K = \{ \text{file [which paper]} \} \)
    \( L = \text{reading} \)

b. \( K = \{ \text{file [which paper]} \} \)
    \( M = \{ \text{reading [which paper]} \} \)

The derivation sketched in (20) is excluded if syntactic computations must proceed from more to less embedded domains, an assumption that is independently made in phase-based approaches to reduce computational complexity. In a well-behaved computation, the system first builds the (embedded) adjunct clause before activating the matrix derivational workspace (cf. (18a)). However, this is not the case in (20), for the matrix domain is activated before the most embedded domain is completed. To put it in general terms, the assumption that derivations unfold from more to less embedded domains has the effect that in the specific case of sideward movement, it must proceed from a “will-be” adjunct to the matrix derivational domain and not vice-versa.6 This assumption not only rules out the derivational step in (20), which would incorrectly allow (17b), but also makes interesting empirical predictions.

---

6 Technically, the directionality from more to less embedded domains can be enforced if we assume with Chomsky (2000) that a numeration is actually composed of subarrays, each of which containing one instance of a (strong) phase head, and that the computational system activates one subarray at a time. Relevant to our current discussion is the following consequence of this phase-by-phase approach: If the maximal projection determined by a subarray must be a phase, then prepositions that select clausal complements must belong to the “subordinating” array in convergent derivations, and not to the array associated with the complement clause—otherwise the CP phase will not be the maximal projection determined by the active subarray, for it will be embedded under a PP (see Nunes and Uriagereka 2000, Nunes 2001, 2004). With this in mind, consider the potential underlying subarrays for the sentence in (i) given in (ii), for instance.

(i) John called Mary after visiting Sue.
(ii) a. \( N_1 = \{ \{ \text{C}, \text{-ed} \}, \{ \text{John}, \text{v}, \text{call}, \text{Mary} \}, \{ \text{C, after}, \text{C}, \text{-ing} \}, \{ \text{v}, \text{visit}, \text{Sue} \} \} \to \)

b. \( N_2 = \{ \{ \text{C}, \text{-ed} \}, \{ \text{John}, \text{v}, \text{call}, \text{Mary} \}, \{ \text{C, after} \}, \{ \text{v}, \text{visit}, \text{Sue} \} \} \to \)

c. \( N_3 = \{ \{ \text{C}, \text{-ed} \}, \{ \text{v}, \text{call}, \text{Mary} \}, \{ \text{C, after} \}, \{ \text{C, -ing} \}, \{ \text{v}, \text{visit}, \text{Sue} \} \} \to \) 

Regardless of which sequence of subarrays is activated, derivations starting with the numeration \( N_1 \) in (iia) are all illegitimate, for the maximal projection determined by subarray C is not a phase. As for \( N_3 \) in (iib), if the computation starts with subarray E, the derivation is doomed because although the complementizer can merge with T, the extension condition prevents T from acquiring a complement later on in the derivation. The same considerations apply to subarrays G and F. In the case of F, the problematic element is the preposition *after*. If the computational system first accesses the subarray F and builds \( \{ [\text{John} [\text{v [call Mary]]}] \} \), the extension condition will later block noncyclic merger between *after* and the clausal
Suppose, for instance, that a given expression becomes inert for purposes of movement in some specific configuration. If derivations proceed from more to less embedded domains, the prediction is that a freezing configuration for an application of sideward movement may be found in the more embedded domain (the launching site), but not in the less embedded domain (the target of movement). Hornstein and Nunes (2002) and Nunes (2004) argue this is what is behind contrasts such as (21), originally noted by Postal (1993).

(21)  
   a. This is the book which, I was [[given t₁ by Ted] [after reading PG₁]]
   
   b. *This is the book which, I [[read t₁] [before being given PG₁ by Ted]]

In order for the parasitic gap constructions in (21) to be derived under a sideward movement analysis, the computational system must copy which from the more to the less embedded domain, as respectively illustrated in (22) and (23) (irrelevant details omitted).

(22)  
   a. K = [reading which]  
      L = given  
   
   b. K = [reading which₁]  
      L = [given which₁]

complement. If the computation starts with subarray H in (iib), building \( [VP \{v \text{ visit Mary}\}] \), no continuation leads to convergence either. If subarray G is activated after subarray H, -ing will successfully merge with the already assembled \( VP \), but the external argument will be prevented from being inserted non-cyclically later on. If subarray F is activated after subarray H, the external argument can be merged to \( [VP \{v \text{ visit Mary}\}] \) in a cyclic manner, but a problem will then arise with after. Assuming that after selects for a clausal complement and not for a \( VP \), it will not have its selectional features satisfied at the phase determined by subarray F and the extension condition prevents noncyclic introduction of the clausal ingredients present in subarray G. By contrast, \( N₃ \) in (iic) can lead to a convergent derivation if the computational system activates the subarrays L, K, J, and I in this order, as sketched in (iii). As the reader can see, if derivations work in a phase-by-phase fashion, and if extension holds, (sideward) movement is bound to proceed from more to less embedded domains.

(iii)  
   a. Computation based on subarray L:  
      \( [VP \{v \text{ visit Sue}\}] \)
   
   b. Subsequent computations based on subarray K:  
      \( [CP C \{TP John \text{-ing } [VP John \{v \text{ visit Sue}\}]\}] \)
   
   c. Subsequent computations based on subarray J:  
      c'. Sideward movement:  
      \( [CP C \{TP John \text{-ing } [VP John \{v \text{ visit Sue}\}]\}] \)
      \( [VP John \{v \text{ call Mary}\}] \)
   
   c''. Merger:  
      \( [VP \{VP John \{v \text{ call Mary}\}\} [PP after [CP C \{TP John \text{-ing } [VP John \{v \text{ visit Sue}\}]\}]]]] \)
   
   d. Subsequent computations based on subarray I:  
      \( [CP C \{TP John \text{-ed } [VP John \{v \text{ call Mary}\}] [PP after [CP C \{TP John \text{-ing } [VP John \{v \text{ visit Sue}\}]\}]]]] \)
(23) a. \[ K = \text{[being given which by Ted]} \]
\[ L = \text{read} \]

b. \[ K = \text{[being given which}^1 \text{by Ted]} \]
\[ M = \text{[read which}^1 \text{]} \]

In (22) and (23) we have sideward movement involving the object positions of *give* and *read*, the only difference being whether they sit in an embedding or embedded domain. When *read* sits in embedded domain, as in (22), sideward movement yields a licit result (cf. (21a)), but when *given* does, as in (23), the result is not well formed (cf. (21b)).

Assuming that the theme of double object constructions receives inherent Case and that inherent Case renders an element inert for purposes of A-movement, Hornstein and Nunes argue that the contrast in (21) independently follows from the directionality of sideward movement. In (23a), *which* is assigned inherent Case by *given* and is therefore frozen for purposes of (sideward) A-movement; hence the unacceptability of (21b). By contrast, in (22) *which* receives inherent Case only after sideward movement takes place. Hence, the movement in (22) may lead to a well-formed result (cf. (21a)). Crucially, if sideward movement could proceed from embedding to embedded domains, (21b) would be incorrectly allowed in a derivation employing the steps in (24) below where the embedding domain is activated before the embedded domain is completed. Thus, the contrast in (21) provides independent support for the assumption that derivations proceed from more to less embedded domains and the corresponding directionality of sideward movement.

(24) a. \[ K = \text{[read which]} \]
\[ L = \text{given} \]

b. \[ K = \text{[read which}^1 \text{]} \]
\[ M = \text{[given which}^1 \text{]} \]

To summarize, if derivations are to proceed in a phase-by-phase fashion and if look-ahead must be minimized, sideward movement (like upward movement) becomes quite constrained and a good number of unwanted derivations involving sideward movement are excluded based on the way derivations unfold (from embedded to embedding domains) and the derivational timing of the Copy operation. An application of Copy is licensed by Last Resort only if its trigger is available to the computation at the derivational step where Copy takes place.

6.3.3 Deletion of copies and linearization of chains

One question that arises in any version of the copy theory of movement is why (in general) it is only one copy that surfaces at PF. Why must the structure in (25), for instance, surface as (26a) and not (26b)?
Nunes (1995, 1999, 2004) argues that linearization considerations prevent (25) from surfacing as (26b). The gist of the proposal is the following. A chain is a discontinuous element, occupying different positions at a time. A PF object, on the other hand, is a linear string. Thus, if the system attempts to realize the whole chain at PF, no linear order will obtain. Consider the linearization of (25) without deletion, for instance. Assuming that linearization is guided by (some version of) Kayne’s (1994) LCA, John must precede was because the upper copy of John asymmetrically c-commands was; by the same token, was must precede John as it asymmetrically c-commands the lower copy of John. However, these two instances of John are nondistinct (they relate to the same material in the initial numeration); thus, the linearization of (25) without an application of deletion yields the contradictory requirement that John must precede and be preceded by was. Likewise, John would be required to precede itself as the higher copy asymmetrically c-commands the lower one. Nunes (1995) proposes that the deletion of chain links, which he refers to as Chain Reduction, allows the computational system to linearize structures containing chains. If the lower copy of John in (25) is deleted, for example, the structure can be trivially linearized as (26a) and no contradiction arises.

Assuming that deletion of copies is performed by Chain Reduction, Nunes (1995, 2001, 2004) further argues that linearization considerations also rule out unwanted instances of sideward movement. Consider, for instance, the well-known contrast in (27), which in GB was taken to show that parasitic gaps must be licensed at S-Structure (see e.g. Chomsky 1982).

(27)  a. [which paper], did you file t, without reading PGi?
     b. *Who filed [which paper], without reading PGi?

Under a sideward movement analysis, the derivation of either construction in (27) involves a licit application of sideward movement from the object position of reading to the object position of filed, as shown in (28).

(28)  a. K = [reading [which paper]]
     L = file/filed
     b. K = [reading [which paper]i]
     M = [file/filed [which paper]i]

However, the derivations underlying the sentences in (27) differ after their final structures are submitted to linearization. In the structure associated with (27a) given in (29a) below, two chains can be formed: CH_1 = (copy^1, copy^3) and CH_2 = (copy^1, copy^3). Crucially, no chain can be formed between copy^2 and copy^3 due to
lack of c-command between them. Applying to CH$_2$, Chain Reduction deletes copy$^3$, yielding (29b). Applying to CH$_1$, Chain Reduction deletes copy$^2$ (cf. (29c)) and the structure surfaces as (27a) after it is linearized.

(29) a. [[which paper]$^1$ [did you [[file [which paper]$^2$] [after reading [which paper]$^3$]]]]

b. [[which paper]$^1$ [did you [[file [which paper]$^2$] [after reading [which paper]$^3$]]]]

c. [[which paper]$^1$ [did you [[file [which paper]$^2$] [after reading [which paper]$^3$]]]]

By contrast, Chain Reduction cannot apply to (30) below (the structure underlying (27b)), because no chain can be formed between the two copies of which paper. Once Chain Reduction is inapplicable, the two nondistinct copies of which paper prevent the structure from being linearized for basically the same reason (25) cannot be linearized: without, for instance, is subject to the contradictory requirement that it must precede and be preceded by which paper (recall that the two copies are nondistinct).

(30) [who [[filed [which paper]$^1$] [without reading [which paper]$^1$]]]

To sum up, sideward movement is drastically constrained by linearization considerations. Its output yields an acceptable result only if further computations allow an additional copy to form an independent chain with each of the copies related by sideward movement. That is the case in (29a) (see also (11g), (13e)), but not in (30).

6.3.4 Summary

In this section we saw how sideward movement can be adequately constrained so that it does not overgenerate. The discussion was not meant to be exhaustive (see Nunes 2001, 2004 for further discussion). Rather, the point was to illustrate that the same conditions that regulate “upward” movement regulate sideward movement.

6.4 Sideward movement and noncanonical phonetic realization of copies

So far we have seen canonical instances of upward and sideward movement, where the highest link of the relevant chain is kept and the lower copies are deleted in the phonological component. However, an increasing body of literature has documented cases where it is the highest chain link that is deleted and even cases where more than one chain link is phonetically realized. Below I show that the general circumstances

---

that allow these two types of exceptions also affect chains that result from sideward movement, as should be expected if the output of both upward and sideward movement is subject to the same conditions on PF realization.

6.4.1 Pronunciation of lower copies

Recall from section 6.3.3 that Chain Reduction deletes lower chain links in order to allow linearization of structures containing chains. The question is why (in the general case) Chain Reduction does not delete the head of the chain and keep a lower copy for phonetic realization. Given the structure in (31), for instance, why must it surface as (32a) and not as (32b)?

(31)  [Johnʰ [was [arrested Johnʰ]]]

(32)  a. John was arrested.

Nunes (1995, 1999, 2004) proposes that the general pattern illustrated in (32) follows from economy considerations. Roughly speaking, as movement allows feature checking/valuation, higher copies have more features checked/valued than lower copies. Thus, all things being equal, the system generally keeps the highest copy as it is the more optimal copy for PF realization for having the greatest number of features checked/valued. However, given that this is a choice based on economy considerations, in case independent convergence requirements are violated if the highest copy is phonetically realized, Chain Reduction deletes the highest copy and keeps the second highest copy.8

A clear case of lower copy pronunciation is presented by the contrasts in (33) and (34) below, as discussed by Bošković (2002b). (33) shows that Romanian is a multiple wh-fronting language. However, the object wh-phrase does not appear to move if it is homophonous with the fronted subject wh-phrase, as shown in (34). Bošković proposes that Romanian has a low-level PF constraint against adjacent homophonous wh-phrases, which rules out (34b). As for the exceptional pattern in (34a), Bošković argues that it also involves multiple wh-fronting in the syntactic component, but in order to comply with the PF constraint on adjacent homophonous elements, the higher copy of the object wh-phrase is deleted and the lower one is pronounced instead, as sketched in (35) (irrelevant details omitted).

---

8 For different technical implementations, see Nunes 1995, 1999, 2004 for an approach in terms of feature elimination in the phonological component and Nunes 2011 for an approach in terms of chain internal probing.
(33) Romanian (Bošković 2002b):
   a. Cine ce precede?
      who what precedes
   b. *Cine precede ce?
      who precedes what
      ‘Who precedes what?’

(34) Romanian (Bošković 2002b):
   a. Ce precede ce?
      what precedes what
   b. *Ce ce precede?
      what what precedes
      ‘What precedes what?’

(35) \[ce_{SUBJ} ce_{OBJ} \quad [ce_{SUBJ} precede ce_{OBJ}]\]

Let us now consider instances of lower copy pronunciation in constructions involving sideward movement. We have seen in section 6.3.3 that Chain Reduction operates with chains and this is what accounts for why a parasitic gap construction such as (36a) below cannot be associated with the structure in (36b), which is derived by sideward movement of which paper. Given that the two copies in (36b) do not form a chain, Chain Reduction is inapplicable and the structure cannot be linearized. Combined with the approach on lower copy pronunciation presented above, this minimalist reanalysis of the S-Structure condition on parasitic gap licensing makes the prediction that a construction superficially similar to (36a) should be well formed if the object wh-phrase cannot surface in the fronted position. Bošković (2002b) shows that this prediction is indeed borne out, as illustrated by (37).

(36) a. *Who filed [which paper], without reading PGi?

   b. [who [[filed [which paper]\textsuperscript{1}] [without reading [which paper]\textsuperscript{1}]]]

(37) Romanian (Bošković 2002b):
    
    Ce precede ce\textsubscript{i} fara sa influenteze PG\textsubscript{i}?
    what precedes what \textsuperscript{SUBF,PRT} influence,3,SG
    ‘What precedes what, without influencing it?’

Following Bošković, we assume that the object wh-phrase undergoes wh-fronting in the overt component, as is the standard case in Romanian, yielding the simplified structure in (38), where an instance of ce moves from the object position of influence to the object position of precede before undergoing wh-fronting.

(38) \[ce_{SUBJ} ce^{3} \quad [[precede ce^{2}] [fara sa influenteze ce^{1}]]\]


In (38) the fronted wh-object forms the chain $CH_1 = (\text{copy}_3, \text{copy}_1)$ and $CH_2 = (\text{copy}_3, \text{copy}_1)$. Applying to $CH_1$, Chain Reduction deletes the lower copy of ce, yielding (39a) below. By contrast, if Chain Reduction deletes the lower copy of ce when applying to $CH_2$, the derivation will not converge due to adjacency between the homophonous subject and object wh-phrases. In order to circumvent this problem, Chain Reduction deletes the higher copy, as shown in (39b), and the structure surfaces as (37), which superficially seems to involve a parasitic gap licensed by a wh-phrase in situ.

\[
\begin{align*}
39 \quad & \text{a. } [ce_{SU} ce^3 [[\text{precede } ce^2] [\text{fara sa influenceze } ce^4]]] \\
39 \quad & \text{b. } [ce_{SU} ce^2 [[\text{precede } ce^2] [\text{fara sa influenceze } ce^1]]]
\end{align*}
\]

Another instance of lower copy pronunciation in constructions involving sideward movement is discussed by Dotlačil (2008) with respect to ATB movement in (non-colloquial) Czech. Nunes (1995, 2001, 2004) and Hornstein and Nunes (2002) argue that ATB extraction also involves sideward movement. Assuming this to be the case, Dotlačil discusses a curious case of ATB extraction of clitics in Czech. First, consider the contrast in (40).

\[
\begin{align*}
40 \quad & \text{Czech (Dotlačil 2008):} \\
40 \quad & \text{a. *Zavolal jsem } Petra \text{ a představil známým.} \\
40 \quad & \hspace{1cm} \text{called aux}_{\text{SG}} Petr_{\text{ACC}} \text{ and introduced friends} \\
40 \quad & \text{b. Petra jsem zavolal a představil známým.} \\
40 \quad & \hspace{1cm} \text{Petr}_{\text{ACC}} \text{ aux}_{\text{SG}} \text{ called and introduced friends} \\
40 \quad & \hspace{1cm} \text{‘I called Petr and introduced him to friends.’}
\end{align*}
\]

(40a) shows that Czech does not allow auxiliary gapping and object drop under coordination. The acceptability of (40b) in turn shows that missing auxiliaries and objects may be licensed if they can be analyzed in terms of ATB extraction; that is, in (40b) the object Petra and the auxiliary have undergone ATB extraction (via sideward movement) and bind a trace/copy in each of the conjuncts. That being so, let us now examine (41) below, which differs minimally from (40a) in that we have a pronominal clitic in place of Petra. Clearly, (41) cannot involve auxiliary gapping and object omission in the second conjunct; otherwise, (40a) should also be acceptable. At first sight, (41) could be derived by ATB extraction of the auxiliary and the pronominal clitics (like (40b)), followed by movement of zavolal to a higher position. However, the latter movement would violate the coordination structure constraint, as zavolal would be moving from just one conjunct. So, how can (41) be derived?

---

9 In consonance with the overall bottom-up nature of the computation (see section 6.3.2), I assume that Chain Formation and Chain Reduction also proceed from more to less embedded domains.

10 For further discussion on how lower copy pronunciation works in these cases, see Niinuma 2010.
Dotlačil (2008) argues that (41) behaves exactly like (40b) in the syntactic component, that is, the auxiliary and the pronominal clitics undergo ATB extraction (via sideward movement), yielding the simplified structure in (42a) below (with English words for presentational purposes), where the highest copy of the clitic cluster forms a different chain with each of the lower copies. Reduction of the chain involving the second conjunct deletes the lower copy, as shown in (42b) (see footnote 9). Similar reduction of the chain involving the first conjunct would leave the clitic cluster in clause initial position, which is not allowed in (noncolloquial) Czech. Chain Reduction then deletes the higher copy, as shown in (42c). As Dotlačil shows, rather than being an apparent exceptional pattern of ATB extraction, (41) is another case of pronunciation of lower copies.

To summarize, the above discussion shows that mismatches between syntactic structures and PF realization that can be solved by appealing to phonetic realization of lower copies do not distinguish upward from sideward movement. Again, this is good news for an approach that treats them in a uniform manner as different instantiations of Copy and Merge.

### 6.4.2 Pronunciation of multiple copies

We have seen in section 6.3.3 that linearization considerations in general prevent a chain from surfacing with all of its links phonetically realized. However, this account predicts that if for some reason a given link becomes invisible to the relevant linearization procedure, a chain may surface with more than one link realized at PF. Nunes (1999, 2004) argues that this may happen if a given copy is morphologically fused (in the sense of Halle and Marantz 1993) with another element. The reasoning goes as follows: if a given copy C fuses with a given element E in the morphological component, the blended result #C-E# (or #E-C#) behaves like an atomic vocabulary item with no internal structure accessible to further morphological or syntactic computations. In particular, the fused copy is no longer visible to linearization (the resulting atomic element is) and no contradictory requirement with respect to other copies will arise. Verb clefting constructions in Vata, as illustrated in (43), can exemplify this process.
Koopman (1984) shows that the two verbal occurrences in (43) cannot be separated by islands, which indicates that they should be related by movement. In terms of the copy theory, the verbal instances seen in (43) can then be analyzed as copies produced by the movement operation. The question that arises is why the presence of more than one copy does not create contradictory requirements for linearization. Nunes (2004) proposes that the highest copy of the clefted verb gets morphologically fused, thereby evading the purview of the LCA. More precisely, he analyzes verb clefting in Vata as involving verb movement to a focus head, followed by fusion in the morphological component between the moved verb and the focus head, as represented by the shaded material in (44a) below. Of the three verbal copies in (44a), the LCA only sees the lower two after the highest copy gets fused with Foc⁰. The lowest copy is then deleted (cf. (44b)) and the structure is linearized as in (43), with two copies of the verb phonetically realized. Evidence for morphological fusion in Vata verbal clefting is provided by the fact that the fronted verb cannot occur with tense or negative particles (see Koopman 1984), which makes sense if these particles render the verb morphologically too complex, thereby preventing the verb from undergoing fusion with the focus head.

\[
\text{(44)} \quad \text{a. Fusion: } \\
\left[\text{FocP} \# \text{[Foc⁰ V [Foc⁰ Foc⁰]]# [TP \ldots [T⁰ V [T⁰ T⁰]] [VP \ldots V \ldots]]}\right]
\]

\[
\text{b. Deletion of copies: } \\
\left[\text{FocP} \# \text{[Foc⁰ V [Foc⁰ Foc⁰]]# [TP \ldots [T⁰ V [T⁰ T⁰]] [VP \ldots V \ldots]]}\right]
\]

With this overall picture in mind, one wonders if sideward movement can also yield outputs with more than one copy phonetically realized. In his detailed analysis of control in Telugu and Assamese, Haddad (2007, 2009) shows that constructions such as (45) and (46) below (CNP stands for conjunctive participle particle) display all the traditional diagnostics of obligatory control and argues that they should also be analyzed in terms of sideward movement and phonetic realization of multiple copies.\(^{11}\)

\[
\text{(45) Telugu (Haddad 2007): } \\
\left[\text{[Kumar sinima cuus-tuu] [Kumar popkorn tinnaa-Du]}\right] \\
\text{Kumar.NOM movie watch-CNP Kumar.NOM popcorn ate-3-M.S} \\
\text{‘While watching a movie, Kumar ate popcorn.’}
\]

\(^{11}\) For a discussion of other instances of adjunct control with more than one copy phonetically realized, see Boeckx, Hornstein, and Nunes 2008.
Assamese (Haddad 2007):

\[
\begin{align*}
[\text{Ram-Or } & \text{khong uth-i}] & [\text{Ram-e } & \text{mor ghorto bhangil-e}] \\
\text{Ram-GEN} & \text{anger raise-CNP} & \text{Ram-NOM} & \text{my house destroyed-3} \\
\end{align*}
\]

‘Having got angry, Ram destroyed my house.’

Given the role of morphological fusion in making the phonetic realization of multiple copies possible, it comes as no surprise that multiple copies are only possible if, in Haddad’s (2007: 87) words, the subject “does not exceed one or two words,” as illustrated by the ungrammaticality of (47) below.

Telugu (Haddad 2007):

\[
\begin{align*}
*[\text{Kumar maryu Sarita sinim cuu-tuu}] \\
\text{Kumar.NOM} & \text{and} & \text{Sarita.NOM} & \text{movie watch-CNP} \\
[\text{Kumar maryu Sarita popkorn tinna-ru}] \\
\text{Kumar.NOM} & \text{and} & \text{Sarita.NOM} & \text{popcorn ate} \\
\end{align*}
\]

‘While Kumar and Sarita were watching a movie, they ate popcorn.’

These restrictions can be interpreted as showing that if the realization of multiple copies is licensed via morphological fusion, it should naturally be very sensitive to morphological information. The first kind of relevant information regards the feature composition of the elements that are to be fused. After all, not any two elements can get fused, but only the ones that satisfy the morphological requirements of one another. The second kind of information concerns morphological complexity. As a rule, the more morphologically complex a given element is, the less likely it is for it to undergo fusion and become part of a terminal. Thus, the addition of specific morphemes (which may vary from language to language) may make the resulting element morphologically “too heavy” to become reanalyzed as part of a word. Of course, if a given copy is syntactically complex, i.e. it is phrasal, as in (47), it is also morphologically complex and not a good candidate to undergo morphological fusion. In turn, once fusion is prevented from applying, the presence of more than one copy induces a violation of the LCA.

To sum up, we again see that the exceptional circumstances and conditions that regulate acceptable instances of pronunciation of more than one copy apply to upward and sideward movement indistinctively.

6.5 Wh-movement and adjunct control

A general property of adjunct control constructions like (48) below, for instance, is that PRO must be controlled by the subject and not the object of the next higher clause. Hornstein (2001) proposes that this subject-object asymmetry results from an economy choice at the derivational step sketched in (49).
(48) John saw Mary after PRO\(_{PRO}\) eating lunch

(49) \(N = \{John, saw, Mary, after, eating, lunch\}\)
\(K = [John]\)
\(L = saw\)

In (49), saw must assign its internal \(\theta\)-role and there are two potential candidates to receive it: Mary, which is still in the numeration, and John in the subject position of the gerundive clause. If Mary is selected and merged with saw, as seen in (50a) below, the derivation converges as a subject control structure, after John undergoes sideward movement to [Spec,VP] (cf. (50b)).

\[
(50) \quad a. \quad N' = \{John, saw, Mary, after, eating, lunch\}
\quad K = [John]\
\quad M = [saw Mary]
\]

\[
(50) \quad b. \quad N' = \{John, saw, Mary, after, eating, lunch\}
\quad K = [John^1 eating lunch]
\quad P = [John^1 saw Mary]
\]

\[
(50) \quad c. \quad [TP John^1 [vP [vP John^1 saw Mary] [PP after John^1 eating lunch]]]
\]

On the other hand, if John is copied and merged with saw, as shown in (51a) below, the derivation should in principle converge as well, this time yielding an object control structure after Mary is plugged in as the external argument, as shown in (51b). Under the assumption that Merge is more economical than Move (see Chomsky 1995b), Hornstein observes that the derivation in (50) is more economical than (51); hence, the subject-object asymmetry in (48).\(^{12}\)

\[
(51) \quad a. \quad N = \{John, saw, Mary, after, eating, lunch\}
\quad K = [after John^1 eating lunch]
\quad M = [saw John^1]
\]

\[
(51) \quad b. \quad N' = \{John, saw, Mary, after, eating, lunch\}
\quad K = [after John^1 eating lunch]
\quad P = [Mary saw John^1]
\]

\[
(51) \quad c. \quad [TP Mary^k [vP [vP Mary^k saw John^1] [PP after John^1 eating lunch]]]
\]

\(^{12}\) Under Nunes’s (1995, 2001, 2004) system, a structure such as (51c) is independently excluded because it cannot be linearized, as the two copies do not form a chain and, accordingly, are not subject to deletion under Chain Reduction (cf. (30)). For the purposes of presentation, I will however put this possibility aside and frame the following discussion in terms of Hornstein’s (2001) original Merge-over-Move approach. The proposal to be suggested below is compatible with either analysis.
Although this preference for subject over object control may be the general case, convergence requirements may lead to the opposite situation. This can be seen in the interaction between wh-movement and adjunct control in (Brazilian and European) Portuguese. In these languages, the subject of infinitival adjunct clauses may be controlled by the matrix subject or the matrix object, depending on whether or not the matrix object undergoes wh-movement, as illustrated in (52) below.\(^{13}\) (52b) has a wh-in situ in the matrix clause and the result is subject control, as in (52a), with no wh-element involved. By contrast, (52c) has wh-movement and now both subject and object control are possible.

\[(52) \text{Portuguese:}\]

\begin{enumerate}
  \item \text{a. [Os alunos]}\(_i\) entrevistaram \text{[os professores]}\(_k\) antes de PRO\(_{i/*k}\)
  \text{the students interviewed the professors before of}
  \text{sair de férias.}
  \text{leave of vacation}
  \text{‘The students interviewed the professors before leaving on vacation.’} \\
  \item \text{b. [Os alunos]}\(_i\) entrevistaram \text{[que professores]}\(_k\) antes de PRO\(_{i/*k}\)
  \text{the students interviewed which professors before of}
  \text{sair de férias?}
  \text{leave of vacation}
  \text{‘Which professors did the students interview before leaving on vacation?’} \\
  \item \text{c. [Que professores]}\(_k\) é que \text{[os alunos]}\(_i\) entrevistaram \text{tk}
  \text{which professors is that the students interviewed}
  \text{antes de PRO\(_{i/k}\) sair de férias?}
  \text{before of}
  \text{leave of vacation}
  \text{‘[Which professors]\(_k\) did [the students]\(_i\) interview before they\(_{i/k}\) left on vacation?’} \\
\end{enumerate}

Assuming with Bošković (2007b) that the strong feature that triggers successive cyclic movement (\(uF\)) is hosted by the moving element, I have proposed (see Nunes 2010) that in languages like Brazilian and European Portuguese, with optional wh-movement, this feature is lexically optional on wh-elements.\(^{14}\) Moreover, the presence of

\[^{13}\text{The matrix clause of (52) involves a plural subject and a plural object, whereas the infinitival clause is uninflected; hence, we are dealing here with regular adjunct control rather than a pro licensed by an inflected infinitive. For original discussion of the finite counterparts of (52) in Brazilian Portuguese, see Modesto 2000 and Rodrigues 2004.}\]

\[^{14}\text{In fact, Bošković (2007b) proposes that wh-elements in English are lexically specified as optionally having a strong feature} uF} \text{and that its interrogative complementizer can only be checked by a wh-element marked with} uF; \text{hence, an interrogative complementizer cannot be checked by a wh-phrase in situ. However, adjunct control in English differs from what we find in Portuguese in that it always involves subject control, regardless of whether or not the object undergoes wh-movement, as illustrated in (i) below.}\]
this feature in the derivation has consequences for economy computations regarding Merge-over-Move. Recall that subject over object control is enforced in adjunct control constructions due to Merge being more economical than Move (see footnote 12). In the case of (52a), for instance, if os alunos “the students” is in the subject position of the adjunct clause, it cannot undergo sideward movement to the complement of the matrix verb, for merger of os professores “the professors” in this position is more economical. So, after os professores is merged, os alunos can only move to the matrix [Spec,vP], yielding subject control (cf. (49)–(50)).

Bearing this in mind, the object control reading of (52b) and (52c) should involve the derivations sketched in (53a) and (53b), respectively.

(53) Portuguese:

a. *Os alunos [[entrevistaram [queaf professores]]] [antes de ti sair de férias]? 
Which professors, did the students interview before they left on vacation?

b. [queaf professores] é que os alunos [[entrevistaram ti] antes de ti sair de férias]? 
Which professors is that the students interviewed before they left on vacation?

English also differs from Portuguese in that the latter allows wh-in situ (in embedded clauses) even when there is a single wh-element, as shown in (ii).

(i) a. Who, greeted who, after PROi/*k entering the room?
   b. Who, did John, greet tk after PROi/*k entering the room?

(ii) a. *John said that Mary is going to travel when?
   b. Portuguese:

   O João disse que a Maria viaja quando?
   When did João say that Maria is going to travel?

In Nunes (2010) I took the general availability of optional wh-movement in Portuguese to indicate that its optional specification for a strong feature is truly lexical. As for English, my conjecture is that uF is optionally assigned not in the lexicon, but in the course of the derivation, when phases are completed (as in Chomsky 2001). Assignment of such feature is however subject to Last Resort: only when the wh-element is not accessible to the computational system (typically, when it is not in a phase edge) can it be assigned a strong feature to undergo successive cyclic movement. Thus, a wh-phrase in the subject position of an adjunct clause will not be assigned uF and Merge-over-Move will be enforced, always yielding subject control (cf. (ib)). Notice that this suggestion is still compatible with Bošković’s (2007b) main proposal that edge features are borne by the relevant moving elements and not by the heads of phases. I leave further development of this suggestion for another occasion.
The wh-element of both derivations in (53) enters the numeration specified with a strong feature $uF$, which in turn requires that the wh-phrase must move if possible. This requirement of the strong feature now overrules Merge-over-Move, for things are not equal anymore. If the wh-element sits in the subject of the adjunct clause and sideward movement to the matrix object position is possible, such movement must take place. Now, if Merge-over-Move is circumvented in the presence of a strong feature $uF$, this strong feature must be checked. Hence, (53a) is unacceptable not because movement of the wh-element from the adjunct clause to the matrix object position violates Merge-over-Move, but because the strong feature of the wh-phrase remained unchecked. When it is checked by moving to [Spec,CP], as in (53b), the derivation converges, yielding an object control reading.15

To sum up. Focusing on successively cyclic (upward) movement, Bošković (2007b) has argued that the computational complexity of syntactic derivations gets substantially minimized if the feature that drives movement is borne by the moving element rather than the probe. I have shown above that this general proposal not only extends to sideward movement, but also makes the correct empirical cut regarding the interaction between wh-movement and adjunct control in languages with generalized optional wh-movement such as (Brazilian and European) Portuguese.

6.6 Morphological sideward movement

Let us finally consider syntax-phonology mismatches involving preposition duplication in colloquial Brazilian Portuguese (henceforth BP). Take the data in (54) and (55), for instance.

15 For the sake of completeness, it remains to show how the subject control reading of both (52b) and (52c) can be obtained. In both cases, the subject control reading results from a derivation in which os alunos “the students” is generated in the adjunct clause and undergoes sideward movement after the wh-phrase is merged in the matrix object position (in compliance with Merge-over-Move). As respectively shown in (i) and (ii) below, the difference between the two derivations involves the lexical specification of the wh-element. In (52b), it is not associated with an $uF$ feature; hence we have wh-in situ (cf. (i)). By contrast, in (52c) the wh-element has an $uF$ feature and therefore must undergo wh-movement to check it (cf. (ii)).

(i) [Os alunos, t]i entrevistaram [que professores]k antes de sair de férias]
   ‘Which professors did the students interview before leaving on vacation?’

(ii) [Que professores]k é que [os alunos, t]i entrevistaram t antes de sair de férias]
    ‘Which professors is that the students interviewed before leaving on vacation?’
6.6 Morphological sideward movement

(54) a. *Eu pensei em o João. [formal/colloquial BP]
    I thought in the João
    'I thought about João.'

    b. Eu pensei no João. [formal/colloquial BP]
    I thought in-the João
    'I thought about João.'

(55) a. Eu pensei em o João fazer esse trabalho. [formal BP]
    I thought in the João do-INF this job
    'I thought that João should do this job.'

    b. Eu pensei no João fazer esse trabalho. [colloquial BP]
    I thought in-the João do-INF this job
    'I thought that João should do this job.'

(54) shows that in BP the preposition em “in” and the definite article o “the” must contract when they are adjacent. In turn, (55) shows that if the definite article belongs to the embedded subject, we have contraction in colloquial BP, but not in its formal registers. Nunes and Ximenes (2009) analyze the difference between (55a) and (55b) as arising from two different structures. In formal registers of BP, the case-marking preposition em precedes the whole infinitival CP, as shown in (56) below, and in this circumstance it is not adjacent to the determiner due to the intervention of C; lack of adjacency then yields lack of contraction (cf. (55a)). As for colloquial BP, Nunes and Ximenes argue that the preposition is realized as C, which renders it adjacent to the determiner, as sketched in (57), and contraction is obligatory (cf. (55b)).

(56) **Formal BP:**

[...X [P [infinitival-CP C [TP [DP D ...]]]]]

(57) **Colloquial BP:**

[...X [infinitival-CP P/C [TP [DP D ...]]]]

A very puzzling paradigm arises in colloquial BP when the contraction patterns depicted in (54)–(55) are combined with coordination, as illustrated in (58) and (59) (see Ximenes (2002, 2004), Ximenes and Nunes (2004), and Nunes and Ximenes (2009)).

(58) a. *Eu pensei no João e a Maria. [formal/colloquial BP]
    I thought in-the João and the Maria
    'I thought about João and Maria.'

    b. Eu pensei no João e na Maria. [formal/colloquial BP]
    I thought in-the João and in-the Maria
    'I thought about João and Maria.'
a. Eu pensei em o João e a Maria
I thought in the João and the Maria
fazerem esse trabalho.
[formal BP]
do-INF.3PL this job
‘I thought that João and Maria should do this job.’

b. "Eu pensei em o João e em a Maria
I thought in the João and in the Maria
fazerem esse trabalho.
[formal/colloquial BP]
do-INF.3PL this job
‘I thought that João and Maria should do this job.’

c. Eu pensei no João e na Maria [colloquial BP]
I thought in-the João and in-the Maria
fazerem esse trabalho.
do-INF.3PL this job
‘I thought that João and Maria should do this job.’

(58) shows that contracting prepositions must be repeated if one of the conjuncts has a determiner that triggers contraction. This suggests that the parallelism requirement on coordinated structures (see e.g. Chomsky 1995b, Fox 2000, and Hornstein and Nunes 2002) also applies to the morphological component. That is, once contraction appears in one conjunct, it must appear in every conjunct. Thus, at first sight, (58) can converge only if there are two prepositions in the underlying numeration and the PPs headed by these prepositions are accordingly coordinated, as sketched in (60).

(60) [Eu pensei [[PP no João] e [PP na Maria]]]
I thought in-the João and in-the Maria

However, this account cannot be extended to the full paradigm of (59). The lack of contraction between the subcategorizing preposition and the determiner of the subject in (59a) is not surprising. It just replicates the pattern seen in (55a), which in formal BP is associated with the structure in (56), where the null complementizer breaks the adjacency between the preposition and the determiner. The ungrammaticality of (59b) in either register is not mysterious either. Given that the coordinated subject must involve DP coordination as it is interpreted as the agent of the embedded verb and triggers plural agreement on the inflected infinitival, (59b) is out due to the presence of a spurious preposition in the second conjunct. If a coordination involving PP is not an option for (59b), the question then is why a sentence analogous to (59b) becomes acceptable in colloquial BP if the prepositions get contracted with the relevant determiners (cf. (59c)).
Nunes and Ximenes (2009) (see also Ximenes 2002, 2004 and Ximenes and Nunes 2004 for discussion) argue that (59c) results from the application of sideward movement in the morphological component, that is, an application of copying\footnote{Such copying can be seen as a subtype of the standard operation involved in morphological reduplication.} and morphological merger. They propose that (59c) indeed involves coordination of DPs, as expected, and that the second preposition is inserted in the morphological component. More specifically, if we have morphological merger (see Halle and Marantz 1993) in the boundary of one conjunct, the parallelism requirement triggers morphological merger in all conjuncts. The derivation of (59c), for instance, proceeds along the lines of (61).

\begin{enumerate}
\item \textit{Spell-out:}
\begin{verbatim}
[ ... pensei [CP \textbf{em} [TP [andP [DP o João] [and e [DP a Maria]]] fazerem ...]]]
\end{verbatim}
\item \textit{Morphological merger:}
\begin{verbatim}
[ ... pensei [CP [TP [andP [DP \textbf{em+o} João] [and e [DP a Maria]]] fazerem ...]]]
\end{verbatim}
\item \textit{Copy and morphological merger:}
\begin{verbatim}
[ ... pensei [CP [TP [andP [DP \textbf{em}+o João] [and e [DP \textbf{em}+a Maria]]] fazerem ...]]]
\end{verbatim}
\item \textit{Fusion:}
\begin{verbatim}
[ ... pensei [CP [TP [andP [DP no João] [and e [DP na Maria]]] fazerem ...]]]
\end{verbatim}
\end{enumerate}

Given that in colloquial BP, case-marking prepositions are realized in C when they take infinitival complements (cf. (57)), the preposition \textit{em} in (61a) is adjacent to the first determiner of the coordinated embedded subject in the spelled-out structure and morphological merger is obligatory in these circumstances, as seen in (61b). Once morphological merger affects the boundary of the coordinated subject, the parallelism requirement on coordinated structures kicks in and demands that the second conjunct also undergo morphological merger. Given that there is no preposition adjacent to the determiner of the second conjunct (recall that the embedded subject involves DP- and not PP-coordination), the preposition morphologically merged with the first conjunct is then copied and the resulting copy merges with the determiner of the second conjunct, as shown in (61c). Finally, the prepositions and the determiners fuse, as shown in (61d), yielding the PF output in (59c), which at first glance appears to involve a quite exotic case of PP-coordination in a canonical subject position. Notice furthermore that the two copies of the preposition do not induce linearization problems as they become invisible for purposes of linearization after they fuse with the relevant determiners (see section 6.4.2).

To conclude, if the morphological component independently has the operations of copying and merger in its inventory, one should not be surprised if these operations...
interact and yield what may be viewed as sideward movement in the morphological component. After all, sideward movement is just a description of the interaction of these basic operations.\footnote{17}

### 6.7 Concluding remarks

In this chapter I have argued that sideward movement is not a novel operation that makes the grammar’s theoretical apparatus heavier. Sideward movement in fact arises as a natural consequence of the interaction among solid pillars of the Minimalist enterprise: the attempt to eliminate non-interface levels, the postulation of syntactic entities in terms of lexical items, and the attempt to minimize derivational complexity, to name a few. In particular, once Move is interpreted within Minimalism as the output of the interaction between the more basic operations of Copy and Merge, sideward movement is not a theoretical primitive, but is simply a mnemonic label for one particular instantiation of applications of Copy and Merge.

We have also seen that the same conditions that restrict upward movement also constrain sideward movement, again stressing the fact that under the copy theory, it does not make sense to attempt to distinguish upward from sideward movement as operations, for Move is not understood as a primitive operation in the system. In other words, it is actually excluding sideward movement that requires complications in the theoretical apparatus and one has to ponder what is gained and what is lost with the required amendments.

Excluding sideward movement by brute force does not seem to lead to profitable gain from a Minimalist perspective. One would be introducing redundancies in the system, as sideward movement is already quite constrained by the conditions that exclude unwanted instances of “upward” movement. Such exclusion would in fact face a considerable loss in empirical coverage. As discussed in the previous sections, sideward movement may not only be seen as cost free within a core set of Minimalist assumptions, but also paves the way for sound analyses of intricate empirical phenomena. The empirical bar, therefore, requires that the exclusion of sideward movement should be accompanied by a uniform account of empirical phenomena such as the ones discussed here.

\footnote{17 However, as a reviewer observes, sideward movement in the morphological component is different from sideward movement in the syntactic component as the former does not involve two independent objects and does not seem to display directionality effects. This distinct behavior may perhaps follow if syntactic merger is subject to the extension condition, but morphological merger is not. This difference in turn may follow if syntactic merger is a more complex operation involving concatenate (which is akin to morphological merger) and label, as proposed by Hornstein (2009). If so, label is the operation that must apply in consonance with the extension condition and the difference between sideward movement in the syntactic and morphological components would result from whether merger is associated with labeling or not. I will leave the discussion of the consequences of this speculation for another occasion.}
Unconventional Mergers

MARK DE VRIES

7.1 Introduction

An essential aspect of a generative grammar is its structure building capacity.* In the Minimalist Program, rewriting rules have been replaced with the simpler and more general operation called Merge. Merge combines syntactic objects into larger syntactic objects, and it can do so recursively, meaning that the output of Merge can be used as input again. This leads to a hierarchical structure, as we will discuss in detail.

Almost from the beginning, the exact definition and workings of Merge have been disputed, which is remarkable at first sight, since the operation was presented in Chomsky (1995b) as a conceptual necessity, and one would perhaps not expect much discussion on such basic issues. Nevertheless, the generative core of grammar has been an important focus of attention since Syntactic Structures (Chomsky 1957), as it determines which kind of representations can be derived and which cannot. And this puts potential analyses of more concrete constructions and grammatical phenomena within certain boundaries. With respect to Merge, we can distinguish at least five issues of discussion:

(i) Labels. To which extent are the insights of X-bar theory to be incorporated in Merge? How “bare” is phrase structure, and do we need labels to begin with? Are the categorial status and the projection status of phrases predictable? For relevant discussion, see Epstein, Thrainsson, and Zwart (1996), Collins (2002), Di Sciullo and Isac (2008), and Boeckx (2008), among others.

* I would like to thank the anonymous reviewers for useful comments and questions. I am also grateful to the editors for kindly allowing me to submit a chapter to this specialized volume. This research was financially supported by the Netherlands Organisation for Scientific Research (NWO).
(ii) The environment. Which boundary conditions play a role for Merge? How is the syntactic workspace organized? Is there a so-called numeration? Which objects are accessible as input for Merge, and where are the results stored? How many objects can be merged at the same time? Is Merge invoked in cycles/phases? See also Bobaljik (1995a), de Vries (2005a), and Chomsky (2008), for instance.

(iii) The nature of the objects handled and created by Merge. According to Chomsky, a phrase is a simple set (or superset if it contains other phrases). This leads to syntactic representations that involve hierarchy but not order. However, it has also been argued that Merge creates ordered pairs instead. This would mean that the relationship between sisters is basic as well, which in turn affects the discussion on linearization. See, e.g., Langendoen (2003), Zwart (2006), and Fortuny (2008).

(iv) Movement. Is movement simply Merge applied to the same element again (remerge)? Is it possible to move an item from one structure to another? What does that mean for the representation of syntactic structure? Do we need copies, chains, multidominance, or all of them? See Starke (2001), Gärtner (2002), Nunes (2004), de Vries (2009b), and the references therein.

(v) Uniformity. Is there just one kind of Merge, or could there be more? Chomsky himself introduced both set-Merge and pair-Merge (for adjuncts); see Chomsky (2004) for some discussion. Additionally, special devices may be needed for parentheses and the like; see also Ackema and Neeleman (2004) and de Vries (2007b).

Needless to say, these issues are interrelated in several ways.

Merge, at any rate, is a tool fit for creating standard syntactic hierarchies. In this chapter, we will discuss the possibility of certain non-canonical mergers. Section 7.2 is about remerge. It is argued that remerge follows from standard boundary conditions (or rather the lack thereof) on the selection of input elements for Merge. Depending on the exact configuration, there are two options: internal remerge, which corresponds to regular movement, and external remerge, which corresponds to the idea of “sideward movement” or “sharing.” Section 7.3 is about Parenthetical Merge. This is a specialized operation necessary for the connection of parenthetical material with the matrix. Section 7.4 tentatively explores the consequences of the application of both external remerge and Parenthetical Merge within one sentence. It will turn out that this correctly predicts a number of remarkable properties of so-called amalgams: partial “invisibility” and apparent non-locality.
7.2 Internal and external remerge

The operation Merge can be viewed as a (mathematical) function with input variables and one output, applied to a linguistic domain: Merge \((a, b) \rightarrow c\), such that \(a\) and \(b\) are syntactic objects, and \(c\) equals the combination of \(a\) and \(b\), in familiar bracket notation \([c\ a\ b]\), which, by definition, is also a syntactic object. By definition, since it is not a logical necessity generally, properties of a whole can be emergent, and not every property of the parts is necessarily preserved in the whole. For instance, the combination of two singulars leads to a plural, like connected atoms form a molecule. Merge, however, preserves syntactic objecthood. Of course, this is the default assumption, and it is intuitively plausible. Crucially, Merge does not tamper with the input. Thus, Merge \(([c\ a\ b], d)\) results in \([e\ [c\ a\ b]\ d]\), but not in \([a\ b\ d]\), which would in fact imply that the group \(c\) is changed or destroyed. Therefore, the iterative application of Merge automatically yields a hierarchy (1b) and not a flat concatenation (1a): a syntactic structure is not a word string.

\[
\begin{align*}
(1)\ a &-b \quad =^+d \Rightarrow \quad a- b-d \\
\text{b.} \quad \overset{=^+d}{\smiley{a\ b}} &\Rightarrow \overset{}{\smiley{a\ b\ d}}
\end{align*}
\]

In other words, if a group is merged with another syntactic object, the participants of the group do not play a role in the outcome: the original group stays intact. If only for practical purposes, groups deserve a label. In the above example, the inner and outer circle can be called \(c\) and \(e\), respectively. An immediate question that arises is the following: are the elements \(a\) and \(b\) themselves still accessible for Merge once they have been merged together? The usual answer is positive, as we will discuss below. But notice right away that this leads to representational complications.

First, we need to separate the core properties of Merge from the boundary conditions. Essentially, Merge relates syntactic objects to each other, and therefore it is structure building: it combines the input objects into a group. As an immediate consequence, it forms a hierarchy: the original input objects are directly included in the output object. It is usually assumed that the number of input objects must be two, which leads to binary branching. Furthermore, both the input and output may be subject to certain constraints. This in particular is where theories converge. Consider the simple model in (2):

\[\text{(2) In handbooks, these are sometimes treated on a par, and even combined into one complex definition. I think this is unfortunate, also from a didactical perspective.}\]
I will assume that material from the lexicon can be transferred to the syntactic work space, which may then contain several independent syntactic objects. Merge operates on objects within this work space. If syntax is done, the resulting phrase structure is passed on to the interfaces. Possibly, syntax works in cycles/phases; this could mean that periodically, embedded material is pushed out of the syntactic work space and will be kept on stack. Much more can be said about all of this, but let us concentrate on what is happening inside the syntactic work space.

Each phrase, word, morpheme, or simply feature (bundle) transferred into the work space from the lexicon is recognized as a syntactic object, and can be used as input for Merge. Notice that it is also a structural root before merger, that is, the “top node,” often trivially so for simplex items. Now, two potential constraints on the input come to mind.

First, Merge could be restricted to roots by a simple and general rule. An object embedded inside a larger object (due to an earlier instance of Merge), a “term,” is no longer accessible. As a consequence, movement is impossible. Though such a limitation clearly has its advantages (mainly conceptual—see Koster 2007 for interesting comments), it is also clear that alternative explanations for (apparent) displacement phenomena then need to be found. Here, I prefer to adopt the standard view that movement, or rather, remerge, exists, and hence that terms of syntactic objects can be used as input for Merge again, at least within the same cycle.

Second, perhaps naively, it could be that the work space has only one “center of attention,” the root of the main structure. Each instance of Merge involves this root, and something else: any other root (normally, an unused lexical item) or a term to be moved. The root of the resulting structure is then the newborn center of attention. Although this has some initial plausibility, the strictest version of it must be discarded, considering that we need to be able to generate structures containing sisters that are both composed phrases. For example, the potential subject the old man and the predicate ate a sandwich are both phrases that must be derived before they can be combined into one clause. Therefore, more than one complex structure can be present in the work space, and in order to derive these and the combination
thereof, the center of attention has to shift at least once from one structure to the other.  

Summarizing so far, Merge is sensibly restricted to syntactic objects figuring inside the syntactic work space, but at first sight there are no further pressing limitations on the input. Thus, an input object can be a root (simplex or complex) or a term of an existing phrase. Now, suppose that A and B are merged at a certain point of the derivation, yielding C. Depending on the initial status of A and B, three possible configurations ensue. The first two are familiar; see (3):

(3) a. first-time merge   b. internal remerge

In (3a), A and B are unrelated roots before merger, each possibly complex (and hence the result of previous mergers). In (3b), A is originally a constituent of B, which implies that it has been input to merger before, in a previous step of the derivation. We can call this situation internal remerge, which corresponds with traditional movement. I will represent it using multidominance, thereby discarding artifacts such as traces or copies (cf. Gärtner 2002, among many others).

The third possibility, external remerge, is perhaps more surprising. It arises if a term is remerged with an external root. The result is depicted in (4):

(4) external remerge

This structure can be derived by two simple sequences of Merge. Either Merge (..., A) ! R is followed by Merge (A, B) ! C, or the other way around. This process has also been called grafting, Parallel Merge, sharing, or (abstracting away

---

2 An alternative perspective in terms of layered derivations involving a complex numeration/resource is sketched in Zwart (2011).
3 If the requirement is dropped that the center of attention is always a root (and hence that one of the input elements for Merge is a root), then two more possibilities follow, which can be called “quirky” internal and external remerge; see de Vries (2009b) for discussion.
from the issue of traces/copies/chains) *interarborial* or *sideward movement* (see, for example, van Riemsdijk 2006a, Citko 2005, Guimarães 2004, Bobaljik and Brown 1997, Nunes 2001), and dates back to earlier ideas by Williams (1978), McCawley (1982), Goodall (1987), and others. According to most authors, the doubly-rooted structure in (4) can only exist as an intermediate stage in the derivation. Eventually, the two roots need to be connected. For further references, see de Vries (2009b), which article also discusses the linearization of structures involving multidomiance—an important issue that I will ignore here for reasons of space.

A more formal characterization of the possible effects of Merge is provided in (5):

\[
\text{(5) Merge } (a, \beta) \rightarrow \gamma \text{ constitutes}
\]

\begin{enumerate}
\item \textit{first-time merge} iff $a$ and $\beta$ are independent roots before merger;
\item \textit{internal remerge} iff $\beta$ is a root and $a$ is included in $\beta$ (or the other way around) before merger;
\item \textit{external remerge} iff $\beta$ is included in some root $\delta$, and $a$ is an independent root (or the other way around) before merger.
\end{enumerate}

Notice that Chomsky’s original dichotomy “external merge” versus “internal merge” corresponds to (5a) and (5b).

At the cost of a complicated boundary condition on the output of Merge, we could exclude (5c) from the grammar, whilst maintaining (5b):

\[
\text{(6) Exclude external remerge}
\]

The output of Merge must be an independent syntactic object.

where \textit{independence} can be defined as follows:

A syntactic object $a$ is independent iff $a$ is not included in any other syntactic object, and no term of $a$ is included in an object that is not also a term of $a$.

However, let us put off such a condition, in line with the authors cited, and investigate to what extent external remerge is a useful concept.

Before we go on, let me add two remarks. First, it is hard to decide which grammar is more minimal: one with a more limited expressive power, or one with less—or less complicated—rules or constraints. Ideally, the two criteria do not interact, or they even enforce each other, but as I just indicated, the situation is not (always) like this. Second, there is a tradition of defining syntactic structure in terms of tree representations (see Carnie 2008 for an overview and references). One could then call upon the simple “single mother condition” in order to prevent the effect of external remerge: syntactic structures are trees and not graphs. However, from the present minimalist perspective, regular movement would also lead to a violation of the single mother condition—see (3b). Moreover, one needs to be careful in not confusing representations with the underlying theory. Consider the derivation of (7), where $A$ is internally remerged (moved).
Merge does not create trees, it creates (a list of) strictly local relationships. Sisterhood correlates with being merge-mates; dominance correlates with inclusion. Non-local relationships all depend on the transitivity of dominance. Of course the whole point of merger is to relate syntactic objects to each other, which leads to semantic composition, and which facilitates all kinds of direct and indirect licensing. Since an object can be involved in more than one direct relationship, it makes sense to assume that it can be remerged. In example (7), A is the merge-mate of both B and E. A rather straightforward (and for most human beings more insightful) representation of the list of basic relationships created by Merge is the graph provided in (7c). As in any representation, there are some arbitrary choices. Lines represent inclusion directionally, for instance. Traditionally, the merge-mate relationship is not directly represented by some kind of symbol—but it can be inferred. Furthermore, the position of remerged elements is arbitrary. In this example, we could also picture A in its remerged position, or in fact anywhere else on the paper, as is illustrated in (8a). It is probably best to put it in the eventual spell-out position. In (8b), I suggest a different type of notation: here, the essential merge-mate relationship is directly represented, namely by solid lines between objects. The inclusion relationship is indicated by dotted arrows, from parental projections to the lines connecting a directly included pair of merge-mates.

The representations in (7c) and (8) are intended as completely equivalent. Which one is to be preferred depends on personal preferences, traditions, and communicative purposes. The graphs in (7c) and (8a) highlight the multidominance effect of remerge. This might distract the attention from the more essential aspect of it,
namely that some object, here A, is involved in two merge-mate relationships; the alternative in (8b) does not have this potential disadvantage.

In Chomsky’s view (see also Hinzen 2006 and many others), the Merge operation is tantamount to set-formation. Though standard hierarchies can of course be represented by complex sets, the idea that sets are the foundation of syntax is problematic for several reasons. Importantly, there is no straightforward way of representing displacement by means of a complex set. It is for this reason that Chomsky proposed the copy theory of movement, which, as was already commemorated in the introduction, has been subject to fundamental critique—and rightly so in my opinion. But even apart from displacement, it is not obvious that sets alone could be the essence of syntax. After merger of A and B, there is a set/group \{A, B\}. This is to be interpreted as a syntactic constituent. From the notation \{A, B\} we may also want to infer that A and B are related to each other as syntactic sisters (with consequences for the interpretation, for syntactic licensing, etc.), but strictly speaking, the existence of a mathematically defined set of objects does neither imply nor facilitate any relationship between those very objects. If the idea is correct that upon merger syntactic objects can “see” each other and interact with each other (and indirectly, each other’s components) because they are sisters because they are merged together, then establishing a sisterhood relationship must be an essential part of what Merge does, next to just group formation. Notice, incidentally, that the notation \{A, B\}—though often referred to as a “set”—is not essentially different from a tree structure A \sqcup B or a bracketed structure [A B]: all are representations that can be interpreted as required, according to certain conventions in a certain application domain.

Another issue is the property of inclusion and projection. According to Chomsky, every group in syntax is a projection of one of the members. This does not follow in any way from a simple set, which is why he proposed that Merge creates a complex set, for instance \{A, \{A, B\}\} in case A projects. From this object, some kind of asymmetry between A and B can indeed be inferred, but which kind of asymmetrical relationship as well as its direction depends entirely on conventional interpretation.4

I conclude that what Merge does, minimally, is to create a group of associated sisters which defines a new, inclusive syntactic object. In other words, it establishes basic relationships between the input and output objects, as indicated in (7a/b), for

4 I am glossing over further complicating issues related to the status of labels and the Wiener-Kuratowski convention here. Independently of that, it is thinkable that there is more than one fundamental asymmetry, and that these do not coincide—which is also highly problematic for the set-only conception of Merge. Concrete candidates next to the asymmetry of inclusion and projection figuring in the literature are asymmetries between sisters in terms of syntactic dependency, selection, or simply precedence. See de Vries (2009b) for some discussion and further references. In the present chapter, I will not be concerned with possible asymmetries between sisters, and I will simply talk about merge-mates in order to avoid unnecessary intricacies.
Such relations ("direct inclusion," "merge-matehood") can be considered theoretical primitives, for which we then need a convenient representation.

Let us now return to external remerge. An abstract, minimal example similar to (4) is given in (9):

\begin{align*}
(9) \quad a. & \quad \text{Merge (A, B) } \rightarrow \text{ C} & \quad \text{A is the merge-mate of B} \\
& & \quad \text{C directly includes A and B} \\
& \quad \text{Merge (E, A) } \rightarrow \text{ F} & \quad \text{E is the merge-mate of A} \\
& & \quad \text{F directly includes E and A}
\end{align*}

Again, A is involved in two mergers, this time with two independent roots, B and E. After Merge, the new roots are C and F, respectively.

The effect of external remerge has been used to analyze head movement (Bobaljik and Brown 1997), parasitic gaps (Nunes 2004), Across-the-Board movement (Citko 2005), coordinated wh (Gračanin-Yuksek 2007), (transparent) free relatives (van Riemsdijk 2006b), cleft and wh-amalgams (Guimarães 2004), and several other construction types. The most widely discussed is Right-Node-Raising (RNR) (Wilders 2008, Chen-Main 2006, Kluck and de Vries to appear, among many others—based on earlier ideas by McCawley 1982). The basic idea is simple: in an example like \textit{Joop bought _ and Jaap sold a car}, the verb phrases in the first and second conjunct share the direct object \textit{a car}; thus, this noun phrase is merged as the complement of one verb, and then remerged with the other verb. The two clauses are completed, and finally joined by means of coordination. A schematic representation of the result is the following:

\begin{center}
\begin{tikzpicture}
  \node (Joop) at (0,0) {Joop};
  \node (and) at (1,-1) {and};
  \node (Jaap) at (2,-2) {Jaap};
  \node (bought) at (0,-3) {bought};
  \node (sold) at (2,-3) {sold};
  \node (a car) at (1,-4) {a car};
  \draw (Joop) -- (and) -- (Jaap);
  \draw (bought) -- (and) -- (Jaap);
  \draw (Jaap) -- (sold) -- (a car);
\end{tikzpicture}
\end{center}

One interesting aspect of RNR is that it shows apparently non-local behavior. An example is (11), where the ellipsis site is inside a relative clause embedded in the first clausal conjunct.\footnote{Note that RNR is coordination-final, not sentence-final. The entire CoP in (11), for instance, can be inserted as a complex subject: \textit{That Joop talked to a man who BOUGHT _ and (that) Jaap talked to a man who SOLD a red car is quite surprising.}}

\footnote{Compare also Koster's (2007) notion of a basic syntactic "triad."}
(11) [Joop talked to [a man who BOUGHT _ ]], and [Jaap talked to [a man who SOLD a red car]].

This possibility turns out to be an emergent property of applying external remerge, which can be used to create a bypass without actually violating locality constraints. For regular internal remerge, this is not the case. Put generally, (12a) can be derived unproblematically, but (12b) cannot.

(12) a. \ldots [\text{CoP} [\text{XP} \ldots [\phi \ldots \alpha \ldots]] \text{ Co} [\text{YP} \ldots [\phi \ldots \alpha \ldots]]] \ldots
b. *\ldots [\text{XP} \ldots [\alpha \ldots [\phi \ldots \alpha \ldots]]] \ldots

Here, \alpha is the shared/moved phrase, \phi is a locality barrier, and CoP in (12a) is the coordination phrase joining the complex parts XP and YP.

In a derivational grammar, locality rules can be implemented as restrictions on the input for Merge: material embedded in some locality barrier \phi is no longer accessible for selection. This view is compatible with the idea of syntactic phases/cycles. As a consequence, \alpha cannot be remerged with a projection dominating \phi in (12b) (but note that the story changes if there is an available edge facilitating successive cyclic movement). In (12a), the situation is different because \alpha can be remerged before \phi is created. In the concrete example (11), the verb bought is merged with the phrase a red car. No locality boundary is reached yet, and the noun phrase can be externally remerged with the other verb, sold. This yields a simple doubly-rooted structure as in (9). Both temporary roots are expanded by Merge, successively leading to relative clauses, complex noun phrases, and matrix clauses. Finally, these are combined by coordination into one complex sentence. The result—strongly abbreviated—is depicted in (13):

(13)
The DP a red car is locally related to both verbs by Merge. The fact that the relevant VPs are deeply embedded inside the conjoined clauses in the final stage of the derivation creates a sense of non-locality, but if the analysis is correct, this is only apparently so.

In section 7.4, we will see that certain other construction types involving external remerge also show signs of non-local behavior.

### 7.3 Parenthetical Merge

The possibility of internal and external remerge depends on the choice of input for Merge. The operation of Merge itself is the same in all cases. Here, I will argue that we nevertheless do need a second type of Merge for another purpose, namely to account for the phenomenon of parenthesis.

Parenthesis is a cover term for a wealth of construction types, including parentheticals, comment clauses, appositions, hedges, and non-restrictive relative clauses (see Dehé and Kavalova 2007 for an overview). The internal structures of these vary greatly, but the important point is that they also have something in common. All behavior that is typical for parentheses, I believe, follows from one essential factor, namely the particular way they are related to the matrix. This general idea is not very controversial, but the opinions differ wildly on how to give it hands and feet. To begin with, it is far from obvious how to define parenthesis either syntactically or phonologically, even though everyone recognizes it intuitively. For this reason I proposed the following working definition: parenthesis is a grammatical construction type that involves a message that is presented or perceived as secondary with respect to the host, where message covers propositions, modal propositions, questions, metalinguistic comments, and so on. This definition ties in with Potts’s (2005) conclusion that appositions and other “conventional implicatures” involve independent lambda terms that are not “at issue.” As an example, consider (14), which contains an appositive (i.e. non-restrictive) relative clause (ARC):

(14) You probably know Joop, who is my neighbor.

The primary message is the host clause you probably know Joop; the secondary message is he (=Joop) is my neighbor. The parenthesis cannot be denied directly. Thus, if the hearer responds “No, that is incorrect,” this means that he or she does not know Joop; it does not mean that Joop is not the speaker’s neighbor. Furthermore, the ARC is outside the scope of the modal operator probably: what is presented as probable is the hearer

---

7 Some of the literature on multidominance, and specifically on external remerge, gives another impression about this, which I think is somewhat misleading. Also, note that the phenomenon of remerge is independent of dimensionality. Multidominance (which results from doing the same thing twice or more) and 3D-structure (which involves another kind of basic relationship, as is the case for Parenthetical Merge discussed below) are independent factors.
knowing Joop, independently of the issue whether Joop is the speaker’s neighbor. What is more, the semantic composition of the host clause is complete by itself. A parenthesis is always a non-restrictive addition. But of course it can have an important pragmatic import. As for the phonology, parentheses normally are prosodically distinguished from the host. How exactly is subject to variation, but a change in pitch and speech rate are common factors. See Dehé (2009) for discussion and a qualification.

A well-known idea is that parentheses are “orphans,” which means that they are not syntactically integrated with the host clause at all (see Haegeman 1991, among others). Given the above, this does not sound entirely unreasonable. However, it also seems like giving up finding an explanation for parenthesis (or at least a formal description of it), and I am more sympathetic to approaches represented by, e.g., Espinal (1991) and Ackema and Neeleman (2004), who define special devices in order to cope with the extraordinary behavior of parentheses.\footnote{Note that these proposals differ substantially from each other, and also from my own, to be spelled out below.} Apart from that, there are several problems with the idea of orphanage. The most straightforward is this. Parentheses, obviously, are both interpreted and pronounced (and thus have an effect on the phonology and the semantics of the utterance as a whole). Therefore, they must be present at the PF and LF interfaces. Given the standard Y-model of grammar (or any variation of (2)), the only way to get there is via the syntax. If a parenthesis were to be added, say, at some unspecified discourse level beyond LF, this would be problematic for the pronunciation (unless, perhaps, a completely different model of grammar is assumed, but that would require elaborate independent justification). In this respect, it is worth stressing that parentheses often interrupt the host clause: they need not be sentence-peripheral. Second, orphanage does not explain the fact that parentheses are secondary information. Third, some parentheses, such as appositions and ARCs, are directly attached to an anchor/antecedent, and they seem to form a constituent with it. In Dutch, they can be topicalized together with the anchor, but stranding in the middle field is impossible. This is illustrated in (15):

\begin{enumerate}
\item \textbf{a.} Ik heb [Joop, \textit{onze buurman},] gezien. \\
I have Joop our neighbor seen \\
‘I’ve seen Joop, our neighbor.’
\item \textbf{b.} [Joop, \textit{onze buurman},] heb ik \_ gezien. (topicalization)
\item \textbf{c.} *Joop heb ik [\_, \textit{onze buurman},] gezien. (stranding)
\end{enumerate}

Such facts suggest that a parenthesis can be added on the constituent level, which implies that syntax plays a role. This is corroborated by the ambiguity of (16), where the comment clause \textit{I think} can be related to the host clause as a whole or to the direct object alone:
Next week, Joop is going to Paris, I think.

(i) "I think that Joop is going to Paris next week."

(ii) "I think that it is Paris that Joop is going to next week."

Fourth, it is remarkable that appositions take over the Case of the anchor in many languages, which mimics the situation in coordination. Example (17) is from Russian (thanks to Evgenia Markovskaya and Herman Heringa). See Heringa (2011) for more examples and discussion.

(17) Oni dali ih Anite, drugomu pauku, tože.

They gave it to Anita, the other spider, as well.

Finally, the simple fact that there can be parenthesis within parenthesis (within parenthesis, etc.) implies recursion—again an indication for the involvement of syntax. An example is (18), where we can distinguish five levels of interpretation:

(18) I still owe Anna—and Anna, who hit Joop, an unpleasant guy, as you know, disappeared last night—250 dollars.

I conclude that parenthesis needs to be represented in syntax.

How can a parenthetical phrase or clause be connected to its host? Clearly, regular Merge is not fit for the job, since it automatically leads to semantic composition and a dominance hierarchy. Scope, and in fact c-command-based relationships in general, are defined over dominance. But parentheses do not interact with the matrix in such terms. Compare, for instance, the contrast in (19a/b), which tests quantifier binding of a variable embedded in a restrictive and appositive relative clause, respectively. Binding is blocked in the latter. Furthermore, Condition C effects can be lifted in parentheticals; see (20), where coreference is perfectly acceptable.

(19) a. [No climber] talked about the mountain he conquered last month.

b. * [No climber] talked about the K2, which he conquered last month.

(20) He said—that is typical for Joop—that he didn’t like veggie burgers.

Thus, phrases in the matrix do not c-command into parentheses. See de Vries (2007b) for a systematic overview and further discussion.

Parentheses are syntactically included in the host sentence, but they are “invisible” for scopal relationships. How do we solve this dilemma? Traditional grammarians distinguished between two fundamental relationships, hypotaxis (subordination) and parataxis (nonsubordination). Similarly, let us assume that there are two basic types

9 Notice that c-command stretches across phases. Not only is this potentially problematic for phase theory in itself, it also shows that simple solutions for the scopal invisibility of parenthesis in terms of syntactic phase domains are not going to work—whence my turn to a more fundamental solution below.
of inclusion: regular inclusion (which corresponds to traditional dominance) and
parenthetical inclusion. These notions do not follow from anything else: they are
simply primitives of the grammar. As a consequence, there are also two kinds of
Merge operations: (regular) Merge, and Parenthetical Merge (which I will abbreviate
as par-Merge):10

(21) a. (regular) Merge (A, B) yields C such that
   (i) C directly (regularly) includes A,
   (ii) C directly (regularly) includes B, \textit{and}
   (iii) A is the merge-mate of B.

b. par-Merge (A, B) yields C such that
   (i) C directly par-includes A,
   (ii) C directly par-includes B, \textit{and}
   (iii) A is the merge-mate of B.

Both types of Merge form a more inclusive object, and hence a constituent. But we
want to be able to say that there are two qualitatively different ways of forming a
constituent. Notably, both a parenthetical and its host clause are formed internally by
regular Merge, but the connection between the two requires an instance of par-Merge.

Before we get to examples, let us discuss the crucial notion of c-command. Put
somewhat informally, every node c-commands its sister and everything in it—a well-
founded idea which we will maintain. We can now define c-command more precisely
over the basic relationships just associated with regular Merge:11

(22) a. A c-commands B iff there is an X such that
   (i) A and X are merge-mates, \textit{and}
   (ii) B = X, or B is regularly included in X.

b. B is regularly included in X iff there is a sequence ("path")
   \((Y_1, \ldots, Y_n)\), where \(n \geq 2\), \(Y_1 = B\), \(Y_n = X\), such that
   for all \(i, 2 \leq i \leq n\): \(Y_i\) directly regularly includes \(Y_{i-1}\).

Here, (22b) simply spells out the transitivity of inclusion.

10 In some previous work (de Vries 2005a, b, 2007b), I referred to the latter as b-Merge, with \(b\) short for
behindance, thereby alluding to a three-dimensional grammar. Since Carnie (2008: 207, fn.) objects to this
neologism (incorrectly attributing it to me, though I took it over from Grootveld 1994), let us maintain the
terminology in the main text, which emphasizes the use instead of the (potential) representation. In this
respect, it is worth mentioning that in Bosveld-de Smet and de Vries (2008), we tested the usability of
different representation types applied to complex coordinate structures (also involving Right-Node-
Raising). The 3D variants did not fare too well.

11 Clearly, c-command can then be viewed as a direct function of regular Merge (cf. Epstein 1999),
which may explain why it is this higher-order relationship that is so important, and not other logically
possible ones. This result was also obtained by Kayne (1984) in terms of unambiguous paths.
Crucially, a parenthetically merged syntactic object is not regularly included (by definition). Consequently, it will not be c-commanded by material merged later in a bottom-up derivation. Consider the abstract example in (23), where F, A, B, and D can be complex. We now need a representational convention in order to distinguish Parenthetical Merge from regular Merge. Here, I will indicate it by stars next to the vertical inclusion line.

(23) \[ \begin{array}{c}
G \\
\text{par-Merge (A, B)} \\
\text{Merge (C, D)}
\end{array} \]

In this structure, F c-commands D, which is part of the host structure, but not A and B because those are not regularly included in F’s merge-mate E. Put differently, par-Merge breaks the transitive line of regular inclusion. More realistic example configurations will follow.

The above is a technical implementation of the invisibility effects illustrated in (19b) and (20). Of course, it is only the beginning of a complete solution for the phenomenon of parenthesis. Let me sum up what I intended to show so far: (i) parenthesis is fundamentally different from hypotaxis; (ii) nevertheless, parenthesis must be represented in syntax; (iii) regular Merge cannot possibly account for this; (iv) therefore, some basic stipulation is inevitable; (v) once we postulate “parenthetical inclusion” as a primitive relationship, the required effects follow straightforwardly and parsimoniously.

There is no fixed position for parentheses in the host clause, and in fact there cannot be such a position, since selection does not play a role. Therefore, let us assume that “independent” parentheses are freely adjoined to the spine of the structure they pragmatically interact with.\(^{12}\) Suppose we would do this directly, as in (24), where XP\(_{\text{par}}\) is the parenthesis, and YP some projection of the host.

\(^{12}\) One may wonder if there is a correspondence between par-Merge and Chomsky’s pair-Merge. The answer is probably negative. There are a number of complications in answering this question. First, it is not entirely clear, to me at least, how the properties attributed to adjuncts follow from the formal characterization of par-Merge (against the background of other assumptions within the Minimalist framework, which is constantly developing). Second, I am not sure that Chomsky’s (2004) empirical description of the behavior of adjuncts is correct in all respects. Third, it is unknown to me how Chomsky would explain the differences between regular adjuncts and parentheses. It is undisputed that adjuncts take part in the meaning composition of the sentence (“predicate intersection”). As we have seen, this is different for parentheses. Furthermore, despite the existence of some complex anti-reconstruction cases Chomsky cites, it seems clear to me that adjuncts do not inherently display invisibility effects such as those illustrated for parentheses in (19b) and (20). Straightforward examples like (i), where the anaphor inside the adjunct is
This, however, has an unacceptable consequence, namely that the lower part of the host becomes invisible for the higher part. A solution is to assume that $\text{XP}_{\text{par}}$ is embedded in an abstract parenthetical phrase $\text{ParP}$. I will argue that this has several advantages. The essential part of the derivation and structure is given in (25):

Here, the line of regular inclusion is unaffected in the host $\text{YP}$, but it stops immediately within $\text{ParP}$, with the required effect on $c$-command.

The head $\text{Par}$, which triggers the application of par-Merge, can be thought of as a specialized discourse connector, a “parenthetical specifying coordinator” (de Vries 2009a), whose semantic effect can be compared to the “comma operator” introduced by Potts (2005). Recall that parentheses act as “independent lambda terms.” As I see it, such semantic effects cannot come out of nowhere: they are evoked by elements in the syntax. I suppose that a similar consideration underlies Potts’s introduction of a “comma feature” in syntax. Just to be clear: there is nothing special about the internal semantics of some parenthesis $\text{XP}_{\text{par}}$; it is only the connection with the host that is of a particular nature. We saw that par-inclusion starts a new $c$-command domain. In semantics, then, this information may trigger the start of a new lambda term. The fact that the parenthetical is syntactically embedded in the host makes sure that it is the parenthetical proposition/message that is perceived as secondary, and not the other

bound by the subject, or (ii), where a pronoun inside an adjunct is bound by a quantifier, or (iii), where Condition C is violated, show convincingly that adjunction does not shield off $c$-command:

(i) John and Mary had breakfast [in each other’s house].
(ii) [Every man], peeled an orange [with his own knife].
(iii) *He, bought a picture [that John liked].

Moreover, adjuncts are phonologically treated as regular constituents of the clause, and as such integrated in the overall intonation contour of the sentence, again, unlike parentheses. It may be, then, that adjuncts have to be integrated into the structure by some operation different from regular Merge (see also Rubin 2003, among others, for comments), but this operation and its consequences must be kept distinct from par-Merge.
way around. Similarly, the information that XP is actively marked as “parenthetical” in syntax can be used by the phonological component to start a new intonational phrase with somewhat different properties.

Usually, a specifying coordinator is asyndetic, but in some cases it can materialize as an actual conjunction. An example is in (18); see further Blakemore (2005) and Kavalova (2007) for a discussion of and-parentheticals. What is more, it is now easier to see the link between independent parentheticals and anchored parentheses like appositions, which show clear signs of structural coordination (cf. Kraak & Klooster 1968, Quirk et al. 1985: 1308, Koster 2000). A simple illustration is Elvish Presley, or The King. Appositions then—and appositive relative clauses alike—are not adjoined to the host, but structurally coordinated to the anchor: $[\text{ParP } \text{anchor} [\text{Par } \text{appositive } \text{XP}]]$; see O’Connor (2008), Cardoso and de Vries (2011), and Heringa (2011) for a more sophisticated discussion of the internal structure. Again, and perhaps more evidently so, we need the parenthetical specifying coordinator Par, as is shown in (26):

$$
\begin{align*}
\text{YP}_\text{anchor} & \quad \downarrow \quad \text{Par'} \\
\text{ParP} & \quad \downarrow \quad \text{par-merge (Par, XP}\text{app}) \rightarrow \text{Par'} \\
& \quad \downarrow \quad \text{Merge (YP}\text{anchor, Par'}) \rightarrow \text{ParP} \\
& \quad \downarrow \quad \text{XP}_{\text{app}}
\end{align*}
$$

Here, ParP is structurally like a regular coordination phrase. The anchor is the first conjunct, and it is visible for the host. The apposition is embedded as the parenthetical second conjunct, and hence invisible for c-command.

On a final note, the involvement of the head Par can be used to reduce the amount of overgeneration (that is, nonsensical, probably crashing, derivations due to the availability of the additional operation par-Merge) by introducing the general heuristic “the parenthetical specifying coordinator Par, and only Par, triggers the application of par-Merge.”

### 7.4 Combining Parenthetical Merge and external remerge in amalgams

This pre-final section is a brief exploration of the combined effects of Parenthetical Merge and external remerge.

There are various ways of intertwining sentences (see Lakoff 1974, Guimarães 2004, van Riemsdijk 2006a, Kluck 2011, among others). The illustrations in (27) contain a cleft amalgam (“Horn’s case”) and a wh-amalgam (“Andrews’ case”), respectively:

$$
\begin{align*}
\text{a. Joop got I think it’s a didgeridoo for his birthday.} \\
\text{b. Joop got you will never guess how many instruments for his birthday.}
\end{align*}
$$
In both examples there is an interrupting clause (printed in italics) with modal import. Unlike the situation in regular parentheticals, there is a phrase (underlined) that seems to function as the pivot between the host clause and the interrupting clause. This pivot (“callus,” “content kernel”) plays a role in both clauses at the same time, which leads to a bracketing paradox. As will be clear from section 7.2, the application of external remerge is a potential solution to this problem: the pivot can then be shared between two clauses. The general idea is depicted in (28):

(28)  

\[
\text{main clause} \quad \downarrow \quad \text{pivot} \quad \downarrow \quad \text{interrupting clause}
\]

Without claiming that this is ultimately the right analysis of amalgams and related construction types, I would like to adopt it for now, and show that it leads to some interesting predictions.\(^{13}\)

What then needs to be established is the relationship between the interrupting clause itself and the matrix (strangely, this issue is ignored in some of the literature on the topic). Since there is no selectional relationship whatsoever, and since the interrupting clause intuitively conveys a secondary message, I will treat the interrupting clause as a parenthetical, which, according to the previous section, is an adjunct involving Parenthetical Merge inside a ParP. This ParP must then be adjoined directly above the pivot. Abstracting away from numerous details, the structure of amalgamated sentences is as sketched in (29):

(29)  

\[
\begin{align*}
\text{matrix} & \quad \downarrow \quad \text{XP} \\
\text{ParP} & \quad \downarrow \quad \text{XP} \\
\text{Par} & \quad \downarrow \quad \text{IC} \\
\text{Y} & \quad \downarrow \quad \text{ZP}
\end{align*}
\]

\[
\begin{align*}
\text{Merge (Y, ZP)} & \rightarrow \text{YP} \\
\text{Merge (X, ZP)} & \rightarrow \text{XP} \\
\text{Merge (..., YP)} & \rightarrow ..., \text{IC} \\
\text{par-Merge (Par, IC)} & \rightarrow \text{ParP} \\
\text{Merge (ParP, XP)} & \rightarrow \text{XP} \\
\text{Merge (..., XP)} & \rightarrow ..., \text{matrix}
\end{align*}
\]

\(^{13}\) Kluck (2011) objects that a sharing analysis does not adequately reflect the semantics of cleft amalgams. She proposes an interesting alternative in terms of ellipsis (sluicing).
Here, IC is the interrupting clause, and ZP the pivot, the shared constituent. More evidence for (29) will follow below. First, let me provide a concrete example derivation of a simple cleft amalgam, as in (27a): Joop got I think it's a didgeridoo for his birthday. First we could merge the noun phrase a didgeridoo with a high functional head associated with noun phrases, D. Immediately, a didgeridoo can be externally remerged with another D. There are now two DPs, which share the noun phrase a didgeridoo. One DP is used as a small clause predicate in a cleft construction, and we generate it's a didgeridoo by regular Merge, and consequently derive the more inclusive clause I think it's a didgeridoo. This is turned into a parenthetical by parmerging it with the specifying coordinator Par. ParP as a whole can then be adjoined inside the main clause (or rather what is to become the main clause) by merging it with the other DP (which is still a root at that stage of the derivation). The now complex DP is merged as the object argument of the verb got. Finally, the subject argument of the verb got can be added, and the main clause is functionally completed. The result is sketched in (30), in a hybrid notation:

\[
(30) \quad [\text{TP} [\text{DP} \text{Joop}]_i [T [\text{vP} \text{t}_i \text{got} \text{DP} ] \text{for his birthday}]]
\]

For practical purposes, movement (internal remerge) of the subjects is indicated by simple traces in the bracketed parts of the representation. Needless to say, the details of English clause structure, small clauses, and noun phrases do not concern us here.

To complete the picture, a structural sketch of the wh-amalgam in (27b)—Joop got you will never guess how many instruments for his birthday—is given in (31). Wh-amalgams are somewhat more complicated than the cleft variant since they contain a sluiced clause; this is indicated by <TP>, in which the displaced wh-phrase is base-merged; in this case it stands for Joop got _ (for his birthday). Moreover, note that in this example how many is a degree phrase related to the pivot NP.

---

14 There are indications that what is shared is just a predicate, and that the XP that the ParP is actually adjoined to is a functional projection on top of this predicate, e.g. a strong determiner phrase turning the predicate into an argument in the matrix.
(31) \[ [TP \{DP \text{Joop}\}\_i [\{T [\{vP t_i \text{ got [D NP]}\} for \text{his birthday}]\}]] \]

Let us now turn to the consequences of such an analysis.

In the previous sections, we derived some important effects of external remerge and Parenthetical Merge, namely non-local behavior and invisibility for c-command, respectively. If (29) or the concrete examples in (30) and (31) are approximately correct, three predictions ensue; see (32).

(32) (i) **Apparently non-local behavior of amalgams**: the pivot can be deeply embedded inside the interrupting clause.

(ii) **Invisibility of the interrupting clause in amalgams**: material inside the interrupting clause (modulo the pivot) is invisible for c-command-based relationships with phrases in the host structure.

(iii) **Visibility of the pivot in amalgams**: the pivot is visible for c-command-based relationships with phrases in the host structure.

Indeed, each of these can be empirically confirmed. First, the examples in (33) and (34) show that the pivot can be embedded in an extended IC.\(^{15}\)

(33) a. Joop got *I think Jaap claimed it’s a didgeridoo* for his birthday.

b. Joop got *I think that it was Jaap who claimed that it’s a didgeridoo* for his birthday.

c. Joop got *I guess I have to convince you that it’s a didgeridoo* for his birthday.

\(^{15}\) In order to avoid processing difficulties, such examples require an increased speech rate within the IC, and emphasis on the pivot. See de Vries (2010) for some examples in Dutch.
(34)  a. Joop got *I am sure you will never guess how many **instruments** for his birthday.
    b. Joop got *I guess there’s nobody here who can even imagine how many **instruments** for his birthday.

As in complex Right Node Raising configurations, the distance between the pivot and the selecting head in the host structure can apparently be large, measured by normal syntactic standards. In (33b), for instance, *a didgeridoo* is deeply embedded inside the interrupting clause: it is part of a cleft construction inside a relative clause inside a subordinate cleft construction. Needless to say, movement (internal remerge) across such a combination of locality domains would be completely impossible. Here, movement is not at stake. Still, *a didgeridoo* is selectionally related to the verb *got* in the main clause, which is at least linearly far away. How is this possible? The answer must be that the main verb does not “see” the object via a long path through the complex interrupting clause, but directly. The object is syntactically shared between the IC and the host, so there is a bypass. The derivation is roughly as follows. The noun phrase *a didgeridoo* is merged as an internal argument in what is to become the host, and then immediately externally remerged as a small clause predicate in what is to become the IC (or the other way around). There are now two roots. The IC is extended by a (long) series of mergers, thereby creating internal locality boundaries. When it is completed, it can, as a whole, be parenthetically adjoined to the other root (which represents an intermediate stage in the derivation of the host). The host is then completed in a normal fashion. Every step of this derivation is local. The result is like (29), with a complex IC.

Next, let us examine the invisibility of the interrupting clause. The examples in (35) through (38) are taken from Kluck (2008b), enhanced with annotation. In (35) it is shown that quantifier binding of a variable inside the IC is excluded. This can be compared to the situation in (19b).

(35)  a. *[No one]*$_i$ is going to *he$_i$ thinks it’s **New York** this Sunday.
    b. *[Every student]*$_i$, bought *he, didn’t realize how many **books** that day.

By contrast, the examples in (36) show that binding of a variable inside the pivot is acceptable:

---

16 Tsubomoto and Whitman (2000) do report an island effect for English cleft amalgams, (not for wh-amalgams). However, their examples (such as “?*John is going to [NP e] I got angry because it was Chicago on Sunday*” 2000: 179) are unacceptable for independent reasons. For instance, the highest predicate of the IC cannot be factive. Note that the cited example considerably improves if one more layer of embedding is added: *John is going to— I suspect I/Mary got angry because it was CHICAGO on Sunday*. See also Kluck (2011), who discusses this and related issues in more detail.

17 The judgments of all the data in this section can be corroborated by similar sentences in Dutch, according to my own intuition (confirmed by some colleagues).
a. [Every student], sold you can imagine how many of his books that day.

b. [No one], is going to I think it’s his girlfriend this Sunday.

As for binding Condition C, (37) and (38) show that the IC is opaque, except for the pivot. Thus, binding causes unacceptability in (38), which seems to imply that the pivot is visible. However, the rest of the IC is opaque, so the results in (37) can be compared to those in (20), a regular parenthetical, which may contain a coreferential R-expression.

(37) a. He had seen—Ed, said it was Anna on TV yesterday.

b. He bought—Ed, didn’t even know (himself) how many books.

(38) a. *He had seen I think it was Ed on TV yesterday.

b. *He bought you can imagine how many books about Ed.

Similar contrasts can be obtained with tests for anaphor binding and other scope-related phenomena. Furthermore, van Maastricht et al. (2010) performed tests with so-called “transparent free relatives” (another type of amalgamated construction) in Dutch, again with very similar results.

To sum up, the analysis of amalgams in terms of external remerge of the pivot plus Parenthetical Merge of the interrupting clause gives rise to a number of specific predictions, which can be tested. The results are very promising.

Needless to say, there are numerous aspects of these construction types that require further inquiry, both empirically and theoretically. On a final note, let me say something about theta role assignment. Since Chomsky (1981), it is generally assumed that each argument bears exactly one theta role. This might be a problem for a sharing approach to amalgamation, but I think this is not necessarily so. In fact, the theta criterion can be seen as support for the proposal above in which external remerge takes place below the DP level. Even apart from that, it is the case that in cleft amalgams (and transparent free relatives) the relevant noun phrase is used as a small clause predicate in the interrupting clause, hence not as an argument. In the examples above, it is an argument in the main clause, where it receives its one and only theta role.

7.5 Conclusions

Reasoning from a Minimalist perspective involving bottom-up derivations, I discussed the nature of the operation Merge and the possibility of certain unconventional mergers. The regular Merge operation has at least three possible effects, depending on the choice of input. Since the difference is relevant, I introduced the rather transparent labels first-time merge, internal remerge, and external remerge. It would be a misconception to think that these stand for different types of
operations: Merge itself does the exact same thing in each case, namely it relates two syntactic objects, thereby forming a new hierarchical layer (which is then a new structural root). The possibility of remerge depends on the degree of freedom in selecting the input. If it is allowed to select a term of a non-trivial syntactic object, which by itself is also a syntactic object, remerge follows. At the cost of an additional rule, this could be prevented, resulting in a grammar of more restricted expressive power. It is hard to tell which choice is more Minimalist in spirit. I argued for the more liberal stance on other grounds. Internal remerge, then, is equivalent to regular movement. And if there is internal remerge, only a complicated rule could prevent external remerge (also described in terms of “Parallel Merge,” “sideward movement,” “grafting,” “sharing,” or “interarboreal movement”).

Discarding traces and copies a priori, a multidominance graph may be used to represent the effect of remerge generally. However, although such graphs as well as standard trees represent the hierarchical effect of Merge well, they do not directly reflect the core aspect of Merge, namely that it relates the syntactic input objects to each other (leading to semantic composition), thereby fulfilling and/or facilitating all kinds of licensing. For this reason, I put forward an alternative representation type that does emphasize the “merge-mate” (sisterhood) relationship. Remerge, whether internal or external, simply means that some object can be directly related to more than one other object. To me, this seems an entirely natural assumption. Multidominance is only a secondary effect. In the past, the aversion of unconventional (complex, clumsy, weird,…) representations may have lead to overhasty decisions concerning what we now call remerge. But what counts, of course, is the underlying theory, and not the practical problem of convincingly representing the results.

External remerge leads to a doubly-rooted structure. During the syntactic derivation this is harmless. However, from a bird’s eye perspective it constitutes a temporary problem that must be resolved before the structure can be processed in the phonological and semantic components. The linearization procedure requires a single-rooted structure (unless, perhaps, complicated additional assumptions are made). As for the semantics, the relationship between the two substructures must be made explicit. In this chapter, two such relationships have been discussed: coordination (namely in Right-Node-Raising constructions) and parenthetical insertion (in amalgams). Furthermore, I showed that derivations involving external remerge potentially give rise to an interesting phenomenon: apparently non-local behavior. By contrast, it follows from the nature of Merge that locality boundaries cannot be bypassed in configurations involving internal remerge.

In section 7.3, I argued that standard Merge, with its three instantiations depending on the input, is insufficient to account for all relevant language data. The problem is that Merge automatically creates a hypotactical structure (standard embedding). This raises the question what to do with parenthesis, a vast phenomenon encompassing dozens of different construction types, which have a number of related
properties in common: they are not selected for by anything in the host structure, they are not embedded in the regular sense, and they are scopally independent. Despite that, there are good arguments to deal with them at the level of syntax: parentheses have effects on both the PF and LF interface, they are linearly integrated with the host, they can apply at the constituent level or the matrix level, they can be recursively added. Therefore, I proposed a second type of merger, Parenthetical Merge. As far as I can see, there is no way of explaining parenthetical inclusion in more basic terms, which is why it needs to be stipulated as a primitive of the grammar. Technically, par-Merge breaks the transitive line of dominance in the host structure, and consequently shields parentheses from c-command-based relationships with material higher up in the host. This type of “invisibility” can be tested on the basis of various phenomena involving scope, such as variable binding by a quantifier, or Condition C effects. Furthermore, I associated par-Merge with the presence of a functional projection ParP, which is a parenthetical specifying coordination phrase. Concretely, a parenthesis is always the complement of the head Par. For anchored parentheses such as appositions, I assumed that they are structurally coordinated with the anchor. “Independent” parentheses (parentheticals) can—in principle—be adjoined to any projection of the host, which explains their variable surface position.

Section 7.4 explored an analysis of sentence amalgamation, in particular cleft amalgams and wh-amalgams, as parentheticals of which the pivot constituent is structurally shared with the host. Thus, the derivation of these construction types involves par-Merge (at the top) as well as external remerge (at the bottom). From this, three predictions follow: first, potentially non-local behavior resulting from the application of external remerge, meaning, in this case, that the pivot can be deeply embedded inside the interrupting clause; second, scopal invisibility of the interrupting clause; and third, the exception of the pivot from this. All three were found to be correct on the basis of examples from English (shown here), and have been confirmed by speakers of Dutch.
Part II

Triggers and constraints
This page intentionally left blank
“Lasnik-Effects” and String-Vacuous ATB Movement*

JUN ABE AND NORBERT HORNSTEIN

8.1 Introduction

The main question we raise in this chapter is why string-vacuous movement is insensitive to locality effects. We argue that even though such movement creates a normal chain under the single cycle hypothesis, it is the bottom copy that is pronounced due to a prohibition to the effect that overt movement must have a PF effect. Given that such locality effects as caused by the Right Roof Constraint (RRC) and island conditions apply only to “overt” movement, it follows that string-vacuous movement does not show locality effects. In this way, we support Sabbagh’s (2007) Across-the-Board (ATB) movement analysis of Right-Node-Raising (RNR); that is, the latter always involves ATB movement but when the movement involved in the right conjunct is string-vacuous, it becomes immune from “overt movement” locality effects while still behaving as if the shared element scopes out of the conjoined domain “covertly.” We implement this idea in a Lasnikian (2001) fashion: copy deletion as it applies within ATB chains results in the elimination of “offending” copies, thereby “saving” otherwise illicit derivations. We show that such a refinement of Sabbagh’s analysis accommodates Japanese Left-Node-Raising (LNR) constructions, the mirror image of English RNR structures.

* We are indebted to Myriam Uribe-Etxebarria and Vidal Valmala, the editors of this book, and two anonymous reviewers for their invaluable comments on an earlier version of this chapter. The first author is grateful to the Department of Linguistics of the University of Maryland for allowing him to spend a year from the summer, 2008, for his sabbatical and for making it possible to work on the material reported here with the second author. The second author is grateful to the first author for deciding to spend his sabbatical at the University of Maryland and for discussing his preliminary ideas with him.

1 This fits with the proposal in Chomsky (1995b) that movement have interface effects. It interprets this requirement as requiring that bottom copies get pronounced in cases of string-vacuous movement.

2 This follows the view of islands proposed in Lasnik (2001), Merchant (2001), Hornstein, Lasnik, and Uriagereka (2003).
The chapter is organized as follows. In section 8.2, we outline Sabbagh’s ATB rightward movement analysis of RNR and point out some conceptual and empirical problems. In section 8.3, we make a new proposal that aims to support Sabbagh’s ATB movement approach to RNR by devising such a mechanism as mentioned above that is free from the problems pointed out in section 8.2. In section 8.4, we argue against the so-called Right Edge Restriction on RNR, which requires that the shared element of RNR must be located at the right edge of each conjunct before ATB rightward movement takes place. In section 8.5, we discuss some consequences of our mechanism of movement regarding (i) argument/adjunct asymmetry of shared elements in RNR and (ii) what Chung et al. (1995) call the merger type of sluicing.

8.2 The ATB movement analysis of Right-Node-Raising: Sabbagh’s (2007) approach

8.2.1 Locality insensitivity of Right-Node-Raising

Ross (1967) observes what has been known as RNR, a typical instance of which is illustrated in (1a) below, as exceptional cases to Coordinate Structure Constraint, and proposes the ATB movement analysis according to which the shared element undergoes movement out of each conjunct in an ATB fashion and is right-adjoined to the whole coordinate structure, as shown in (1b):

(1) a. John likes __ and Bill hates __, the linguistic professor teaching Ling 101.

b. [Clause [Clause John likes __] and [Clause Bill hates __] [the linguistic professor teaching Ling 101]]

Since Ross proposed this analysis, a major obstacle to it has been how to characterize the fact that RNR is immune to the bounding conditions, including island conditions, imposed upon rightward movement. Thus, it is free from the clause-boundedness condition on rightward movement; compare (2) and (3):

(2) *Max said [that he was going to return __ to the library] yesterday each of the books that he checked out last week. (Sabbagh 2007: 350)

(3) Josh said that he thought that I should sell __, and Jamie said that she thought that she might want to buy __, each of the Rambaldi artifacts that I have in my attic. (ibid.: 358)

Further, it has been well known since Wexler and Culicover (1980) that RNR is insensitive to islands in general, as shown below:
(4) I know someone who wants to buy __, and you know someone who wants to
sell __, a copy of this manuscript. (ibid.: 352)

(5) Josh wonders who bought __, and Bill will find out who sold __, pictures of
Fred. (ibid.: 382)

(6) Politicians win when they defend __, and lose when they attack __, the right of
a woman to an abortion. (ibid.: 382)

Relating to this immunity to locality conditions is the fact that RNR allows
P(reposition)-stranding, though rightward movement does not allow it; compare (7)
and (8):

(7) *Jamie walked into __ suddenly, the dean’s office. (ibid.: 350)

(8) Joss walked suddenly into __, and Maria stormed quickly out of __, the dean’s
office. (ibid.: 351)

However, it is not the case that RNR is always immune to the locality conditions in
question or the constraint on P-stranding; this only holds for those cases where the
original position of the shared element occupies the right edge of each conjunct.
Compare (3) and (8) with the following RNR examples:

(9) *Joss said [that he was going to donate __ to the library] yesterday, and Jamie
claimed [that she would donate __ to the museum] last week, a large collec-
tion of ancient texts.

(10) *I sent one of the books to __ in perfect condition, and sent the other ones to
__ in poor condition, the Somerville public library. (ibid.: 355)

Further, Sabbagh (2007) observes that even though a shared argument of RNR
originates at the right edge of each conjunct, application of ATB rightward move-
ment to it is bounded if it crosses a modifier that modifies the whole coordinate
structure, as shown below:

(11) a. Joss will [sell __ to a library, and donate __ to a shelter] on the same day, all
of his old manuscripts. (Sabbagh 2007: 356)

b. *Joss [said that he was going to fire __, and insisted that no one would ever
consider rehiring __] on the same day, the crazy guy from accounting.

(ibid.: 358)

Here on the same day modifies the whole coordinate structure, and the contrast in
acceptability between (11a) and (11b) indicates that the RNR of this type behaves just
like typical rightward movement regarding sensitivity to locality conditions.

Given this state of affairs, Sabbagh gives the following generalization: “Movement
that violates the RRC [=Right Roof Constraint] is licit whenever it does not show overt
effects (viz-à-viz word order) of having violated the RRC.” (2007: 359) A similar
generalization holds true for P-stranding cases as well; cf. (8) vs. (10).
In order to explain this generalization, Sabbagh adopts the ATB-movement analysis of RNR and assumes that rightward movement is unbounded in principle. There are mainly two constraints that restrict the unboundedness of this movement. One is Fox and Pesetsky’s (2005) Order Preservation Condition, which Sabbagh informally states as follows:

\[(12) \quad \text{Order Preservation}\]

The linear ordering of syntactic units is affected by Merge and Move within a spell-out domain, but is fixed once and for all at the end of each spell-out domain.

Here, spell-out domains correspond to phases in the sense of Chomsky (2000) and Sabbagh assumes that among the spell-out domains are CP, vP, and PP. He further follows Chomsky (2000) in that when spell-out applies to these domains, it applies to the complement of the head of each domain. The important effect (12) gives rise to is that the linear orders fixed in a spell-out domain cannot contradict those fixed in the following spell-out domains. The other constraint is concerned with possible landing sites of rightward movement. Sabbagh stipulates the following:

\[(13) \quad \text{Landing sites for rightward movement}\]

Rightward movement may move an element X: (i) to the right edge of the first spell-out domain that contains X’s base position, or (ii) to the right edge of the matrix CP.\(^3\)

Given these constraints on rightward movement, it follows that in order to undergo an ATB rightward movement, the shared element needs to be located at the right edge of each conjunct; otherwise, the ATB movement necessarily violates the Order Preservation Condition. For illustration, let us consider the P-stranding cases of RNR given in (8) and (10), repeated below:

\[(14) \quad \begin{align*}
&\text{a. Joss walked suddenly into __, and Maria stormed quickly out of __, the dean’s office.}
&\text{b. *I sent one of the books to __ in perfect condition, and sent the other ones}
&\text{to __ in poor condition, the Somerville public library.}
\end{align*}\]

In (14a), the shared phrase the dean’s office sits at the right edge of each conjunct. Hence it is linearly ordered as following all the other elements in each spell-out domain, i.e. PP, vP, and CP, until the derivation reaches the final spell-out domain in which the two clauses are conjoined, as shown below:

\(^3\) Sabbagh considers only the cases that involve matrix coordination in RNR. He notes in his footnote 34 that (13), as it stands, gives rise to the problem of how to deal with those cases involving embedded coordination in RNR, and makes a suggestion to the effect that any Conj(unction) Phrase is a possible landing site for rightward movement. As far as we see, replacement of (ii) in (13) by this statement will make the whole condition on landing sites more accurate. See section 8.2.3 for relevant discussions.
At this final step, it undergoes rightward movement in an ATB fashion, adjoining to the matrix CP, which corresponds to the whole coordinate structure, and again it is linearly ordered as following all the other elements in this domain. Hence, no contradiction arises regarding the linear ordering of this shared phrase. In (14b), on the other hand, the shared phrase the Somerville public library does not sit at the right edge of each conjunct. Hence, in order to undergo ATB rightward movement without inducing contradiction with respect to linear orders, the shared phrase needs to be shifted to the right edge in each conjunct, as shown below:

(16) \[
\text{[CP I Past [vP send one of the books [PP to _] in perfect condition _]] and [CP (I) Past [vP send the other ones [PP to _] in poor condition _]] [the Somerville public library]}
\]

However, this is not allowed according to (13), since no landing site is available to it; the first spell-out domain for it is the PP immediately dominating it, but in order to move across the PPs in perfect condition and in poor condition, it should be adjoined to the right edge of the second spell-out domain, i.e. vP. Hence, there is no derivation for (14b) in which the linear ordering of the shared phrase does not contradict in any pair of the spell-out domains.

Much the same explanation holds for the fact that RNR is locality insensitive when the shared element is located at the right edge of each conjunct. Take (3) and (9), repeated below, for illustration:

(17) a. Josh said that he thought that I should sell __, and Jamie said that she thought that she might want to buy __, each of the Rambaldi artifacts that I have in my attic.

b. *Joss said [that he was going to donate __ to the library] yesterday, and Jamie claimed [that she would donate __ to the museum] last week, a large collection of ancient texts.
In (17a), the shared phrase \textit{each of the Rambaldi artifacts that I have in my attic} sits at the right edge of each conjunct. Hence, it is linearly ordered as following all the other elements in each spell-out domain, i.e. the embedded vPs, the embedded CPs, the matrix vP, and the matrix CP in each conjunct, until the derivation reaches the final spell-out domain in which the two clauses are conjoined, as shown below:

\begin{equation}
\text{CP \text{[CP Josh Past \text{vP say [CP that he Past \text{vP think [CP that I should \text{vP sell _]}]]]} and \text{[CP Jamie Past \text{vP say [CP that she Past\text{vP think [CP that she might \text{vP want to buy _]}]]]]] [each of the Rambaldi artifacts that I have in my attic]} }
\end{equation}

At this final step, it undergoes ATB rightward movement, adjoining to the whole coordinate structure, and again it is linearly ordered as following all the other elements in this domain. Hence, no contradiction arises regarding the linear ordering of this shared phrase. Notice that under this mechanism, the shared phrase does not have to move rightward in each conjunct before ATB movement applies as long as it sits at the right edge of each conjunct. This is the way to capture the fact that RNR is locality insensitive as long as the shared element sits in such a position. Thus, much the same account as given above applies to the island-violation cases given in (4)–(6).

In (17b), on the other hand, the shared phrase \textit{a large collection of ancient texts} does not sit at the right edge of each conjunct. Hence, in order to undergo ATB rightward movement without inducing contradiction with respect to linear orders, the shared phrase needs to be shifted to the right edge in each conjunct, as shown below:

\begin{equation}
\text{CP \text{[CP Josh Past \text{vP say [CP that he Past \text{vP be going to donate _ to the library]}]} and \text{[CP Jamie Past \text{vP claim [CP that she would \text{vP donate _ to the museum]}]} last week _] [a large collection of ancient texts]]}
\end{equation}

However, (13) prohibits it from moving across the modifiers \textit{yesterday} and \textit{last week} to be adjoined to the right edge of the matrix vP or CP in each conjunct, since the first
spell-out domain for it is the embedded vP (so that it is able to move across the PPs *to the library* and *to the museum*), and hence the matrix vP or CP in each conjunct provides no landing site for it. Hence, when the shared phrase moves across these modifiers when it undergoes ATB rightward movement, it violates the Order Preservation Condition: in the spell-out domain prior to the ATB movement, the shared phrase precedes these modifiers, and yet after it, the relevant linear orders are reversed.

Sabbagh (2007) claims that his mechanism of movement correctly captures the so-called Right Edge Restriction, which requires that the shared element of RNR must be located at the right edge of each conjunct before ATB rightward movement takes place. Notice that under his mechanism, this generalization is derived from the Order Preservation Condition given in (12); since the shared phrase ends up being located rightmost in the output of RNR, it needs to sit at the right edge of each conjunct to preserve its linear order relative to the other elements. This means that his RNR mechanism requires “intermediate” steps when the shared element undergoes ATB rightward movement. We will show in section 8.4 that our alternative mechanism does not require such “intermediate” steps and that this brings some advantages to our analysis.

8.2.2 Scope-out facts of shared QPs in RNR

Sabbagh (2007) supports the ATB movement analysis of RNR by presenting data that demonstrate that shared QPs in RNR are scoped out of coordinated sentences. He observes that (20a) allows the reading on which the shared QP *every patient who was admitted last night* takes scope over the indefinite *some nurse*, so that a different nurse may be involved in treating each patient. This is not the case with (20b).

(20)  

a. Some nurse gave a flu shot to __, and administered a blood test for __, *every patient who was admitted last night*. (every > some)

b. Some nurse gave a flu shot to *every patient*, and administered a blood test for *every patient*. (*every > some) (Sabbagh)

Under the ATB movement analysis of RNR, the wide scope reading of the shared QP can be straightforwardly accounted for, since it is moved out of the coordinate structure, hence occupies a position higher than the matrix subject, a CP-adjointed position according to Sabbagh. On the other hand, Sabbagh claims, neither the deletion analysis nor the multidominance analysis of RNR can provide any straightforward account for the difference between (20a) and (20b) with respect to the availability of the wide scope reading of the shared QP.

---

4 It is crucial under Sabbagh’s mechanism of movement that the matrix CP referred to by (13ii) as a landing site does not include *the matrix CP in each conjunct*, but only refers to the category corresponding to the whole sentence. See fn. 3 for a related point.
Crucially, Sabbagh further observes that this scope-out fact still holds even though the shared QP originates within an island, as shown below:

(21) a. John knows [someone who speaks __], and Bill knows [someone who wants to learn __], every Germanic language. (someone > every; every > someone)

b. A different doctor asked [who last used __], and a different nurse will find out [who sold __], every stethoscope in the ER. (a different doctor/nurse > every; every > a different doctor/nurse)

Since the shared QPs in (21) occupy the right edge of each conjunct, they can undergo ATB rightward movement and be adjoined to the matrix CPs without causing contradiction regarding the linear order statement. As expected, they can take scope over the indefinites that appear in the coordinated sentences. Thus, this fact lends strong support to the ATB movement analysis of RNR.\(^5\)

### 8.2.3 Problems of Sabbagh’s theory of movement

Though we find that Sabbagh’s (2007) argument for the ATB movement analysis of RNR on the basis of the scope facts outlined in the last subsection is convincing, his mechanism of movement in terms of Fox and Pesetsky’s (2005) Order Preservation Condition does not seem to be well designed. For one thing, the condition on landing sites on rightward movement given in (13), repeated below, which plays a crucial role in his mechanism of movement, looks like a sheer stipulation:

(22) *Landing sites for rightward movement*

Rightward movement may move an element X: (i) to the right edge of the first spell-out domain that contains X’s base position, or (ii) to the right edge of the matrix CP.

Even though he stresses that rightward movement is unbounded in itself, restricting the landing sites for rightward movement in such a way as stated in (22) in effect forces it to be bounded for normal cases of rightward movement, since no further landing site except the matrix CP is available once an element is moved to the right edge of the first spell-out domain. Let us consider (2) and (7), repeated below:

(23) *Max said [that he was going to return __ to the library] yesterday each of the books that he checked out last week.*

(24) *Jamie walked into __ suddenly, the dean’s office.*

\(^5\) See Sabbagh (2007) for more evidence to the same effect, which includes the possibility of Antecedent Contained Deletion and the availability of the “distributive” reading of the same/different phrases. Sabbagh also discusses cases relating to Condition C, negative polarity licensing, and bound variable anaphora that suggest that the bottom copy of the RNRred phrase is also visible for such a condition and licensing.
In (23), the right-shifted phrase *each of the books that he checked out last week* originates in the embedded object position. It is able to be shifted to the right edge of the embedded vP, the first spell-out domain, according to (22i), as shown below, and it is marked as linearly following all the other elements in this domain.

(25) Max said [that he was going to [vP return _ to the library [each of the books that he checked out last week]]] yesterday

Then, it must be adjoined to the matrix CP, as shown below:

(26) [CP [TP Max said [that he was going to [vP return_ to the library_]] yesterday]] [each of the books that he checked out last week]]

But in this spell-out domain, it has already had spell-out applied to its complement, i.e. TP, hence the modifier *yesterday* has been registered as linearly following all the other elements in this domain. This leads to a contradiction about the linear ordering statement after the phrase in question is right-shifted after *yesterday*. Notice that this account crucially exploits the unavailability of any landing site other than the first spell-out domain and the matrix CP for this right-shifted phrase; if it could be adjoined to the matrix vP, it should be moved across any modifier that is also hung to the same vP. In (24), the right-shifted phrase *the dean’s office* will also induce a contradiction about the linear order statement, since in the matrix CP, spell-out applies to the complement of the head C and the modifier *suddenly* is registered as linearly following all the other elements in this domain, as illustrated below:

(27) [CP [TP Jamie walked [PP into_] suddenly] [the dean’s office]]

Hence, the phrase *the dean’s office* cannot be right-shifted without violating the Order Preservation Condition. Again, if this phrase could be adjoined to the vP and *suddenly* could appear in the same vP, then the sentence should be grammatical. Sabbagh provides no motivation for restricting the landing sites for rightward movement as stated in (22) and, as far as we can see, there seems to be no reasonable conceptual motivation for this restriction. Hence, (22) seems to be nothing but an artifact that is made up exclusively for capturing the boundedness of rightward movement.
Empirically, it does not seem that Sabbagh can explain the fact that the ATB rightward movement involved in RNR shows locality effects even when a modifier that takes scope over the whole coordinate sentence intervenes in that movement, as shown in (11), repeated below:

(a) Joss will [sell __ to a library, and donate __ to a shelter] on the same day, all of his old manuscripts.

(b) *Joss [said that he was going to fire __, and insisted that no one would ever consider rehiring __] on the same day, the crazy guy from accounting.

According to Sabbagh’s mechanism for rightward movement, the grammaticality of these sentences crucially relies on where the modifier on the same day is attached. In (28a), the shared phrase all of his old manuscripts can be adjoined to the right edge of the vP in each conjunct, since the latter is the first spell-out domain, hence moving across the PPs to a library and to a shelter, as shown below:

\[
\begin{align*}
\text{Joss will } & \left[\text{vP sell } \_\text{ to a library [all of his old manuscripts]}\right] \text{ and } \left[\text{vP donate } \_\text{ to a shelter [all of his old manuscripts]}\right] \text{ on the same day} \\
\end{align*}
\]

Suppose that on the same day is attached within the whole TP. Then, in the matrix CP, spell-out applies to the complement of the C head, and hence this modifier is marked as linearly following all the other elements in this domain, just like yesterday in (23). This leads to a contradiction about the linear order statement when the shared phrase is shifted after on the same day, as shown below:

\[
\begin{align*}
\text{[CP TP Joss will [vP sell } & \_\text{ to a library [} \text{all of his old manuscripts} \text{]} \text{ and [vP donate } \_\text{ to a shelter [} \text{all of his old manuscripts} \text{]} \text{ on the} ] \\
\end{align*}
\]

Hence, in order to account for the grammaticality of sentence (28a), it is necessary to assume that this modifier can be attached higher than the whole TP, so that it belongs to the same spell-out domain as the right-shifted phrase. However, this incorrectly lets in even an ungrammatical sentence such as (28b), where the shared phrase is located at the right edge of each conjunct and moves across a modifier when it undergoes ATB rightward movement. Nothing should go wrong when the derivation
of (28b) reaches the stage where each conjunct is separately created and yet to be combined, since the shared phrase sits at the right edge of each conjunct and hence no contradiction occurs with respect to the linear order statement. If on the same day were attached higher than the whole TP, as is necessary to account for the grammaticality of (28a), then the relative order of this modifier and the right-shifted phrase would be determined in the very last spell-out domain, hence no contradiction about the linear order statement. Hence, as far as we can determine, Sabbagh’s mechanism for rightward movement fails to accommodate the contrast shown in (28).

A further empirical problem, though somewhat indirect, comes from Japanese Left-Node-Raising (LNR), about which Nakao (2009) demonstrates that it shows some properties typical of RNR. (31) below illustrates a typical instance of Japanese LNR:

(31) Zibun-no syasin-o [Taroo-va Mary-ni _ okuri],
    self -GEN picture-ACC -TOP -DAT send
    [Ziroo-va Susan -ni _ ageta].
    -TOP -DAT gave

‘Self’s picture [Taroo sent _ to Mary] and [Ziroo gave _ to Susan].’

Here zibun-no syasin “self’s picture” can have a so-called distributive reading, taking different values in the two conjuncts; that is, it can be interpreted as Taroo’s picture in the first conjunct and as Ziroo’s picture in the second. This is one of the typical properties of such a construction as RNR that involves an ATB dependency, as witnessed by the following English RNR:

(32) Every man i loves __, but no man j wants to marry __, his(i, j) mother.
    (Jacobson 1999)

Importantly, this Japanese construction shows locality insensitivity, just like RNR, when it does not show overt effects (vis-à-vis word order) of having violated locality conditions. Compare the following Japanese LNR examples:6

(33) Taroo-va zibun-no syasin-o [v_p [ _ suteyoo to suru] hito-o
    -TOP self -GEN picture-ACC throw away try person-ACC
    sikari], [v_p [ _ totte-okoo to suru] hito-o hometa].
    scold keep try person-ACC admired

‘Lit. Taroo, self’s pictures [scolded the person who tried to throw _ away] and [admired the person who tried to keep _].’

---

6 Here the intended reading of zibun-no syasin-o “self’s picture-ACC” is its distributive one. We make this as a rule throughout the chapter; thus, when a given sentence involves zibun-no __ “self's __” as the shared phrase of LNR, the intended reading is always its distributive one.
(34)  *Taroo-wa zibun-no syasin-o \[vp [Mary-ga _ suteyoo to sita] -TOP self -GEN picture-ACC -NOM throw away tried basyo-o sagasi], \[vp [Bill-ga _ totte-okoo to sita] basyo-o sagasita].

place-ACC look for -NOM keep tried place-ACC looked for

‘Lit. Taroo, self’s pictures [looked for the place where Mary tried to throw _ away] and [looked for the place where Bill tried to keep _].’

(33) is a case in which the shared phrase zibun-no syasin-o “self’s picture-ACC” sits at the left edge of each vP conjunct and the seemingly string-vacuous ATB movement is insensitive to the relative clause island. In (34), on the other hand, the original position of the shared phrase is not located at the left edge in each conjunct and hence the phrase in question involves non-string vacuous ATB movement. This gives rise to a relative clause island effect.

It is interesting in the present context to consider this Japanese construction, since it involves leftward movement and such movement is free from the clause-boundedness condition, unlike rightward movement in English, as illustrated below:

(35)  Zibun-no syasin-o John-wa [Bill-ga _ ki-ni itteiru] to itta.

self -GEN picture-ACC -TOP -NOM like Comp said

‘Self’s picture, John said that Bill liked t.’

Hence it is obvious that the landing sites available to leftward movement are different from those given in (22) for rightward movement. Given that leftward movement can take place unboundedly, it is natural to assume that every left edge of the categories a moved element passes through is available for leftward movement or if we follow the lines of Sabbagh’s (2007) idea of landing sites, it may be more natural to assume that every left edge of the spell-out domain is available for leftward movement.

Keeping this in mind, let us consider possible derivations for (33) and (34). As for (33), it is possible to provide much the same explanation as given for the relevant RNR cases. Since zibun-no syasin-o “self’s picture-ACC” sits at the left edge of each conjunct in all the spell-out domains until the two conjuncts are combined, it is always marked as linearly preceding all the other elements, hence no contradiction arises with respect to the linear order statement. Next, combining the two conjuncts and shifting zibun-no syasin-o to the left edge of the resulting coordinate structure takes place in the same spell-out domain, hence respecting the Order Preservation Condition.

Let us now consider (34). Recall that we mentioned in the last paragraph of section 8.2.1 that Sabbagh’s (2007) mechanism of rightward movement tries to capture the Right Edge Restriction, which requires that the shared element of RNR must be located at the right edge of each conjunct before ATB rightward movement takes place; otherwise, the Order Preservation Condition will be violated. It is then
expected under his mechanism that what we may call Left Edge Restriction should hold true for Japanese LNR. However, the ungrammaticality of (34) indicates that it is not the case that we can derive the legitimate LNR cases if the shared elements sit at the left edge of each conjunct before ATB movement takes place. This is because nothing seems to prevent the shared phrase *zibun-no syasin-o* from being scrambled to the left of the subjects *Mary-ga* and *Bill-ga* within the relative clause island of each conjunct before ATB movement takes place, as schematically shown below:

\[(36) \quad \ldots \left[ vP \left[ \text{Relative Clause} \ zibun-no syasin-o \right] \text{Mary-ga} \_ \text{suteyoo to sita} \text{ basyo-o} \ldots \right], \]

\[\left[ vP \left[ \text{Relative Clause} \ zibun-no syasin-o \right] \text{Bill-ga} \_ \text{totte-okoo to sita} \text{ basyo-o} \ldots \right] \]

Such an application of scrambling is in fact possible, as witnessed by the following example:

\[(37) \quad \text{Taroo-wa} \left[ \text{zibun-no syasin-o Mary-ga} \_ \text{suteyoo to sita} \text{ basyo-o} \right. \text{sagasita.}
\]

\[\text{looked for}
\]

‘Taroo looked for the place where Mary tried to throw away self’s picture.’

If this movement step were involved in the derivation of (34), as indicated in (36), the shared phrase *zibun-no syasin-o* would be located at the left edge of each conjunct before ATB movement takes place, and hence no contradiction would arise with respect to the linear order statement. Thus, in order to account for the ungrammaticality of (34), it is necessary to come up with a reasonable way of banning any “intermediate” step from being involved when ATB movement is applied.

### 8.3 Proposal

We follow Sabbagh (2007) in his claim that ATB rightward movement is involved in deriving RNR cases. We find that the data concerning the scope-out facts of shared QPs in RNR, discussed in section 8.2.2, constitute strong evidence for this approach. We depart, however, from his way of deriving the locality insensitivity of the ATB movement when a shared element sits at the right edge of each conjunct. We propose instead that when the ATB rightward movement does not show any overt effect with respect to word order, it must be “covert” rather than “overt” due to a PF bare output.
condition such as Chomsky’s (1995b) to the effect that overt movement must have a PF effect. We then derive the locality insensitivity of the ATB movement that does not induce a PF effect from the following assumption:

(38) The RRC and the island conditions apply only to “overt” movement.

We understand (38) in the context of current attempts to relate island phenomena to PF well-formedness conditions. In this context, phonetic gaps disrupt linearization conditions and island effects are ultimately related to this fact. As “covert” vacuous movement strings do not result in PF gaps, they do not induce island effects.

In this chapter, we will not try to derive the RRC and the island conditions from independent principles, just as Sabbagh does, leaving these conditions as they are. Though one might regard this as a regress when compared with Sabbagh’s attempt, we believe that the arguments given in the last section against his proposal suffice to show that his attempt to derive the locality conditions in question is not successful and hence that it is not unreasonable to leave this task to the side.

Chomsky (1995b) characterizes the relevant PF condition as follows:

(39) $a$ enters the numeration only if it has an effect on output. (Chomsky 1995b: 294)

Here $a$ is intended to be a strong feature, which triggers overt movement. Without committing ourselves to exactly what triggers overt movement, let us simply restate the essence in (39) as something like the following, under the single-cycle hypothesis according to which overt vs. covert movement is distinguished by which position of a chain is pronounced, the head or the tail:

7 A note on terminology: we use “overt” and “covert” as merely descriptive terms. “Overt” movement will be one in which the higher part of the chain is pronounced, “covert” where the lower copy is phonologically expressed. We take island effects as prohibiting gaps due to overt movement (deletion of lower copies) within islands. This follows the definitions for islands in Ross (1967) and Chomsky (1977). Both define islands as involving overt phonetic gaps.


9 A defining property of movement operations subject to Subjacency in Chomsky (1977) is that it leaves a phonetic gap. This is clearly consonant with the present proposal.

10 (40) may be independently supported by the fact that string-vacuous scrambling is prohibited in Japanese, as observed by Hoji (1985). Further, given the claim made by Saito (1989) that scrambling is semantically vacuous, it may follow that “covert” scrambling is also prohibited, since such movement does not have any effect on LF output.

This proposal is also clearly related to the one in Chomsky (1986) as regards parasitic gaps in subject relative clauses like (i):

(i) This is the man who everyone who meets $e$ admires $t$.

Chomsky proposed that in cases such as this, who is in situ, making the local CP available for the parasitic gap chain. Chomsky restricts this in situ possibility to cases where movement is string-vacuous.
The head of a chain created by Move cannot be pronounced unless it has an effect on PF output (i.e. the string linear properties of the output). Let us now consider the schematic structure of RNR in which a shared element originates at the right edge of each conjunct:

Here DP₃ is an occurrence of the shared DP that has undergone ATB rightward movement and DP₁ and DP₂ are occurrences in the sites from which it has undergone this movement. Thus, we have two chains whose head is shared: (DP₃, DP₁) and (DP₃, DP₂). The question is which occurrence is pronounced. Suppose, following the suggestion made by Sabbagh (2007, fn. 13), that when more than one chain shares a head, they are taken as identical for the purpose of Nunes’s (2004) mechanism of linearization, according to which the members of the same chain are taken as identical for linearization when Kayne’s (1994) Linear Correspondence Axiom (LCA) is applied to the structure containing the chain. Then, no more than one occurrence of the chains involved can be pronounced; otherwise, pronunciation of any two occurrences of the chains leads to contradiction regarding linearity. This automatically derives the fact that there is no LF (covert) ATB movement, as noted by Bošković and Franks (2000) and Citko (2000), since in such a case, application of covert ATB movement creates a configuration like (41) in which the tails of two chains sharing a head, namely DP₁ and DP₂, are pronounced, hence violating Nunes’s mechanism of linearization.

11 As the editors pointed out to us, a property of RNR in English is that the material on the right edge is intonationally separated from the rest of the conjoined material. Thus, it would appear that RNR always has an effect on PF output. We here interpret (40) as saying that it must have an effect on the string properties of the sentence. Two further points are worth making regarding this point. First, it does not appear that Japanese cases of LNR also require intonational breaks despite their other parallels to RNRs. Second, the intonational break may actually be the residue of something else entirely, e.g. the fact that shared material in a conjunction when not deleted must be deaccented in English. As the shared material in a RNR structure is the material on the right edge, it is deaccented. We speculate that this obligatory deaccenting yields the apparent effect of intonational separation. If this is correct, the intonational separation need not reflect structural properties of the RNRed material, such as being adjoined to the right edge. Moreover, it seems plausible to us that the deaccenting is not regulated by (40) but by some more overarching principle like the recoverability of deletion.

12 This idea might be more straightforwardly implemented on the assumption that ATB movement involves sideward movement; that is, in (41), DP₃ moves from the position occupied by DP₁ through the position of DP₂. This may be a more plausible alternative to the one assumed in the text, though we will not pursue this possibility any further in this chapter.
Keeping this in mind, let us now consider which occurrence is pronounced among DP₁, DP₂, and DP₃. Suppose, following the standard assumption that movement needs a trigger, that the ATB movement that is involved in RNR takes place to license some feature; let us mark it with [+F] without committing ourselves to the question of exactly what this feature is. We conjecture that such a movement to license [+F] is also involved in heavy NP shift and extraposition and that it functions to make an overt shift of a phrase to the right edge. Suppose then that a phrase bearing [+F] is licensed at an edge position and must be pronounced unless it violates any other condition such as (40). Given this, in the configuration given in (41), DP₃ bears [+F], as indicated below:

\[
\begin{array}{c}
[\cdots \text{DP}_1] \text{ and } [\cdots \text{DP}_2] \quad \text{DP}_3 \quad \text{[+F]}
\end{array}
\]

Then, in the chain <DP₃, DP₁>, DP₁ cannot be pronounced, since this chain is not string-vacuous, hence (40) is not applicable. Thus, DP₃ is eligible for pronunciation. In the chain <DP₃, DP₂>, however, DP₃ cannot be pronounced due to the ban on no PF effect, stated in (40). This leaves DP₂ as the only copy eligible for pronunciation, and it is pronounced.

Note, if DP₂ is pronounced in the configuration given in (42), the chain (DP₃, DP₁) has no phonetic realization. As a result, given (38), we predict that this chain as well as the chain (DP₃, DP₂) should not exhibit either RRC or island effects. Or put another way: as long as DP₂ is the copy pronounced, which follows if in string-vacuous movement chains, the lower copy is fed to PF, in the chain (DP₃, DP₂) DP₂ gets pronounced and in the chain (DP₃, DP₁) nothing gets pronounced. Thus the (DP₃, DP₂) chain is not subject to RRC or island effects as none of its members are

---

13 It may be that this feature is identified as some sort of focus feature, but there is no strong reason to suppose so. The editors provide us with an interesting example that may indicate that the right-shifted phrase in RNR is not really focused; the relevant example is given below:

(i) Only on Fridays does John cook and Mary eat fish.

Given the uniqueness condition on focus, this RNR example shows that it is on Fridays rather than the right-shifted phrase fish that is focused.

14 This case is consonant with those cases of multiple wh-fronting languages Bošković (2002b) deals with, in which the lower copy of a wh-chain is pronounced in order to avoid a sequence of homophonous wh-words.

One may derive the fact that DP₁ is not pronounced from Fox and Pesetsky’s (2005) Order Preservation Condition, following the lines taken by An (2007b). Given that each conjunct constitutes a spell-out domain, DP₁ and DP₂ are registered as following all the other elements before ATB movement applies. Then, on the assumption that DP₁ and DP₂ are regarded as just two occurrences of the same entity after ATB movement has applied, DP₁ cannot be pronounced since it induces a contradiction with respect to the linear order statement: DP₁ [=DP₁] will precede all the elements in the second conjunct. Such a contradiction does not arise when DP₂ is pronounced.
pronounced and the \((\text{DP}_3, \text{DP}_2)\) chain is similarly immune as only the lower copy is pronounced.

This reasoning clearly echoes Chomsky’s (1982) functional definition of empty categories. Whether the chain \((\text{DP}_3, \text{DP}_1)\) is taken as “overt” or “covert” depends upon the status of the other chain \((\text{DP}_3, \text{DP}_2)\) that it is related to via ATB movement. This situation only arises in contexts where chains share a common head, e.g. in coordination structures that license ATB movement.\(^{15}\) In these configurations, the partly overlapping chains only allow one occurrence to be pronounced, according to Nunes’s (2004) linearization mechanism. In this context, Nunes’s proposal has the interesting effect that in some chains none of the copies need be (or can be) pronounced and this in turn has the effect of liberating them from conditions that regulate “overt” movement. In what follows, we will see that this characterization is consistent with the data under consideration and further in section 8.4 that it makes predictions that contradict what the Right/Left Edge Restrictions would expect and yet are correct ones.

Let us illustrate this logic in an RNR example like (43) below.

(43) John wanted to buy __ and Mary had sold __, \textbf{one of the Rambaldi artifacts that I have in my attic}.

It has the structure in (44), where among the members of the ATB chains, the one that is bold-faced is pronounced and the ones that are marked with angled brackets are unpronounced:\(^{16}\)

\[
\text{[TP [TP [TP John wanted to buy <one of the Rambaldi artifacts that I have in my attic>] and [TP Mary had sold [one of the Rambaldi artifacts that I have in my attic]] <one of the Rambaldi artifacts that I have in my attic>]} \]

\(^{15}\) It should be part of any theory of RNR constructions that it explains why the properties we see only appear manifest in coordinate structures, and not more widely.

\(^{16}\) We will not be committed to the issue of what is the right structure for coordination, and will simply assume the traditional structure where coordination of XPs constitutes an XP.
In this structure, the shared phrase *one of the Rambaldi artifacts that I have in my attic* has undergone ATB movement and is right-adjointed to the whole coordinate structure. Since the occurrence of this phrase in the second conjunct is adjacent to the one shifted in an ATB fashion, the latter cannot be pronounced due to the PF condition given in (40), hence the former is the one that is pronounced.

Given the assumption in (38), it follows that in such a case as (44), the ATB rightward movement is immune to the RRC and the island conditions, since no “overt” movement is involved. That this is in fact the case was demonstrated in (17a) for a violation of the RRC and in (4)–(6) for violations of a variety of island conditions; all of the examples are repeated below:

(45)  Josh said that he thought that I should sell __, and Jamie said that she thought that she might want to buy __, each of the Rambaldi artifacts that I have in my attic.

(46)  I know someone who wants to buy __, and you know someone who wants to sell __, a copy of this manuscript.

(47)  Josh wonders who bought __, and Bill will find out who sold __, pictures of Fred.

(48)  Politicians win when they defend __, and lose when they attack __, the right of a woman to an abortion.

On the other hand, when a shared argument does not sit at the right edge of each conjunct, it undergoes “overt” ATB movement. Thus, such a RNR case as (49) below will have the structure given in (50):

(49)  John wanted to buy __ yesterday and Mary had sold __ the day before, one of the Rambaldi artifacts that I have in my attic.

(50)  \[
\begin{array}{l}
\text{[TP [TP [TP John wanted to buy <one of the Rambaldi artifacts that I have in my attic> yesterday] and [TP Mary had sold <one of the Rambaldi artifacts that I have in my attic> the day before]] [one of the Rambaldi artifacts that I have in my attic]]}
\end{array}
\]

Here, the right-shifted occurrence of the shared phrase *one of the Rambaldi artifacts that I have in my attic* is pronounced, since in this case, the ATB rightward
movement is not string-vacuous, and so the PF condition (40) is not applicable. It is then predicted that in such a case, the ATB movement is sensitive to the locality conditions in question. That this is the case is shown in (17b), repeated below, which involves non-string-vacuous ATB movement, violating the RRC:

(51) *Joss said [that he was going to donate __ to the library] yesterday, and Jamie claimed [that she would donate __ to the museum] last week, a large collection of ancient texts.

Furthermore, the present mechanism of ATB movement can accommodate the P-stranding cases of RNR in an appropriate way. The relevant examples in (14) are repeated below:

(52) Joss walked suddenly into __, and Maria stormed quickly out of __, the dean’s office.

(53) *I sent one of the books to __ in perfect condition, and sent the other ones to __ in poor condition, the Somerville public library.

Notice that the possibility of P-stranding has to do with overt movement and further that it also depends upon in which direction the movement in question takes place, as witnessed by the fact that English allows P-stranding for leftward movement, but not for rightward movement. Given this, the grammaticality of (52) is attributed, under the present theory, to the fact that the shared phrase the dean’s office sits at the right edge of each conjunct before ATB rightward movement takes place, hence no “overt” ATB movement is applied, as shown below:

(54) \[ \text{TP} \left[ \text{TP} \left[ \text{TP} \text{Joss walked suddenly into <the dean's office>}, \text{and} \left[ \text{TP} \text{Maria stormed quickly out of } <\text{the dean's office}> \right] \right] \right] \] \text{<the dean's office>}

On the other hand, (53) involves a case of RNR in which the shared phrase does not sit at the right edge of each conjunct before the ATB rightward movement takes place, hence the latter movement must be “overt.” The ungrammaticality of (53) is then attributed to whatever condition rules out such a case as below where overt rightward movement leaves a preposition stranded:

(55) *I sent one of the books to __ in perfect condition, the Somerville public library.
In the last section, we pointed out that Sabbagh (2007) cannot explain the fact that the ATB rightward movement involved in RNR shows locality effects when a modifier that takes scope over the whole coordinate sentence intervenes in that movement, as shown in (28), repeated below:

(56)  a. Joss will sell __ to a library, and donate __ to a shelter on the same day, all of his old manuscripts.

         b. *Joss [said that he was going to fire __, and insisted that no one would ever consider rehiring __] on the same day, the crazy guy from accounting.

Under the present mechanism of movement, this fact is straightforwardly explained: since such cases involve “overt” ATB movement due to the presence of the modifier on the same day, they should show locality effects just in the same way as normal rightward movement does. Notice that this fact holds true, no matter whether the shared phrase sits at the right edge of each conjunct before ATB movement takes place, as witnessed by (56b).

Further, the present ATB movement approach to RNR can correctly capture the scope-out facts of the shared QPs as shown in section 8.2.2; the relevant example (20a) is repeated below:

(57) Some nurse gave a flu shot to __, and administered a blood test for __, every patient who was admitted last night. (every > some)

The fact that this sentence allows the reading on which every patient takes scope over some nurse straightforwardly follows under the present analysis of RNR, since even though the shared QP is originated at the right edge of each conjunct, it undergoes “covert” ATB movement, so that (57) has the following structure:

(58) [TP [TP some nurse [vP gave a flu shot to <every patient who was admitted last night>] and [vP administered a blood test for [every patient who was admitted last night]]] <every patient who was admitted last night>]

Given that the shared QP is right-adjoined to a category no lower than TP, as indicated in (58), it occupies a position higher than that of some nurse, which thus
correctly captures the fact that the former takes scope over the latter. The same explanation carries over to such cases as given in (21) where the shared QPs are originated within islands.

As may be clear from the above discussion, our approach to RNR can be regarded as a refinement of Sabbagh’s (2007) ATB movement approach in such a way that “covert” as well as “overt” movement takes care of the scope-out facts of the shared phrases, whereas the locality insensitivity of string-vacuous ATB movement is attributed to the fact that the movement in question is “covert”, hence immune to the locality conditions that are assumed to apply only to “overt” movement.

8.3.1 Abels’s (2004) objection
Abels (2004) observes that leftward ATB movement and VP ellipsis are compatible with each other, as shown below:

(59) ?Who did you say that John had invited long ago but that Mary hadn’t [VP ___] until yesterday?

In this sentence, who undergoes leftward ATB movement and the VP invited <who> is erased in the second conjunct. Abels then argues that the fact that RNR does not allow VP deletion, as shown below, constitutes counterevidence to the ATB rightward movement approach to RNR:

(60) *Jane talked about and/but Frank didn’t [VP ___], the achievements of the syntax students.

However, notice that this particular instance involves string-vacuous movement, and hence under the present analysis of RNR, the shared argument the achievements of the syntax students is pronounced inside the VP of the second conjunct, thereby not able to survive VP deletion. Thus, the present analysis correctly blocks such a VP deletion case as (60) from being generated. Further, it leads to the prediction that if the ATB movement involved is not string-vacuous, the application of VP deletion should result in a grammatical sentence. This is in fact borne out:

(61) Jane discussed yesterday and Fred did/but Fred didn’t ___ the day before, the achievements of the syntax students.

Thus, even though (60) appears to constitute counterevidence to the ATB movement approach to RNR, when contrasted with such data as (61), it is turned into strong support to our revised version of the ATB movement approach to RNR.

8.3.2 PF output condition on string-vacuous movement
Recall that we have made crucial use of the PF output condition given in (40), repeated below, to ban string-vacuous movement:
The head of a chain created by Move cannot be pronounced unless it has an effect on PF output.

Notice that this condition is an output condition rather than a derivational condition that constrains an application of Move. This characterization makes a prediction different from Sabbagh’s (2007) with respect to the interaction of string-vacuous movement and locality effects, when a given instance of movement involves more than one step. Sabbagh would predict that an application of string-vacuous movement always evades locality effects, since such an application would preserve the linear order statement that has been created up to the current spell-out domain. On the other hand, under our theory of movement, whether or not a given instance of movement is string-vacuous should be measured in terms of the whole chain rather than that step of movement. This means that whether a given step of movement is taken as “overt” or “covert” is decided in terms of the PF effect of the entire chain that includes it, and hence whether a given step of movement violates the RRC or an island condition is only decided when the whole chain is evaluated. We can technically code the relevant condition as follows: if a given movement step violates a locality condition (e.g. the RRC or island condition), mark the moved phrase with “*”. At PF evaluate the chain as in (63).

(63) a. “*” is removed when the bottom copy of the chain of its carrier is pronounced.17
   b. If the chain retains “*”, it leads to ungrammaticality.

Let us apply this mechanism to an ATB configuration:

(64) [... DP₁] and [... DP₂] DP₃ [+F]

If the chain (DP₃, DP₁) or (DP₃, DP₂) violates any locality condition, DP₃ is marked with “*”. Further, if there is no PF effect in the chain (DP₃, DP₂), DP₂ is the copy pronounced and DP₃ is deleted at PF. This eliminates the “*” according to (63a) and the structure is well formed. In effect, we have here one more “Lasnik Effect”: an instance of a PF deletion process saving an otherwise illicit derivation (cf. Lasnik 2001).

Now, a prediction: when a chain created by a series of steps of movement includes a string-vacuous step of movement, it should nonetheless show locality effects if the

---

17 Strictly, the ATB-moved phrase is involved in two chains, namely (DP₃, DP₁) and (DP₃, DP₂) in (64). The chain relevant for (63a) is intended to be (DP₃, DP₂).
whole chain is created in such a way that the order of the words involved is changed. This is in fact borne out when we consider Japanese LNR and its clefted counterpart. (65) is a LNR example and (66) is its clefted counterpart:\footnote{NL in the glosses stands for nominalizer here and in what follows.}

\begin{center}
\begin{tabular}{llll}
\hline
(65) & pro & \textit{zibun-no syasin-o} & \textit{[\_ suteyoo to suru] hito-o sikari},
\hline
& \hspace{1cm} & \textit{self} \hspace{0.5cm} & \textit{picture-ACC} \textit{throw away try person-ACC scold}
\hline
& \hspace{1cm} & \textit{[\_ totte-okoo to suru] hito-o hometa}.
& keep \textit{try person-ACC admired}
\hline
& \hspace{1cm} & \textit{‘Lit. pro, self’s pictures [scolded the person who tried to throw \_ away] and [admired the person who tried to keep \_]’}.
\end{tabular}
\end{center}

\begin{center}
\begin{tabular}{llll}
\hline
(66) & *pro & \textit{[\_ suteyoo to suru] hito-o sikari}, & \textit{[\_ totte-okoo to suru]}
\hline
& \hspace{1cm} & \textit{throw away try person-ACC scold keep try}
\hline
& \hspace{1cm} & \textit{hito-o hometa} & \textit{no-wa zibun-no syasin-o da.}
\hline
& \hspace{1cm} & \textit{person-ACC admired NL-TOP self \hspace{0.5cm} picture-ACC be}
\hline
& \hspace{1cm} & \textit{‘It was self’s pictures that pro [scolded the person who tried to throw \_ away] and [admired the person who tried to keep \_]’}.
\end{tabular}
\end{center}

The grammaticality of (65) immediately follows if we adopt the same ATB movement approach to Japanese LNR as we did for English RNR. In this case, the shared phrase \textit{zibun-no syasin-o “self’s picture-ACC”} undergoes string-vacuous movement with respect to the first conjunct, and hence is pronounced in its original position within that conjunct, as schematically shown with English glosses below:

\begin{center}
\begin{tabular}{llll}
\hline
(67) & pro \textit{<self’s picture-ACC>} & \textit{[[self’s picture-ACC…] \ldots]}, & \textit{[[<self’s picture-ACC>…] \ldots]}
\hline
\end{tabular}
\end{center}

\textit{Zibun-no syasin-o “self’s picture-ACC”} is marked with “*” when it undergoes ATB leftward movement, crossing the relative clause island. However, by (65a), “*” is deleted since the bottom copy of the chain in the first conjunct is the one pronounced.

(66) is derived from this sentence by clefting the shared phrase \textit{zibun-no syasin-o “self’s picture-ACC”}. This further movement renders the derivation ungrammatical. Why? According to Hasegawa (1997), the Japanese cleft is derived by first moving a focused phrase out of a given clause and then moving the remnant clause to the topic phrase. The clause in the topic phrase functions as the presupposition. (68) illustrates the derivation.
In (68a), *self’s picture* sits at the left edge in each conjunct, hence its ATB leftward movement should be free from island conditions according to Sabbagh’s (2007) mechanism of movement in terms of the Order Preservation Condition. The derivation in (68) should be grammatical, contrary to fact.

On the present account, the ATB leftward movement of *self’s picture* at stage (68a) is marked with “*” due to a relative clause island violation. Further application of remnant movement from stage (68b) results in the configuration in which the chain of *self’s picture* created by the ATB movement has a PF effect. Hence, the higher copy in what amounts to the focused position in the cleft construction is retained. Since the ATB movement in question ends up affecting PF output (i.e. it is “overt”), the “*” marked on *self’s picture* remains, and the derivation in (68) is illicit.

19 That moving across the matrix subject pro still behaves like string-vacuous movement is confirmed by the contrast in grammaticality between (34) and the following LNR, in which the subject of the relative clause is replaced by pro:

(i) Taroo-wa zibun-no syasin-o [[pro _ suteyoo to sita] basyo-o sagasi],
-TOP self - GEN picture-ACC throw away tried place-ACC look for
[[pro _ totte-okoo to sita] basyo-o sagasita],
keep tried place-ACC looked for
‘Lit. Taroo, self’s pictures [looked for the place where pro tried to throw _ away and
[looked for the place where pro tried to keep _].’
8.3.3 Absence of the scope-out facts in Japanese LNR

It is well known that QPs in Japanese, in contrast to English, show scope rigidity as structurally higher surface QPs take wider scope. Thus, while the English (69) is scopally ambiguous, its Japanese counterpart, (70), must be interpreted with the subject QP taking scope over the object QP.

(69) Someone loves everyone.

(70) Dareka-ga daremo-o aisiteiru.
    someone-NOM everyone-ACC love

Since May (1977), the wide scope of the object QP in (69) is derived by applying covert QR and deriving (71).

(71) [TP <everyone> [TP someone loves everyone]]

Given this, the rigidity of Japanese QPs observed above can be captured by assuming that while different copies can be retained at AP and CI interfaces in English, copy deletion applies uniformly at the interfaces in Japanese. If so, we expect Japanese LNR and English RNR to display different scoping-out effects. This expectation is borne out. Let us first note that when non-string-vacuous movement is involved, the shared QP in LNR can scope over an indefinite in each conjunct:

(72) Subete-no syasin-o dareka otokonoko-ga _ kirai, dareka onnanoko-ga _ kiniitteiru.
    every -GEN picture-ACC some boy -NOM hate some girl -NOM like

'Every picture, some boy hates and some girl likes.'

This sentence allows the reading on which for each picture, a different pair of boy and girl is involved as an experiencer of the verbs expressed. However, when the shared QP in LNR sits at the left edge of an island in each conjunct, as in (73), it cannot scope over an indefinite that is outside the island. This contrasts with English RNR, e.g. (21) above.

(73) Taroo-wa subete-no syasin-o [[_ suteyoo to suru] hito-ni nanika -TOP every-GEN picture-ACC throw away try person-DAT some batu-o atae], [[_ totte-okoo to suru] hito-ni nanika punishment-ACC give keep try person-DAT some hoobi-ageta].
    oreward-ACC gave

'Lit. Taroo, every picture [gave some punishment to a person who tried to throw _ away] and [gave a reward to a person who tried to keep _].'
Observe: (73) is grammatical since the shared QP subete-no syasin-o “every picture-ACC”, though originating within the relative clause island, sits at the left edge of each conjunct and the ATB movement involved is string-vacuous in the first conjunct, hence must be “covert” with the bottom copy pronounced. However, (73) does not allow the reading on which this shared QP takes scope over the indefinites nanika batu-o “a punishment-ACC” and nanika hoobi-o “a reward-ACC”, which are located outside the relative clause island. This follows if the PF copy and the LF copy must be the same in Japanese.20

8.4 Against the Right/Left Edge Restriction

In this section, we argue against the empirical accuracy of the Right Edge Restriction, which requires that the shared element of RNR must be located at the right edge of each conjunct for ATB rightward movement to be licit. Recall that the analysis in Sabbagh (2007) rests on the assumption that this generalization is correct and thus needs to be derived from independent principles. Sabbagh (2007) derives it from the Order Preservation Condition; unless the shared element sits at the right edge of each conjunct before the ATB movement is applied, this condition will be violated. On the other hand, our mechanism of ATB movement does not necessarily require it to have “intermediate” steps to observe the Right Edge Restriction. Let us take the following example for illustration:

(74) John will donate __ to the library, and Maria will donate __ to the museum, each of these old novels. (Sabbagh 2007: 354)

According to Sabbagh’s mechanism of rightward movement, the shared phrase each of these old novels must be right-adjoined to vP in each conjunct before ATB

---

20 Even in English RNR, there is a case in which the PF copy must be the same as the LF copy. The editors provide an example such as the following:

(i) Mary didn’t sell __ and John wouldn’t buy __, many books about linguistics.

It is very hard to get the reading on which the RNRRed quantifier takes scope over negation with this sentence. This suggests that negative scope is determined in terms of relative position among PF copies, though we have no idea why this must be so. Under the present assumptions, it will be predicted that if a RNRRed quantifier is produced by non-string-vacuous movement, unlike (i), then it is easier to obtain the wide scope reading of that quantifier since in such a case, the top copy of the quantifier chain is pronounced. That this is in fact the case can be witnessed by the fact that the wide scope reading of the many-phrase in (ii) is about as available as the corresponding case, given in (iii), in which the many-phrase undergoes heavy NP shift:

(ii) Mary didn’t sell __ to her students and John wouldn’t buy __ for his colleagues, many books about linguistics.

(iii) Mary didn’t sell __ to her students many books about linguistics.
movement takes place, since otherwise, the linear order statement created in this
spell-out domain will contradict that created after ATB rightward movement
has applied. On the other hand, our mechanism of movement does not
require such an intermediate step before ATB rightward movement takes
place, since there is no intervening boundary relevant to the locality conditions
applying to the ATB movement in question. Hence the latter movement can
apply when the shared argument each of these old novels is located in its
original position, i.e. the object position of donate, in each conjunct, and must do
so if an economy condition prohibits any superfluous steps. Thus, if our proposal
is on the right track, it casts doubt on the empirical accuracy of the Right
Edge Restriction. In the next subsection, we argue that when the restriction
in question is generalized so as to accommodate LNR cases, it becomes too strong,
thus indicating that both Left and Right Edge Restrictions are incorrect. Finally, we
address some apparent problems concerning the data that Sabbagh (2007) claims
show the need for intermediate steps when ATB rightward movement is applied in
English RNR.

8.4.1 Japanese LNR

We demonstrated in section 8.2.3 that Sabbagh’s (2007) RNR mechanisms do not
extend in a natural way to accommodate Japanese LNR cases, though these also
exhibit locality insensitivity when the shared elements involve string-vacuous move-
ment, as shown in the contrast between (33) and (34), both of which are repeated
below:

(75) Taroo-wa zibun-no syasin-o [[_ suteyoo to suru] hito-o sikari],
    -TOP self -GEN picture-ACC throw away try person-ACC scold
    [_[_ totte-okoo to suru] hito-o hometa].
    keep try person-ACC admired
    ’Lit. Taroo, self’s pictures [scolded the person who tried to throw _ away]
    and [admired the person who tried to keep _].’

(76) *Taroo-wa zibun-no syasin-o [[Mary-ga _ suteyoo to sita] basyo-o sagasi],
    -TOP self -GEN picture-ACC -NOM throw away tried place-ACC look for
    [_[Bill-ga _ totte-okoo to sita] basyo-o sagasita].
    -NOM keep tried place-ACC looked for
    ’Lit. Taroo, self’s pictures [looked for the place where Mary tried to
    throw _ away] and [looked for the place where Bill tried to keep _]’.

Sabbagh (2007) can properly account for the grammaticality of (75), since the shared
phrase zibun-no syasin-o “self’s picture-ACC” though included within the relative
clause island, is located at the left edge of each conjunct, and hence its order relative to other elements is not changed after ATB leftward movement takes place. By contrast, the ungrammaticality of (76) causes a problem to Sabbagh’s mechanism of movement. Recall that under his mechanism, nothing seems to prevent the shared phrase *zibun-no syasin-o* from being scrambled to the left of the subjects *Mary-ga* and *Bill-ga* within the relative clause island of each conjunct before ATB movement takes place. In fact, this step of movement will be necessary to derive such a LNR case as the following without inducing a violation of the Order Preservation Condition:

(77) *Zibun-no syasin-o [Mary-ga kinoo _ suteyoo to si], [Bill-ga kyoo self -GEN picture-ACC -NOM yesterday throw away try -NOM today _ totte-oku koto-ni sita]. keep fact-DAT did

’Self’s pictures, Mary tried to throw _ away yesterday and Bill decided to keep _ today.’

In order to apply ATB leftward movement to *zibun-no syasin-o* “self’s picture-ACC” without violating this condition, it should shift to the top of each conjunct. This in turn suggests that in (76), the shared phrase needs to be shifted in the same way. However, if this happens, no contradiction in the linear order statement arises, exactly like (77). This indicates that a Left Edge Restriction parallel to the Right Edge Restriction in English does not hold for Japanese LNR. Indeed, it seems necessary to ban any “intermediate” step from being involved when ATB movement is applied.

There is a reasonable way to achieve this result under our mechanism of movement. Notice that this mechanism requires no intermediate step to derive such a LNR case as (77); since no boundary relevant for the locality conditions exist in the derivation of this sentence, the shared phrase *zibun-no syasin-o* “self’s picture-ACC” can undergo ATB movement in one fell swoop. Suppose that an economy condition prohibits any superfluous step in the derivation. Then, one fell swoop ATB movement is mandatory in deriving (77). Going back to (76), we can reasonably claim under our mechanism of movement that the intermediate step of movement of the shared phrase *zibun-no syasin-o* “self’s picture-ACC” to the top of the relative clause in each conjunct is superfluous, hence ATB leftward movement must take place in one fell swoop. 21 Since this movement crosses the relative

21 One might object that since Japanese has access to scrambling, an operation usually claimed to be free from any economy condition, nothing will prevent this operation from being involved in the intermediate step of movement in question. See Abe (1993), however, for the claim that scrambling is subject to such an economy condition as the fewest steps condition, hence cannot be used as an intermediate step of another application of movement.
clause island, the resulting chain is marked with “*” according to the mechanism mentioned around (63). Further, since it is not string-vacuous movement, the head copy of the chain is pronounced, hence the “*” is retained and this leads to ungrammaticality.

Further evidence against the Left Edge Restriction and for our mechanism of movement comes from the fact that LNR shows asymmetrical sensitivity to localities between the two conjuncts involved. Notice that under our mechanism, whether a given instance of ATB leftward movement is taken as “overt” or “covert” is determined in terms of the chain created between the shared element and its copy in the first conjunct, and that the chain created between the shared element and its copy in the second conjunct is irrelevant for this decision. Thus, it is predicted under this system that a given instance of ATB leftward movement is locality insensitive, as long as the chain involving the first conjunct is created by string-vacuous movement and, crucially, irrespective of the status of the other chain with respect to string-vacuity and locality. Recall that once it is decided to pronounce the bottom copy of the chain involving the first conjunct, the other chain has no phonetic realization, hence should be immune to locality effects. Let us then consider the following schematic structures:

\[(78)\]

\begin{enumerate}
\item \( \text{DP}_1 [[\text{Island} \; \text{DP}_2 \ldots] \ldots] \text{ and }[\ldots[[\text{Island} \; \text{XP} \ldots \text{DP}_3 \ldots] \ldots] \]
\item \( \text{*DP}_1 [[\ldots\text{Island} \; \text{XP} \ldots \text{DP}_2 \ldots] \ldots] \text{ and }[\ldots[[\text{Island} \; \text{DP}_3\ldots] \ldots] \]
\end{enumerate}

In these structures, \( \text{DP}_1 \) is the head of the chains created by ATB leftward movement, and \( \text{DP}_2 \) and \( \text{DP}_3 \) are its bottom copies. (78a) represents a configuration where \( \text{DP}_2 \) is adjacent to \( \text{DP}_1 \) but \( \text{DP}_3 \) is not located at the left edge of the second conjunct. (78b), by contrast, represents a configuration where the opposite relationships hold in the relevant sense: while \( \text{DP}_3 \) is located at the left edge of the second conjunct, \( \text{DP}_2 \) is not located in a parallel way in the first conjunct. As indicated in (78), our mechanism of movement predicts that (78a) should be grammatical while (78b) should not. The relevant data seem to verify this prediction:
a. ?Taroo-wa zibun-no syasin-o [[_ suteyoo to sita] hito-o sagasi],
   -TOP self -GEN picture-ACC throw away tried person-ACC look for
   [[[Bill-ga _ totte-okoo to sita] basyo-o sagasita].
   -NOM keep tried place-ACC looked for
   ‘Lit. Taroo, self’s pictures [looked for the person who tried to throw _ away
   and [looked for the place where Bill tried to keep _].’

b. *Taroo-wa zibun-no syasin-o [[Mary-ga _ suteyoo to sita] basyo-o
   -TOP self -GEN picture-ACC -NOM throw away tried place-ACC
   sagasi], [[_ totte-okoo to sita] hito-o sagasita].
   look for keep tried person-ACC looked for
   ‘Lit. Taroo, self’s pictures [looked for the place where Mary tried to
   throw _ away] and [looked for the person who tried to keep _].’

(79a) illustrates the case given in (78a) and (79b) that given in (78b), and there is a
clear contrast in acceptability between them.22 This follows under our assumptions.
It also constitutes evidence against a Left Edge Restriction and whatever mechanism,
such as Sabbagh’s (2007), tries to derive this restriction. Lastly, this casts doubt on the
validity of the Right Edge Restriction as applied to English RNR cases.

In fact, there is a piece of evidence for this final conclusion when we consider the
possibility of P-stranding in English RNR. Consider the following examples:

(80) a. ?Joss walked into __ suddenly, and Maria stormed quickly out of __, the
dean’s office.

b. *Joss walked suddenly into __, and Maria stormed out of __ quickly, the
dean’s office.

Our proposal predicts the contrast in (80). In (80a), the chain involving the second
conjunct is formed by string-vacuous movement, thereby its bottom copy being
pronounced, and hence no P-stranding takes place. And this is basically not affected
by the fact that the shared argument the dean’s office is not located at the right edge of
the first conjunct. In (80b), on the other hand, the chain involving the second
conjunct is formed by non-string-vacuous movement, hence the head copy being
pronounced and leaving the preposition of behind. This violates whatever constraint
prohibits P-stranding by rightward movement and it is not affected by the fact that
the shared argument was located at the right edge of the first conjunct. Thus, the
contrast in (80) constitutes strong evidence against the Right Edge Restriction.23

22 Though the former sentence does not sound so good, to begin with. This might suggest that some
superficial parallelism requirement contributes to the acceptability of coordinate structures involving ATB
movement. In fact, (79b) sounds worse than (76).
23 We are unable to construct relevant RNR examples regarding locality effects that show a clear
contrast between (79a) and (79b) cases, probably because of some parallelism requirement imposed
upon ATB structures, mentioned in the previous footnote. In this context, it is interesting to note that
8.4.2 Some apparent problems

In footnote 7, Sabbagh (2007) gives evidence for the intermediate step of movement occurring in each conjunct in those cases where rightward ATB movement takes place in a non-string-vacuous way. This comports with the Right Edge Restriction, while it is at odds with the present proposal. Sabbagh (2007) provides two pieces of evidence. One is concerned with the following paradigm of English RNR and its interaction with wh-movement:

(81) ??Who did John send __ to Mary [a picture of ___]?
(82) Who did Josh buy __ and Mary sell __, [a picture of ___]?
(83) ??Who did Josh buy __ from Sam and Mary sell __ to Max, [a picture of ___]?

When non-string vacuous ATB movement takes place, as in (83), a freezing effect on wh-movement occurs as in (81). This effect disappears in cases like (82) that involve string-vacuous ATB movement. Sabbagh (2007) takes this as indicating that the derivation of (83) involves rightward movement of the shared phrase a picture of who to the right edge of each conjunct before ATB movement takes place and hence that the shared phrase constitutes a barrier, exactly like (81). This gives rise to a freezing effect.

The present proposal also derives the contrast. The difference between (82) and (83) can be attributed to the fact that only in (83) is the shared highest DP copy pronounced. The relevant generalization will then be that locality effects are products of “overt” syntactic representations and thus it matters which copy is pronounced. Whether a structure is island-like depends on its pronounced position. We can code this as in (84).24

24 This looks more cumbersome than it is. What is at issue is how islands are determined. The “*” simply records which elements are relevant for computing barriers and relevant copies. Note, it is intrinsic to a Lasnikian approach to islands that they are output conditions on pronounced representations. The “*”s code the relevant copies.
a. “*” is removed when either (i) or (ii) holds:
   (i) the bottom copy of the chain is pronounced.
   (ii) the boundary that has caused “*”-marking is not pronounced.

b. If the chain retains “*”, it leads to ungrammaticality.

Then, in (82), the ATB shifted DP does not constitute a barrier for the extraction of who since it is not pronounced, while in (83), it does constitute a barrier since it is pronounced.

A similar account can be given to the second piece of evidence Sabbagh provides for the Right Edge Restriction. This is concerned with the heaviness of shared arguments in RNR:

(85) Josh returned __ for Jamie several books *(that she borrowed last month).

(86) Josh gave Jamie __, and Sue gave Sandy __, several books (that she borrowed last month).

(87) *Josh will sell __ to the library today, and Maria will donate __ to the museum tomorrow, several books.

The heaviness requirement only holds for those RNR cases that involve non-string-vacuous ATB movement, as illustrated in the contrast between (86) and (87). Again, Sabbagh (2007) takes this fact as indicating that (87) involves an intermediate step of movement of the shared DP to the right edge of each conjunct before ATB rightward movement takes place. The present proposal accounts for this pattern on the natural assumption that “heavyness” is a PF restriction on adjoined constituents and that in contrast to (85) and (87) the pronounced copy in (86) is not the one in adjunct position.

8.5 Speculations on some further consequences

8.5.1 Argument vs. adjunct asymmetry

Recall that we have made the following assumption above:

(88) The RRC and the island conditions apply only to “overt” movement.

Notice that those RNR and LNR cases we have so far dealt with all involve DP arguments for their shared phrases. Given Huang’s (1982) generalization, according to which arguments show locality effects only when they undergo overt movement, whereas adjuncts show such effects when they undergo covert movement as well, one may wonder in light of our mechanism of movement whether this generalization also holds true for RNR/LNR cases. This conjecture is supported by both English RNR and Japanese LNR. Let us first consider relevant English RNR cases:
(89)  
  a. John got fired _ and Bill had his salary reduced __, because he talked back/because of office politics.

  b. *Mary helped the person who got fired _ and comforted the person who had his salary reduced __, because he talked back/because of office politics.

(89a) shows that adjunct phrases/clauses such as because _ can serve as shared phrases for RNR. The unacceptability of (89b) indicates that unlike shared arguments, shared adjuncts cannot evade island effects even though string-vacuous ATB movement is involved; in other words, “covert” movement of adjuncts in RNR is sensitive to such locality conditions. The same pattern of facts also obtains for Japanese LNR, as illustrated below:

(90) Zibun-no titoya-no sei-de [John-wa _ kubi-ni nari], [Mary-wa self -GEN father-GEN fault-for -TOP be-fired -TOP _ genkyuu-sareta].

  was-salary-reduced

  ’For self’s father’s fault, John got fired _ and Mary had her salary reduced _.’

(91) *John-wa zibun-no titoya-no sei-de [ _ kubi-ni natta] hito-o tasuke, -TOP self -GEN father-GEN fault-for was-fired person-ACC help

  [ _ genkyuu-sareta] hito-o nagusameta.

  was-salary-reduced person-ACC comforted

  ’Lit. John, for self’s father’s fault [helped the person who got fired _] and [comforted the person who had his salary reduced _].’

The acceptability of (90) shows that Japanese LNR allows adjuncts to serve as shared phrases that undergo ATB movement. Unlike those LNR cases that involve shared arguments, (91) does not have the intended reading, even though here string-vacuous ATB movement of zibun-no titoya-no sei-de “for self’s father’s fault” is involved. This again suggests that adjuncts are sensitive to island effects even when they undergo “covert” movement.

If our mechanism of movement for RNR/LNR is correct, the above data clearly support Huang’s generalization, and (88) should be amended in the following way:

(92) The island conditions apply only to “overt” movement for arguments whereas they apply not only to “overt” but also to “covert” movement for adjuncts.

Given our formulation of “*”-marking, given in (84), this suggests that the condition concerning removal of “*”, stated in (84a), does not apply to adjuncts; hence, we need to modify (84) into something like the following:
(93) a. “*” is removed when either (i) or (ii) holds:
   (i) the bottom copy of the chain is pronounced.
   (ii) the boundary that has caused “*”-marking is not pronounced.

b. If the chain retains “*”, it leads to ungrammaticality.

c. (a) does not apply to adjunct chains.

This recapitulates the old observation that adjuncts are subject to more severe ECP-like island restrictions than arguments are. We can code this asymmetry, but we have no current explanation for why it should hold. As the reader no doubt knows, this is not a property unique to this analysis. We can incorporate the difference into our “*”-marking convention, and we do so in (93), but this is theoretically questionable. We have seen that the correct conditions for overt movement of arguments concerns the PF outputs produced. In this context notions like “heavyness” and “string-vacuity” make theoretical sense. They make considerably less sense in contexts where there are no phonetic effects, viz. LF effects that, in GB days, were the province of the ECP. Adjunct scope falls squarely into this second group. Thus, though (93) serves as a reasonably good description of the phenomena, it should not be understood as explaining the effect.25

8.5.2 The in situ analysis of the merger-type of sluicing

The present theory of movement is in consonance with the so-called Vacuous Movement Hypothesis, defended by George (1980), Chomsky (1986), and Agbayani (2006), among others. Thus, a sentence such as (94) below will have the rough structure given in (95) under our theory of movement:

(94) Who left?

(95) \[
\text{CP} <\text{who}> \ [\text{TP}\ WHO\ left]\]

Since the wh-movement to the Spec-CP is string-vacuous in (95), the bottom copy of the resulting chain must be pronounced according to the condition given in (62), which bans a given operation from having no PF output effects.

In this context, it is interesting to observe that what Chung et al. (1995) call the merger type of sluicing exhibits island insensitivity, as shown below:

---

25 See Nakao (2009) for a plausible approach to this effect in dealing with the argument-adjunct asymmetry concerning sluicing.
(96)  a. The administration has issued a statement that it is willing to meet with one
     of the student groups, but I’m not sure which one.
     (Chung et al. 1995: 272)

     b. Sandy was trying to work out which students would be able to solve a
certain problem, but she wouldn’t tell us which one. (ibid.)

(97)  a. They want to hire someone who speaks a Balkan language, but I don’t
     remember which language. (Merchant 2001: 87)

     b. Ben will be mad if Abby talks to one of the teachers, but she couldn’t
remember which. (ibid.: 88)

Recently, Kimura (2010) makes a very interesting proposal to deal with the island
insensitivity of this type of sluicing; according to her analysis, a remnant wh-phrase
in this construction stays in situ. She adopts Agbayani’s (2006) theory of movement
in terms of Move-F, according to which pied-piping functions as a repair operation
that recovers isolated features created by Move-F. Further, this operation is subject to
what Kimura calls PF Adjacency Condition, as stated below:

(98)  PF Adjacency Condition

     Features isolated by movement and the remnant wh-category must be phono-
logically adjacent.

This mechanism of movement tries to capture the fact that no vacuous movement
takes place. Thus, in such a case that involves wh-extraction of a subject, as in (94),
the wh-category who does not move to the Spec-CP, since it can satisfy the PF
condition (98) without undergoing pied-piping.

Given this mechanism of movement, Kimura (2010) claims that not only pied-
piping but also deletion can be exploited to meet the PF condition (98), and that
sluicing is exactly a case where this condition is satisfied by deletion. Thus, the sluice
of (99), for instance, has the derivation given in (100).

(99)  She’s reading something, but I can’t imagine what. (Chung et al. 1995: 241)

(100) a. I can’t imagine C_Q [she’s reading what]
     ↓ Movement of wh-feature

     b. I can’t imagine wh+C_Q [she’s reading [what ]]

     c. I can’t imagine wh+C_Q [she’s reading [what ]]

At the stage of (100b), what must undergo pied-piping so as to be adjacent to the
isolated wh-feature if the intervening material is not deleted. On the other hand, if it
is, *what* does not have to undergo pied-piping, as indicated in (100c), since it is adjacent to the isolated wh-feature even if it stays in its original position. This is how the in situ analysis of sluicing works under the mechanism of movement advocated by Agbayani (2006).

Our mechanism of movement enables us to dispense with the two-step operation of movement in terms of Move-F while maintaining the basic insight of Kimura’s analysis; that is, under our mechanism, a remnant wh-phrase undergoes “covert” movement in the merger type of sluicing. Thus, the sluice of (99) has roughly the following derivation:

(101) a. I can’t imagine \( [\text{CP} \ <\text{what}> \ C_{[+\text{wh}]} \ [\text{TP} \ \text{she’s reading} \ <\text{what}>] \)

b. I can’t imagine \( [\text{CP} \ <\text{what}> \ C_{[+\text{wh}]} \ [\text{TP} \ \text{she’s reading} \ \text{what}] \)

(101a) represents the structure created in narrow syntax, in which *what* constitutes a two-membered chain; that the two members of *what* are both put with angled brackets indicates that no decision is made yet with respect to which member is pronounced. The surface form indicated in (101b) is derived from (101a) by deleting the embedded TP except the wh-phrase *what*. Since the two members of *what* are adjacent to each other thanks to deletion, the bottom member must be pronounced according to the PF condition (62). This in situ analysis, then, explains the island insensitivity of sluicing, as shown in (96) and (97), since the island conditions are operative only to “overt” movement for arguments, as formulated in (93).

The most significant prediction of this in situ analysis is that unlike wh-arguments, wh-adjuncts should exhibit island sensitivity when they function as remnant wh-phrases in sluicing. The prediction is borne out, as noted by Merchant (2001):

(102) a. She’s practicing her serve so that she’ll be able to hit the ball in a certain deadly way, but her trainer won’t tell us in what way/??how.

b. He wants to interview someone who works at the soup kitchen for a certain reason, but he won’t reveal yet ??what reason/*why. (Merchant 2001: 129)

---

26 This analysis raises the question of how it is possible to delete a material that does not make a constituent. See Abe (2008) for a suggestion of how to approach this question and for independent empirical evidence for the need of such a non-constituent deletion in dealing with sluicing in such a wh-in-situ language as Japanese.
These facts are straightforwardly captured under the present assumption that adjuncts show island sensitivity even when they undergo “covert” movement.27

8.6 Conclusion

We have derived the fact that string-vacuous movement is locality insensitive from the following two main claims: (i) even though such movement creates a normal chain, it is the bottom copy that is pronounced due to a PF condition that bans the pronunciation of moved copies that have no PF effect, and (ii) such locality effects as caused by the Right Roof Constraint (RRC) and island conditions apply only to overt movement in which higher copies are pronounced. We have shown how these assumptions combine to account for the properties of RNR structures in English and analogous mirror image LNR constructions in Japanese.

27 See Abe (1993) and Tsai (1994) for the claim that adjunct wh-phrases in situ in regular questions must undergo covert movement to Spec-CP to be licensed, unlike argument wh-phrases. Though they attribute the island insensitivity of argument wh-phrases in situ to the availability of unselective binding, hence no need to undergo covert movement, our mechanism of movement re-opens the possibility, pursued by Huang (1982) and many others, that wh-arguments in situ actually undergo “covert” movement. We leave for future research the consideration of its consequences in the recent setting of the Minimalist Program, however.
Disharmony, Antisymmetry, and the Final-over-Final Constraint*

THERESA BIBERÄUER AND MICHELLE SHEEHAN

9.1 Introduction

Linearization, and more specifically, how hierarchical structure is related to linear order, has emerged as a central research topic in Generative Grammar in recent years. One highly influential theory of the relationship between the two was proposed by Kayne (1994):

(1) Linear Correspondence Axiom (LCA; Kayne 1994: 6)

For a given phrase marker P, [where \( d \) is the non-terminal to terminal dominance relation, \( T \) the set of terminals and \( A \) the set of ordered pairs \(<X_j, Y_j>\) such that for each \( j \), \( X_j \) asymmetrically c-commands \( Y_j \)], \( d(A) \) is a linear ordering of \( T \).

The basic insight of the LCA is that asymmetric c-command maps onto linear precedence.\(^1\) If the LCA holds universally at the mapping to PF, and we assume

* We gratefully thank the following for their comments, questions, and criticisms: our past and present colleagues on the Structure and Linearisation in Disharmonic Word Orders Project, Laura Bailey, Glenda Newton, Sten Vikner, and particularly Anders Holmberg and Ian Roberts; the audiences at the Ways of Structure Building workshop (Vitoria-Gasteiz, November 2008), at the DGfS workshop “Linking of sentential arguments in crosslinguistic perspective” (Osnabrück, March 2009) and at the SyntaxLab seminar (Cambridge, March 2009), particularly Joe Emonds, Ángel Gallego, Antje Lahne, Terje Lohndal, Neil Myler, Peter Oehl, Daniel Seely, Barbara Stiebels, Tonjes Veenstra, and George Walkden. Further thanks are due to two anonymous reviewers and the editors of this volume for their insightful comments. Finally, we also acknowledge the financial support of the UK’s Arts and Humanities Research Council (AHRC), which funds the on going research to which this chapter relates (Grant No. AH/E009239/1). As ever, we take responsibility for remaining errors and misunderstandings.

\(^1\) Initially, (1) was intended as a constraint operative in narrow syntax, aimed at deriving (a restrictive version of) X-bar theory. More recently, however, the LCA has been taken to apply only at the mapping to PF (cf. Chomsky 1995a,b, Moro 2000, and the Minimalist debate regarding the locus of linearization in the architecture of grammar, particularly Chomsky 2007, 2008, and Berwick and Chomsky 2011).
that the underlying hierarchy of phrase structure is universal (adopting Baker’s 1988
Universal Theta role Assignment Hypothesis (UTAH), and the general conclusions
of the Cartographic approach of Rizzi 1997 et seq., Cinque 1999 et seq. and others),
then it follows that head-final orders must be derived via movement of some kind.
This follows essentially because heads are (usually) non-branching whereas comple-
ments are (usually) branching; as such, a head will generally asymmetrically
\textit{c}-command into the complement it selects.\footnote{More must clearly be said about the deepest right branch under a Bare Phrase Structure (BPS)
approach (cf. Chomsky 1995a for a discussion of the problem and some potential solutions).}

Since its inception, the LCA has remained a controversial hypothesis. First,
empirical challenges have been raised by those working on predominantly head-
final languages (cf. Kural 1997 on Turkish, Murasugi 2000 on Japanese, and Elordieta
opposite view on Turkish, Basque, and Malayalam respectively). Second, theoretical
objections have been raised to the kind of movement which the LCA necessitates in
order to derive observed surface word orders (cf. Abels 2007, and Abels and Neele-
man 2009 for recent discussions). Nonetheless, the many striking typological asym-
metries observed by Kayne and proponents of Antisymmetry remain, and demand
explanation. Below is a representative, but by no means exhaustive, list taken from
2009, and Cinque forthcoming):

\begin{itemize}
\item[a)] Specifiers surface to the left, not the right;\footnote{It has been suggested that sign languages may constitute an exception to this generalization (cf. Neidle et al. 1998, Cecchetto, Geraci, and Zucchi 2006, but see Göksel et al. 2009, and Abner 2010 for counter-
arguments).}
\item[b)] There are no V-penultimate or clitic-penultimate languages, but there are V2
and clitic-second languages;
\item[c)] AuxV and VAux differ in that VAux order typically requires adjacency between
V and Aux; and
\item[d)] There are AuxOVX languages (e.g. many Mande languages), but there seem to
be no XVOAux languages.
\end{itemize}

Interestingly, it is not immediately clear, in relation to (a – d) that the LCA is actually
a necessary or sufficient condition to block the relevant gaps. Most analyses of V2, for
example, crucially rely on the assumption that specifiers surface to the left of heads,
reducing (b) to (a). Even in the absence of the LCA, then, the stipulation that
specifiers precede X’ rules out the possibility of V-penultimate languages regardless
of how the order of head and complement is regulated.\footnote{As Abels (2007) points out, moreover, V-penultimate could actually be derived via remnant move-
ment if such movement, which typically plays a prominent role in LCA-based derivations, is not
constrained; the same is true of superficially “rightward” wh-movement and, as Abels (2007) points out,
potentially of any conceivable order.} Similarly, the asymmetries in
(c)—(d) do not seem to fall out from the LCA directly. Nonetheless, we will argue that some version of the LCA is a necessary condition if we are to account for these asymmetries in an explanatory way. In particular, both (c) and (d) reveal the same basic asymmetry, imposed by the so-called Final-over-Final Constraint (FOFC). We argue that, while FOFC can apparently be stated without recourse to the LCA, an explanatory account of the effect is only possible if asymmetric c-command regulates word order in disharmonic (mixed head-initial/head-final) languages, and so the core of the LCA is crucial to an understanding of FOFC.

The chapter is structured as follows. Section 9.2 introduces the Final-over-Final Constraint and provides a summary of previously noted evidence in its favour. Section 9.3 outlines Biberauer, Holmberg, and Roberts’ (2010) LCA-based analysis of FOFC, before attempting to give a formalization of the constraint that does not make reference to the LCA. In section 9.4, however, we introduce data from CP placement which suggests both that an LCA-free account is inadequate and that aspects of BHR’s analysis may profitably be rethought. Section 9.5 provides a novel account of the CP placement facts, drawing in particular on Sheehan’s (2010b, forthcoming a,b) “complement stranding” proposals regarding the manner in which syntactic structure is linearized. Section 9.6 concludes, raising some issues for future research.

9.2 The Final-over-Final Constraint

A number of related papers highlight an until recently unnoticed asymmetry in attested disharmonic or mixed word orders: the so-called Final-over-Final Constraint (FOFC; cf. Holmberg 2000, Biberauer, Holmberg, and Roberts/BHR 2007, 2008, 2009, 2010, Biberauer, Newton, and Sheehan/BNS 2009a,b, Biberauer, Sheehan, and Newton/BSN 2010). While structures with a head-initial phrase dominating a head-final one are fairly well attested in certain domains, the reverse pattern, schematized in (2), appears to be extremely rare or, once certain provisos are taken into account, impossible:

(2) \[ XP [YP Y ZP] X \]

This fact is particularly striking in languages which permit a range of disharmonic word orders, (e.g. West Germanic languages, Finnish, Basque), but specifically disallow the order in (2).\(^5\) The empirical situation that we are faced with can be schematized as in (3):

\(^5\) We do, of course, acknowledge the fact that gaps in word-order patterns may, in fact, have a range of explanations (historical, processing-related, etc.). Nonetheless, the fact that we observe the same gap in divergent syntactic contexts suggests to us that FOFC is a deep property of language, if not syntax. See Sheehan (forthcoming b) and Biberauer, Holmberg, and Roberts (2010) for some objections to Hawkins’ (1994) processing-based approaches to FOFC.
(3) Harmonic and disharmonic combinations

\[
\begin{array}{cccc}
\text{a.} & \beta' & \alpha P & \gamma P \\
\text{b.} & \beta' & \beta & \alpha P \\
\text{c.} & \beta' & \alpha P & \gamma P \\
\text{d.} & * & \beta' & \alpha P \\
\end{array}
\]

very common  very common  less common  unattested


There are two main types of evidence for FOFC, which can be summarized as follows:

I. Synchronic evidence: the (virtual) cross-linguistic lack of the following word orders, even in languages permitting variable word orders:

(a) \([V>\text{O}]\rightarrow\text{Aux}\)  (BHR 2007 et seq.)
(b) Finnish: \([\text{N}>\text{Comp}]\rightarrow\text{P}\) vs. \([\text{P} \rightarrow \text{Comp}>\text{N}]\)  (BHR 2007 et seq.)
(c) \([\text{Asp}>\text{V(P)}]\rightarrow\text{T}\)  (Julien 2002, 2007, Myler 2009a)
(d) \([V>\text{O}]\rightarrow\text{C}\)  (Hawkins 1994, Kayne 1994, Dryer 2009a, BHR 2007 et seq.)
(e) \([\text{Pol}>\text{TP}]\rightarrow\text{C}\)  (BNS 2009a, BSN 2010)

II. Evidence that word-order change obeys FOFC-determined “pathways:”

(a) The history of English, French, Afrikaans OV\rightarrow\text{VO}  (BNS 2009a,b, BSN 2010)
(b) The history of Ethiopian Semitic VO\rightarrow\text{OV}  (BNS 2009b)
(c) The history of certain Niger-Congo languages VO\rightarrow\text{OV}  (Nikitina 2008)

We first provide a brief overview of two instances of type I evidence. We then discuss some apparent counterexamples before considering how FOFC should be captured theoretically. We leave type II evidence aside here, noting only that the historical record of languages investigated to date indicates that changes from head-final to head-initial ordering (“OV to VO”) proceed “top-down,” as indicated in (4a), whereas the reverse change (from head-initial to head-final, i.e. “VO to OV”) proceeds “bottom-up,” as shown in (4b).\(^{6}\)

\(^{6}\) Word-order change, therefore, does not seem to introduce FOFC-violating structures, exactly as we would expect if FOFC is a universal constraint and synchronic constraints on grammar shape diachronic developments (cf. Kiparsky 2008; see BNS 2009a,b, BSN 2010 for detailed discussion).
9.2.1 Polarity heads and complementizer placement

BNS (2009a) and BSN (2010) show that the order *[[Pol TP] C], where Pol is an initial polar question particle, is virtually unattested cross-linguistically. Assuming, following Laka (1994), that the head housing polarity is lower than that housing (finite subordinating) complementizers, this represents a FOFC-effect. Most suggestive that this is a formally imposed syntactic gap rather than an accident or diachronic effect is the fact that in the closely related Indo-Aryan languages, exactly those languages which have an initial question polarity (Pol) head lack a final complementizer (cf. Davison 2007). As an illustration, compare Hindi-Urdu and Marathi. Hindi-Urdu has an (optional) initial polarity head *kyaa and lacks a final complementizer of any kind (the forms included here to illustrate this are equivalent to final complementizers found in related languages, which have been sourced from demonstratives and “report” verbs respectively; cf. Bayer 2001):

a. *kyaa aap wahaaN aaeNgii?
   ‘Are you going there?’

b. *usee [[vee aa rahee haiN] yah/ kah-kar] maaluum hai
   ‘He/she knows [that they are coming]’

Marathi, on the other hand, has a final polarity head and both initial and final complementizers, which can co-occur:

a. [[to kal parat aalaa kaa(y)] mhaaNun/asa] raam maalaa
   ‘Ram was asking me [whether/if he came back yesterday]’

b. raam maalaa witSaarat hotaa [ki to kal parat
   ‘Ram was asking me [whether/if he came back yesterday]’

[Marathi, Davison (2007: 184), attributed to R. Pandharipande]
Furthermore, while there are many languages cross-linguistically with a final polarity head and an initial complementizer, there are remarkably few with the FOFC-violating combination (cf. Dryer 2005e, 2005f, and BSN 2010 for further discussion).\footnote{Interestingly, question particles do not seem to conform to FOFC in other syntactic contexts: VOQ, for example, is possible in a number of languages. We return to this issue at the end of section 9.2.3.}

9.2.2 An asymmetry in complementizer placement

Further support for FOFC comes from a well-known asymmetry in complementizer placement, namely the lack of final complementizers in VO languages. This can be seen to be another FOFC-effect once it is observed that FOFC holds transitively from head to head. Because a head-initial phrase cannot be \textit{immediately} dominated by a head-final phrase, it follows that it cannot be dominated \textit{at all} by a head-final phrase, meaning that head-finality must begin at the bottom of a given spine. As such, FOFC also accounts for the fact that many OV languages have clause-initial complementizers, but no VO languages have clause-final Cs (cf. Dryer 1992: 102, Hawkins 1994: 256–7, 2004, Cinque 2005, Zwart 2009b, 2010). The structures below show how VO...C structures are ruled out transitively by FOFC:

\begin{enumerate}
\item[(7)] a. *CP
\begin{itemize}
\item TP C
\item T VP
\item V Obj
\end{itemize}
\item[(b)] *CP
\begin{itemize}
\item TP C
\item VP T
\item V Obj
\end{itemize}
\end{enumerate}

FOFC violated between TP and CP \hspace{1cm} FOFC violated between VP and TP

9.2.3 Apparent counterexamples

It is important to note, however, that there are some obvious apparent counterexamples to FOFC, even in well-studied languages. One class of counterexamples involves head-initial DPs in OV languages such as German:

\begin{enumerate}
\item[(8)] Er hat [VP [DP ein Buch] gelesen]
\item[(9)] V'
\begin{itemize}
\item DP V
\item D NP
\end{itemize}
\end{enumerate}

‘He read a book.’
There are actually relatively few OV languages with clear determiners/articles which allow the configuration in (9).\(^8\) Most OV languages either (i) lack determiners distinct from demonstratives\(^9\) or (ii) have final determiners (cf. Dryer 1992: 104, 2005a,b). Nevertheless, OV languages with phrase-initial articles are no less common (taking into account areal and genetic factors) than VO languages with phrase-final articles. There is therefore no evidence of a FOFC effect in this domain. This needs to be acknowledged by any approach to FOFC. The pattern with prepositional phrase complements of V is more complex, and we leave it aside here (cf. Sheehan forthcoming b for details).

The second class of counterexamples comes from final particles in many VO languages.

\[
\begin{align*}
\text{(10) a. } & \text{Tā huì shuō zhōngwén } ma? \quad \text{[Mandarin Chinese]} \\
& \text{3SG can speak Chinese Q} \\
& \text{‘Can he speak Chinese?’} \\
& \text{(cited in Paul forthcoming)} \\
\text{b. } & \text{Znàso baasé Ranti } \text{yé} \quad \text{[Mumuye (Niger-Congo)]} \\
& \text{Znasó mimic Ranti PERF} \\
& \text{‘Znasó mimicked Ranti’} \\
& \text{(cited in Dryer 2009b: 344)}
\end{align*}
\]

Interestingly, it seems to be a property of particles generally that they fall outside typological word-order generalizations, and for this reason Greenberg (1963) simply eliminated them from his sample. We leave a full explanation of the behavior of these elements to future research (cf. also Biberauer and Sheehan 2011, and Biberauer, Holmberg, and Roberts 2010 for further discussion). Three observations are, nonetheless, worth making here.

First, though apparently FOFC-violating particles such as that in (10a) are often labeled C-elements, they are very clearly not involved in subordination, and are often limited to matrix clauses. As such, they do not constitute an exception to the robust pattern described in section 9.2.2. Second, in many cases, the language in question has a non-particle equivalent of a certain category which is “well-behaved” with respect to FOFC. A relevant case is illustrated below:

---

\(^8\) This is a FOFC violation only if articles are heads, as proposed by Abney (1987). As this is the standard assumption in the literature, we will assume for the moment that it holds, but see section 9.5.7 for added complications.

\(^9\) According to standard assumptions (cf. Giusti 1997, 2002), demonstratives are actually specifiers rather than heads and so pre-nominal demonstratives should not count for the formal account of FOFC. See, however, Guardiano (2009) for detailed consideration of the cross-linguistic distribution and properties of demonstratives.
(11) a. Tân mua gì the?
    Tan buy what PRT
    ‘What did Tan buy?’

   b. Anh dã nói (rằng) có ta không tin
    PRN ANT say that PRN PRT NEG believe
    ‘He said that she didn’t believe (him).’

In (11), we see that Vietnamese features both an optionally realized subordinator (rằng) and clause-final force particles, such as the question particle in (11a). Striking here is the fact that the subordinator shows the expected behavior—head-initial placement—while the particle appears to compromise FOFC. This more generally evident pattern (cf. also Northern Italian dialects and West Flemish, as discussed in Munaro and Poletto 2004, and Haegeman forthcoming respectively) suggests that “full” and particle elements expressing similar sorts of information are formally distinct. A final point worth noting in connection with particles is that they may, in some cases, be subject to FOFC in one direction, but not in another. For example, we have seen that initial question particles block the presence of final complementizers (cf. section 9.2.1), but, curiously, question particles do not seem to be subject to FOFC “downwards,” as they are relatively common in VO languages (cf. (10a) above). Particles, then, evidently pose a range of particularly interesting word-order-related questions, including a degree of immunity to FOFC. There are a number of hypotheses which might explain the special behaviour of particles, but we put them to one side here, for reasons of space (cf. again Biberauer and Sheehan 2011, and BHR 2010 for discussion).

9.3 FOFC and the LCA

The most recent analysis proposed by BHR (2009, 2010) views FOFC as a Relativized Minimality effect (cf. Rizzi 1990a et seq.). The central idea here is that the “default” order of head and complement is indeed head>complement, as Kayne (1994) proposes.10 Head-final orders are then triggered by the presence of a movement
diacritic, designated ^ by BHR. More specifically, BHR assume that the association of ^ with a head’s c-selection feature triggers comp-to-spec movement within the relevant phrase, giving head-final structure (cf. also Holmberg 2000 and Julien 2002).\footnote{This relies on the not uncontroversial assumption that comp-to-spec movement is actually possible in certain circumstances, contra Pesetsky and Torrego (2001) and Abels (2003). Julien (2002, 2007), however, provides empirical arguments that comp-to-spec movement must be permissible in the specific case where it is driven by c-selection; Agree-driven comp-to-spec movement such as that discussed by Pesetsky and Torrego (2001) could still be ruled out on this view. Note, furthermore, that from BHR’s perspective, comp-to-spec movement alters the c-command relations of a phrase, meaning that a specifier asymmetrically c-commands the projecting head. As such, inasmuch as movement to a specifier position can be justified more generally, local movement to spec is no different, and no more problematic from a theoretical perspective. If Julien, BHR, and others who assume comp-to-spec movement are correct, an alternative explanation will, therefore, clearly have to be found for Abels’ observations.}

Crucially, BHR argue that: (i) ^ can only be present in an extended projection of the lexical head, X, if it is also present on X; and (ii) the distribution of linearization (i.e. c-selection)-related ^ is constrained such that a higher (projecting) head on a given spine cannot bear ^ if the lower head(s) on that same spine do(es) not do so. This is stated in (12):

\begin{equation}
\text{(12) If a head } X^{n} \text{ in the extended projection } E \text{ of a lexical head } L \text{ has } ^{\wedge} \text{ associated with its c-selection feature for a lower head } X^{n-1}, \text{ then so does } X^{n-1}.
\end{equation}

Thus v, for example, can only bear ^ if V does so too. Since auxiliaries are typically assumed to be in the extended projection of V, VOAux orderings are immediately ruled out.\footnote{Where auxiliaries fail to inflect and do not appear to be integrated with the clausal spine via the usual c-selection and Agree-relations observed in familiar European languages, VOAux orderings may, however, arise. This is arguably the case with non-inflecting auxiliary particles of the type found in many Central and West African languages (cf. Dryer 2009b for discussion). See Cecchetto and Donati (2010) for recent discussion of the role of c-selection and Agree features in the context of projection, and, Toivonen (2003) for the earlier idea that particles may be defective, being unable to project. Marcel den Dikken (p.c.) additionally points out that, in the specific context of particle auxiliary-elements which appear to exhibit very specific co-occurrence restrictions (e.g. those in the Chinese varieties), it may in fact be the case that these elements instantiate secondary predicates, i.e. lexical items, which therefore cannot, in terms of (12), form part of the extended projection of the lexical verb. This proposal has the appealing consequence that VOAuxPart structures are ruled in on the same basis that particle-verb structures featuring the order VOPart are ruled in, i.e. because PartP does not dominate VP.}

On extended projection definitions of the type assumed by Grimshaw (1991, 2001, 2005) and van Riemsdijk (1998a), V-O-C orders are also ruled out, as C is contained in V’s extended projection. Consider the following from Grimshaw (2000):

\begin{quote}
Hawkins 1999) and it also seems to be the case that the original EPP (Chomsky 1981) holds less systematically in OV and V-initial languages that have been said to employ “roll-up” movement than in familiar VO languages (cf. Pearson 2000, Julien 2002, Richards and Biberauer 2005, and Öztürk 2009), suggesting that what is true in other areas of language—namely, that complexity in one (sub)domain is off set by lesser complexity in another—may also be true in the movement context.
\end{quote}
X is a head of YP, and YP is a projection of X iff:

1. YP dominates X
2. the categorial features of YP and X are consistent
3. there is no inconsistency in the categorial features of all nodes intervening between X and Y (where a node N intervenes between X and YP if YP dominates X and N, N dominates X, and N does not dominate YP) and
4. no node intervening between X and YP is lexical.

In the context of this proposal, then, head-finality either starts with the lexical head at the bottom of a given extended projection (V in this case), or else it is impossible. Just as Relativized Minimality rules out “skipping” in the domains of different types of A-, A’-, and head-movement, BHR argue that it also rules out “skipping” in the distribution of linearization-related ^. A formalization of FOFC in these terms can be schematized as follows:

(14) Parameter 1: ^ is present on L
    Parameter 2: ^ is present on Hn (where Hn selects L^ or Hn-1^).

Let us now consider how this proposal can account for the discrepancy between FOFC-compliant and, at first sight, FOFC-non-compliant structures. The analysis provides a potential account for why FOFC fails to hold between V and a DP complement in that V and D are contained in distinct extended projections and the lexical head V may therefore freely bear or lack ^ independently of the directionality of the DP. Whether Ps are lexical or functional extensions of N, we expect them, likewise, not to restrict the directionality of V. To the extent that particles are

---

13 We will see in the next section, however, that this cannot be the whole story, as CP complements of V pattern differently: while DPs can both precede and follow V, independently of whether they are head-initial or head-final structures, head-initiality vs. head-finality clearly matters for CPs since only head-final CPs may precede V. Head-initial CPs therefore differ from head-initial DPs in that they cannot precede V in languages with otherwise head-final VPs.

14 Holmberg (2000), however, observes that the FOFC-violating order is barred within Finnish PPs, which permit both types of harmonic orders and also the non-FOFC-violating disharmonic order. This implies that these postpositions must form part of the extended projection of N, as proposed by Grimshaw (1997, 2001, 2005). More problematic is the status of circumpositional structures in Germanic and
in some sense categorially deficient, unable to project (cf. Toivonen 2003), lacking in c-selection features (cf. BHR 2010 for discussion of so-called “syncategorematic” particles) or to which they are lexical items, specifically non-projecting lexical heads defining trivial extended projections, it may also be possible to understand why particles do not appear to be subject to FOFC in the same way as ordinarily projecting heads.

Returning to the role of the LCA in the context of this proposal: BHR explicitly state their analysis in terms of the LCA, but one might question whether it is in fact necessary to view ^ as a syntax-internal movement-trigger; couldn’t it, instead, be interpreted as a PF-interpretable instruction, which does not in fact play a role in narrow syntax? In other words, what precludes ^ being a linearization feature which is interpreted by PF as signalling that the head with which it is associated is to be linearized to the left of its complement rather than to the right as would be the case if the head in question lacked ^? This would be compatible with a model in which specifiers are consistently located on the left, with the result that movement must be leftwards, but head-complement order is determined on the basis of a PF-parameter (cf. Abels and Neeleman 2009, Richards 2004, 2008 for ideas of this kind).

In the following sections, we examine the placement of embedded CPs, which exhibit what we term a “FOFC-compliance effect.” It is our contention that these effects cannot be adequately explained unless some version of the LCA is assumed and ^ is interpreted as a syntactically interpretable diacritic. More specifically we argue that the PF-parameter interpretation of the constraint underlying FOFC cannot account for the phenomenon of clausal extraposition, whereas a movement feature-based one can. The crucial distinction between CP and DP arguments, we claim, is that the latter behave like “atoms” from a layered derivation perspective.

9.4 CP extraposition

In this section, we focus on a phenomenon which has important repercussions for our understanding of FOFC: the embedding of clausal constituents under V. According to the definition of extended projection given in (13), FOFC effects are not predicted to hold across clausal boundaries. By (d), merger of matrix V necessarily initiates a new extended projection, as was the case for the DP and PP complements discussed above. BHR’s definition of EP, however, removes this proviso, thus elsewhere, which appear to feature head-initial PPs dominated by head-final phrases of the same type. We leave these matters to future research.
allowing EPs to extend across clauses. As presented in section 9.3, though, BHR’s proposal, leads us to expect that OV languages with initial complementizers should also not exist: if the V-initiated spine of embedded complements is indeed continuous with the verbal spine of the matrix clause, (12) entails that it will not be possible for matrix V to bear ^, while C, a lower head c-selected by V, does not. In these terms, FOFC therefore makes the apparently correct prediction that head-initial CPs will not surface preverbally in OV languages, as we discuss below, but it also makes the manifestly incorrect prediction that OV languages will systematically lack head-initial CPs. In fact, Dryer (2009a) registers 54 of the 640 OV languages in the WALS sample as having initial Cs. Furthermore, these head-initial CPs are systematically extraposed. The purpose of this chapter is to consider in detail the structure building considerations that underlie this empirical fact. In particular, we will show that BHR’s (2009, 2010) Relativized Minimality-based version of feature distribution coupled with Sheehan’s (2010, forthcoming a,b) complement stranding (CS) analysis combine to account for all the data. Crucially, this is only true if we adopt the version of BHR’s model whereby ^ equates to a trigger bringing about narrow syntactic movement.

9.4.1 Embedded clauses in OV languages

It is well known that many (otherwise) OV languages obligatorily place CP complements to the right of V (cf. Hawkins 1994, 2004, Dryer 2009a, Cinque 2009). In keeping with tradition, we will call this “extraposition” without committing to any analysis at this point. Thus German, an otherwise OV language, obligatorily extraposes (non-scrambled) CP complements to the right of V:

15 BHR’s precise formulation is given in (i):

(i) The Extended Projection of a lexical head L (EP(L)) is the sequence of categories EP = \{\alpha_1, \ldots, \alpha_i, \ldots, \alpha_n\} such that:

(a) \alpha_i is in the spine defined by L;
(b) and for each pair of heads <Hi, Hi+1> in EP, Hi c-selects Hi+1;
(c) Hi is categorially non-distinct from Hi+1.

Since, in the case we are concerned with, matrix V selects CP-complements and is also categorially non-distinct from C (which would instantiate the “top end” of Grimshaw’s extended projection), matrix V and its extended projection count as part of the same extended projection as that initiated by the selected clause’s V.

16 Interestingly, BHR’s proposal does not rule out the possibility of mixed languages in which matrix clauses are systematically head-initial (i.e. rigidly “VO”), but embedded clauses are consistently head-final (i.e. rigidly “OV”). To the best of our knowledge, however, languages of this type do not exist. Neither Cinque (2005) nor Dryer’s (2009a) extensive surveys of clausal placement in VO and OV languages mention such languages.

17 As BHR and the current paper have an overlapping chronology, BHR in fact refer readers to the current paper for an account of the mechanics of CP extraposition.

18 Cf. Büring and Hartmann (1997) and Barbiers (2000) for discussion of the clearly scrambled, i.e. A’, nature of CPs surfacing to the left of V.
... weil er gesagt hat [\(\text{CP dass} \) Schnaps gut schmeckt] because he said has that Schnaps good tastes

‘… because he said that Schnaps tastes good’

[Büring and Hartmann 1997: 32]

The debate surrounding the correct analysis of this phenomenon is vast and what follows is a somewhat crude summary of the main empirical facts and their implications. Crucially, we focus only on the obligatory extraposition of complement CPs. Adjunct and relative clause extraposition from NP has very different properties, so much so that it appears to be a totally distinct phenomenon (cf. Fox and Nissenbaum 1999 and Sheehan 2010b for the same claim about the corresponding structures in English).

9.4.2 Empirical facts about extraposed CPs

Obligatorily “extraposed” complement CPs in many OV languages show Condition C effects in relation to an indirect object in the matrix clause (cf. Aghaei 2006 on Persian, Bayer 2000 on Bangla, and Haider 1993 and Büring and Hartmann 1997 on German):

(16) *Dāvar [be ‘u] goft [\(\text{CP ke [Ali]} \) bāzi- ro mi- bar-e] referee to him said that Ali game- ACC IMPF win-3SG

\(\not=\) 'The referee said to him (Ali) that he, Ali, had won the game’

[Persian, Aghaei 2006: 156]

(17) *Die skeidsregter het vir hom gesê dat Piet die spel gewen het the referee has for him said that Pete the game won has

\(\not=\) 'The referee said to him (Pete) that he, Pete, had won the game’

[Afrikaans]19

They therefore pattern with their counterparts in English and other VO languages, which show the same effect:

(18) *I told heri that Maryi was right

(19) *Disse- lle, que a Maria, tinha razão said.1SG her that the Maria had reason

\(\not=\) 'I told her (Mary) that Mary was right’ [Portuguese]

(16)-(19) suggest that the indirect object in both types of languages c-commands CP. This impression is reinforced by the fact that a quantified indirect object can bind a pronominal contained in an extraposed CP in the same languages:

(20) Dāvar-ā [be har bāzikon]i -i, qowl dād-an [ke faqat [‘un]i referee-PL to every player- INDEF promise gave-3PL that only he/she bāzi-ro be-bar-e]

game-ACC SBJ-win-3SG

‘The referees promised every player that only he/she will win the game’

[Persian, Aghaei 2006: 157]

19 Examples cited without references to the work of other scholars were collected by the authors.
The referees have every player promised that he/she will win.

‘The referees promised every player that he/she will win’  [Afrikaans]

Postverbal CPs both in VO and in OV languages of the type illustrated above therefore appear to occupy a VP-internal position c-commanded by indirect objects. Moreover, if Huang’s (1982) Condition on Extraction Domain (CED) holds, the fact that these extraposed CPs are not strong islands in languages like German, Dutch, Afrikaans, Hindi-Urdu, and Persian further suggests that they must be in a complement position:

Which team did Ram say that Mohan thought would win?20

[German]

According to the CED, or rather Uriagereka’s (1999) minimalist reinterpretation of it, extraction is only possible from complement positions. It would appear, then, that “extraposed” CPs occupy a complement position inside VP.21

An obvious complication for the view that complement CPs are located inside VP arises from the fact that a complement CP cannot interrupt a verb cluster in German or Dutch, and so always surfaces in final position, often separated from its selecting head:

‘… that Jan had said that he had done it’

20 This example, and others like it, appear to indicate that Hindi-Urdu lacks that-trace effects. This is not so surprising given that it has been shown that movement to Spec-TP in Hindi-Urdu does not involve Agree-driven movement of the type found in languages associated with an English-style canonical subject position (cf. Nevins and Anand 2003). Hindi-Urdu subjects, then, may well be extractable from a lower position, as is the case in familiar Romance null-subject languages, or it may be the case that non-Agree-driven movement to Spec-TP does not result in the “freezing” effect observed in languages like English (cf. Rizzi and Shlonsky 2006 for recent discussion).

21 Ian Roberts (personal communication) asks why certain extraposed clauses behave like weak islands, a fact noted by Cinque (1999). We have no explanation for this fact, in as far as it holds. Note, however, that weak islandhood, unlike strong islandhood, displays no interaction with Huang’s CED effects, and, as such, does not bear on the status of the CP as a complement.
If the CP does indeed occupy its base-position, the V selecting it must somehow have raised higher, either via head or phrasal movement. As an anonymous reviewer points out, verb clusters clearly cannot be formed via verb movement, as the stranding of a particle in post-auxiliary position is ungrammatical in such contexts:

(26) a. *... dat Jan hem vaak gebeld hat op
   that Jan him often called has up
   ‘... that Jan has often called him’

As VP-adverbials generally also precede Aux (cf. (26)), we take the relevant movement to be phrasal VP movement, rather than head movement (cf. Biberauer and Roberts 2005, 2006 for discussion and references). We therefore reach the apparently paradoxical conclusion that CPs are spelled out in their base complement position inside VP, but that the verb itself has raised higher via VP movement. The explanation for this paradox, we would like to suggest, stems from the availability of scattered deletion, and the following tree represents a (simplified) version of the structure we will ultimately propose for CP-extraposition in OV languages such as German (see section 9.5.4):

(27)

As (27) shows, we propose that CP remains in a complement position inside VP, moving along with it to Spec-AuxP (to satisfy Aux’s ^ feature), but then being spelled out at PF in its base position (for principled reasons to be elaborated in sections 9.5.3 and 9.5.4). V, on the other hand, is spelled out in its derived position, which may, of course, be higher than indicated here (for example, in the context of biclausal
In addition to explaining why CP is obligatorily stranded under VP-movement, we must also explain why CP fails to precede V in such cases, given that V in the languages in question (e.g. German, Dutch, Hindi, Persian) bears ^, as illustrated by the fact that all DP and PP arguments must surface preverbally.

Before we turn to our proposal, we consider one particularly influential account of the different distributions of DPs and CPs: that based on Case. We show, however, that an account of this type is not able to account for the full range of data. Consequently, we propose an alternative account, which draws on a distinction between DPs and CPs when these are viewed from a layered derivation perspective (cf. Uriagereka 1999, Johnson 2002, and Zwart 2009a,b, 2010, 2011).

9.5 Accounting for CP extraposition

9.5.1 Case resistance vs. directionality

Many previous analyses of CP extraposition have drawn on Stowell’s (1981) Case-Resistance Principle (CRP; cf. van den Wyngaerd 1989, various contributions in Lutz and Pafel 1996 and Beerman et al. 1997, Büring and Hartmann 1997, Simpson and Bhattacharyya 2003, Oehl and Lofti to appear). The idea in such approaches is that CP, in contrast to DP, does not need or even cannot bear Case.

There are a number of reasons to reject such an account of clausal placement.

First, it is important to note that not all OV languages extrapose CP complements of V. Japanese allows embedded head-final CPs, as well as clausal nominalizations headed by light nouns such as koto (thing), no (event) or tokoro (place) (cf. Horie 2000 and Suzuki 2000). Unlike clausal nominalizations, embedded CPs are never case-marked, but they can occur in the immediate preverbal position, following VP-adverbials (they can also be scrambled to pre-adverbial or initial position, but, unlike what we see in West Germanic—cf. note 18—this is only an option in Japanese; the neutral position for Japanese embedded CPs is in the immediately preverbal position,

CRP approaches take two main forms, depending on the assumptions that are made about phrase structure:

(i) Head parameter-based theories: being non-nominal, the clause resists Case by moving out of the preverbal Case position; or
(ii) LCA-based theories: being non-nominal, the clause resists Case by failing to move into the preverbal Case position.

We abstract away from these differences here as both approaches will be shown to be empirically problematic.

It is potentially important to differentiate between the Indo-European languages under discussion in which only CP arguments surface post-verbally, and a number of other languages which also display a mixed VO/OV character, but in which only one DP argument surfaces preverbally, even in ditransitive constructions (e.g. many Niger Congo languages) (cf. Biberauer and Sheehan 2011).
i.e. the unmarked position for DPs, PPs, etc. which is not available to West Germanic
finite CPs).

(28) Watashi-wa (chuibukaku) [kare-ga sore-o yatta kadooka to]
I-TOP carefully he-NOM that-ACC did Q C
(*-o) (chuibukaku) zunemashita
ACC carefully asked
'I (carefully) asked if he did that' [Japanese]24

Moreover, as also pointed out by Bayer, Schmid, and Bader (2005), partially defec-
tive25 clausal complements to V can optionally surface in the canonical preverbal
object position in languages with CP-extraposition (cf. also Kornfilt 1997, 2001, and
Özsoy 2001 on Turkish, and Biberauer and Roberts 2008 on West Germanic; as we
will see in following discussion, more radically reduced clausal complements are
necessarily preverbal):

(29) a. . . dass Max mir empfohlen hat [PRO das Lexikon zu kaufen
that Max me recommended has the dictionary to buy
b. . . dass Max mir [PRO das Lexikon zu kaufen empfohlen hat
that Max me the dictionary to buy recommended has
‘. . . that Max recommended I buy the lexicon’
[German, Bayer et al. 2005: 1]

There is nothing to suggest that the complement in (29b) is nominal (and so Case-
bearing) whereas that in (29a) is not. Although -en originated as a nominal ending
(cf. Los 1999, Roberts and Roussou 2003), zu-clauses, unlike dass-clauses, cannot be
coordinated with nominals synchronically, suggesting that they are not nominal (cf.
(30), and cf. Bayer 2009 on dass-clauses, which we return to in section 9.5.4):

a cup coffee and books to buy are
zwei meiner Lieblingsbeschäftigungen
two my-GEN favourite-pastimes

24 Thanks to Makiko Mukai for these judgments and discussion of the data.
25 Control clauses of the kind illustrated in (29) are described here as "partially defective" owing to the
fact that they are generally thought to be CPs (cf. Martin 1996 for discussion and references), while,
however, being unable to license nominative Case. Other infinitival clauses are typically thought to
instantiate smaller sub-parts of the clausal spine, i.e. to entail smaller extended projections (cf. Wurmbrand
2001 for discussion in relation to Germanic specifically). Clausal complements may, therefore, be defective/
reduced to varying degrees, and, as we proceed to show in this chapter, building on Biberauer and Roberts
2008, degree of defectiveness also appears to correlate with clausal placement in a manner which invites a
structural explanation.
The problems with CRP approaches are even clearer in Turkish, which has a range of nominalization and embedding strategies. Some Turkish clauses appear to conform to the predictions of the CRP, as nominalized clauses occur in a preverbal position and are marked with accusative Case, whereas (more marked) non-nominalized CPs headed by *ki* (a Persian borrowing) are extraposed, and have no overt case-marking.26

(31) a. (Ben) siz-in Ankara-ya git-tiğiniz] -i duy-du-m
   I you-GEN Ankara-ya go-2PL-NOM-POSS-ACC heard-PAST-1SG
   ‘I heard that you went to Ankara’
   [Özsoy 2001: 216]

b. Anla -dī -m [CP *ki onun bir derdi var
   understand PAST-1SG-3SG.GEN one problem. POSS.3SG exists
   ‘I realized that he had a problem’
   [Haig 2001: 201]

However, there are also certain predicates whose “direct complements” must surface in the preverbal position, despite the fact that they lack nominal properties (i.e. genitive subjects and nominal morphology) and are not marked with accusative Case:

   we you-NOM Ankara-DAT go -PAST -2SG consider -PAST -1PL
   ‘We consider you to have gone to Ankara’
   [Özsoy 2001: 217]

b. *Biz san -dī -k [sen- ∅ Ankara-ya git -ti -n]
   we consider-PAST-1PL you-NOM Ankara-DAT go -PAST -2SG

These direct complements show no evidence of a Case requirement. In German too, certain reduced clauses must surface in the canonical object position without displaying nominal properties (cf. Biberauer and Roberts 2008):

26 In actual fact, accusative case-marking with nominal objects presents a problem for the idea that movement to a preverbal position always correlates with accusative Case assignment as non-specific nominals occur in the immediate preverbal position and yet bear no case-marking (cf. Kornfilt 2003). We abstract away from these complications here, but note that they serve to reinforce our point that OV/VO order in Turkish cannot be explained by patterns of Case-assignment.
Rather than Case, what appears to be crucial in regulating the placement of embedded clauses is directionality and "deficiency" (cf. also Biberauer and Roberts 2008, who cite Wurmbrand’s (2001) generalization in this connection). Let us consider these factors in turn. That "deficiency" is relevant becomes clear if we compare (29) and (33): whereas control complements (i.e. CPs) may surface both pre- and postverbally, TP-complements like those of raising scheinen ("seem") necessarily surface preverbally, and the same is true for other reduced clausal complements (cf. Biberauer and Roberts 2008 for detailed discussion, and Cardinaletti and Starke 1999 for discussion of the relevance of structural deficiency in determining placement). Directionality is also clearly relevant since harmonically head-final clauses may surface both pre- and postverbally, whereas clauses with a head-initial CP are obligatorily extraposed.

Further evidence that this is correct comes from Indic languages, where it is the position of C (initial or final) which regulates CP-placement. In the languages with both initial and final Cs, the placement of embedded clauses depends wholly on the position of the embedding complementizer (cf. Pandharipande 1997, and Nayudu 2008 on Marathi, Bayer 1999, 2001 on Bengali, and Davison 2007 on Indic more generally). We illustrate on the basis of Bengali, in which head-initial CPs are obligatorily postposed (34a) and cannot surface either in the immediate preverbal position (34b) or clause-initially (34c):

(34) a. chele-Ta Sune-che [CP je [or baba aS-be]]
      boy-CF hear-PTS3 C his father come-FUT
      ‘The boy has heard that his father will come’
      S V CP

b. *chele-Ta [CP je [or baba aS-be]] Sune-che *S CP V

c. *[CP je [or baba aS-be]] chele-Ta Sune-che *CP S V

[Dec 1999: 17]

Head-final CPs, in contrast, can be preverbal, surfacing either in the immediate preverbal position (35a) or in sentence-initial position (35b) (the preferred option). Worth noting in this case, though, is that "extraposition" is said to be marginally available in (35c). Clearly, then, head-final CPs may surface in the position canonically reserved for head-initial CPs, but the reverse possibility (34b) is not available:
Crucially, the etymological source of the complementizer (final C/QUOT > participial verb of “saying” and initial C > relative pronoun) does not appear to affect the placement possibilities of the clause it heads in any obvious way; therefore, it does not seem feasible to look for an account of the CP-placement facts which appeals to diachronic considerations.27 Processing considerations, such as those highlighted by Hawkins (1994, 2004), offer a plausible explanation for the tendency to extrapose head-initial CPs, but are challenged by the possibility of extraposing harmonically head-final CPs (e.g. so-called coherent infinitives in German—cf. Bech (1955/1957) and (29a) above—and (more marginally) CPs in Indic, cf. (35c) above), and by the fact that the avoidance of center-embedding is not always obligatory: preposing head-final CPs in languages like Japanese is optional, whereas post-posing head-initial CPs in OV languages appears to be obligatory (for a discussion of some of the problems facing a processing-based account of FOFC see Sheehan forthcoming b; for a more general critique of Hawkins’ approach see Mobbs 2008 and Walkden 2009).

To summarize what we have established in this section, then: clausal complement placement cannot plausibly be attributed to Case-related, diachronic, or processing considerations; instead, the relevant factors appear to be the “deficiency” of the clausal complement and its headedness (initial or final).

9.5.2 FOFC and layered derivations

The purpose of this section is to argue that the fact that head-initial complement CPs in many OV languages surface in their base position to the right of V (and everything else) is related to FOFC, and follows from the way in which derivations proceed. (36) illustrates the superficial relevance of FOFC:

27 In this regard, note that Marathi has two final-Cs, one derived from a demonstrative and the other from the verb “to say.” Clauses headed by both types of final C mirror the Bengali facts in (34) (cf. Pandharipande 1997, Nayudu 2008, chapter 2).
As (36) shows, a head-initial CP surfacing in preverbal position violates the FOFC generalization in (2), repeated here as (37):

\[(36) \quad \ast \quad V' \quad CP \quad V \quad C \quad TP \]

Crucially, (37) remains mysterious if it is conceptualized as following from a constraint on the distribution of a PF-interpretable linearization feature, rather than from the operation of a syntactically visible movement diacritic. This is because a PF-parameter interpretation of BHR’s formalization of FOFC ((12)), mooted as a possibility in section 9.3, is evidently not sufficient to account for the positional discrepancy between head-initial DPs and CPs in OV languages of the type we are considering here: in both cases, V will bear \(^\wedge\), with the result that we would expect PF to spell out unattested \([CP]\) V structures alongside the attested \([DP]\) V ones. Postulating two Vs, with DP-selecting V bearing \(^\wedge\) and CP-selecting V lacking \(^\wedge\) obviously just replicates the observed facts, without offering an explanation of any kind. Moreover, even if this option is taken, the additional FOFC effect observed with verb clustering fails to be explained: as the discussion surrounding (24)–(25) shows, \([V[CP][Aux]]\) fails to arise in neutral cases; instead, the relevant structures necessarily exhibit \([V][Aux][CP]\) ordering, suggesting that extraposition in these contexts is also triggered as a FOFC-compliance strategy. The problem is further exacerbated by the availability of three- and even four-verb clusters, none of which permit FOFC-violating interpolations (cf. Bader and Schmid 2009 for recent discussion of West Germanic verbal clusters). Viewing \(^\wedge\) as a diacritic which is blindly interpreted at PF as “linearize this head’s complement to its left” can rule out these further FOFC effects only by stipulation. We therefore reject this interpretation of the constraint underlying FOFC given in (12).

As noted in section 9.3, BHR in fact interpret \(^\wedge\) as a movement trigger which is visible during the syntactic computation. For BHR, then, the impossibility of (36) follows from the fact that this structure would entail a configuration in which (a) a higher head (V), bearing a c-selection-related \(^\wedge\), selects a category lacking this species of \(^\wedge\) and (b) selector and selectee form part of the same Extended Projection (EP). Importantly, (b), which is central to BHR’s formal characterization of FOFC, only follows if Grimshaw’s intuitively appealing notion of EP, given in (13) above, is modified such that EPs extend across clauses, with the most deeply embedded verb initiating a verbal “spine” that terminates at the top of the matrix clause (cf. note 15 for discussion). Equally importantly, this modification alone only accounts for why (36)-type structures are ruled out; it does not offer an explanation for why languages with head-final VPs systematically extrapose head-initial CP complements, and, more importantly, why such structures become legitimate.
Here, we uphold BHR’s proposal that \(^\rightharpoonup\) in fact instantiates the movement diacritic which is more generally assumed to be visible to the computational system and which has been variously referred to as i.a. an EPP-feature, an Edge Feature and \(*\).\(^{28}\) We, however, depart from BHR in retaining Grimshaw’s definition of EP, in terms of which EPs feature only one lexical head, namely that initiating the EP at the bottom of the relevant “spine.” This departure naturally reintroduces the complication which BHR’s reformulation sought to circumvent, namely that it is no longer clear why structures like (36) are ruled out: if V, being a lexical head, begins a new extended projection, it should be free to bear \(^\rightharpoonup\), regardless of the directionality of its complement. The fact that we do not observe a FOFC effect between V\(^\rightharpoonup\) and the head-initial DP it selects in languages like German (cf. (8) and (9) above) therefore follows straightforwardly, but the fact that CP complements of V behave differently from DPs, and display an apparent FOFC-compliance effect emerges as problematic. In what follows, we will show that if one adopts Zwart’s (2009a, 2010, 2011) proposal that derivations proceed in “layers,” a potential difference between the prototypical instances of DPs and CPs emerges: generally speaking, DPs behave like strong islands, whereas CPs do not. We return to the apparent counterexamples to this simplified picture in section 9.5.6, adopting a simplified perspective in the first instance for expository convenience.

Let us first consider how the manner in which CPs and DPs are constructed contributes to the discrepancies in their islandhood and also linearization behaviour.

Uriagereka (1999), Johnson (2002), and Zwart (2009a,c, 2010, 2011) have all proposed versions of what Zwart calls a “layered derivation” approach to structure-building. The basic idea in all cases is that certain phrases are constructed on the basis of lexical arrays distinct from that associated with the main clausal structure, being “atomized” (or converted into complex lexical items) before being inserted into larger, clausal structures. In the context of the subsequent derivation, these “atomized” structures then behave roughly like lexical items, subject to lexical integrity.

It is important to note that this kind of “Multiple Spell-Out” (Uriagereka’s term) differs from Chomsky’s notion of “derivation by phase” in that it is thought to apply

\(^{28}\) It follows from the LCA that all movement is necessarily leftwards; thus c-selection-related comp-to-spec movement will result in rigidly head-final structures (subject to the spell-out mechanism to be discussed in this chapter), and we can think of linearization movement (L-movement in BHR’s terms) as a movement-type parallel to A-, A’-, and head-movement. Aside from the empirically motivated objections to interpreting \(^\rightharpoonup\) as a PF diacritic noted above, it could therefore be argued that interpreting it as non-distinct from the syntactically visible movement diacritic is also conceptually preferable: being, by hypothesis, required to facilitate Internal Merge, \(^\rightharpoonup\) is arguably necessary to the computational system generating human language; as such, harnessing it to signal head-finality simply entails maximal exploitation of an available device. If this reasoning is correct, we see that there may in fact be more linearization information present in narrow syntax than is often thought to be compatible with the Strong Minimalist Thesis and considerations of “optimal design” (cf. Chomsky 1995: 335, Boeckx 2013: 57, and Berwick and Chomsky 2011: 15 for strong statements in favor of the notion that linearization be viewed as a property of syntactic structures that would optimally be imposed at PF, and cf. BHR 2009, 2010 for further discussion).
obligatorily only to “satellites” (i.e. externally merged specifiers). Thus, while the material in v’s complement domain (VP) may undergo Transfer to PF upon completion of a vP-phase, vP as a whole is not then renumerated; rather, it remains in the workspace and the CP phase (itself defined by a distinct lexical array) is simply constructed on top of the existing structure. By contrast, the derivation of a “satellite” takes place independently of the derivation of the main clausal structure (“in parallel”, as suggested by some authors—cf. Zwart 2011 for discussion). It will then be renumerated, i.e. added to a lexical array which defines a larger structure, within the context of which the atomic element will be a satellite. In section 9.5.3, we offer a slight revision of this model, based on Sheehan (2010b, forthcoming a), but the revision is not crucial at this point.

Following Sheehan (2010b, forthcoming a), we propose that, as well as applying to satellites, atomization applies more generally to DP islands, whether they are externally merged in a specifier or in a complement position. This is motivated by the well-known, but poorly understood, fact that the strong islandhood of most DPs is not geometrically conditioned: they are strong islands in both complement and specifier position (cf. Ross’s (1967) complex NP constraint). Note that this is a logical step to take if the one-way implication between atomization and strong islandhood in Uriagereka (1999) is strengthened to a bi-conditional. We return more specifically to a plausible explanation of the behaviour of DPs in section 9.5.3, restricting our attention for the moment to the question of the difference between CPs and DPs.

Given the above-mentioned (independently motivated) assumptions, a relevant distinction between DP and CP complements emerges. This is illustrated in (38) (outline font indicates an atomized structure):

(38) a. V^ DP V^ DP b. V^ CP V^ CP
   DP V^ CP C TP C TP

More accurately, Nunes and Uriagereka (2000) argue that all eventual specifiers are atomized prior to movement, but Sheehan (2010a,b) takes issue with this claim on empirical grounds, given that extraction from derived left branches is possible in many languages (consider, for example, derived subjects in English, cf. Kuno 1973). In fact, it appears that only underlying specifiers are necessarily atomized; derived specifiers are not. We return to this issue below.

We return to the analysis of non-opaque non-specific nominals in complement position in section 9.5.7.
As (38a) shows, DP complements of V are atomized and are therefore non-branching; CP, by contrast, (as a prototypical non-island) is not atomized, giving us a potential reason why the latter might behave differently from the former with respect to FOFC.

Crucially, however, it is not the branching property of CP per se which forces CP-extraposition in languages such as German: left-branching non-nominalized clauses in Japanese can surface preverbally, as illustrated in section 9.5.1, and these clauses do not behave like islands, indicating, in our terms, that they have not been atomized. Rather than branching structure, then, the consideration determining CP-extraposition in German-type languages would appear to be the fact that these CPs are right-branching. As such, the difference between Japanese and German/Dutch/Hindi/Persian appears to replicate the (by now) familiar FOFC effect (repeated here in descriptive terms):

(39) Harmonic and disharmonic combinations of V and CP

a. \( V' \)
   \( \quad \) CP
   \( \quad \) V
   \( \quad \) TP

b. \( V' \)
   \( \quad \) V
   \( \quad \) C
   \( \quad \) TP
e.g. Japanese
c. \( V' \)
   \( \quad \) CP
   \( \quad \) V
   \( \quad \) C
e.g. Hindi
d. \( V' \)
   \( \quad \) CP
   \( \quad \) CP
   \( \quad \) V

   (potentially) unattested\(^{31}\)
e.g. Bangla

Here we see that head-final CPs may surface either preverbally (39a) or postverbally (39c), whereas head-initial CPs only have the option of surfacing postverbally (39b); preverbal placement, as in (39d), leads to ungrammaticality.

It is important to note here, as Jack Hawkins (personal communication) points out, that (39c) is also a rare word order, possible only surfacing in variable VO/OV languages or as the result of optional extraposition. This fact is, however, independently explained by FOFC. Given that final complementizers are banned in VO languages because of FOFC (see section 9.2.2 above), it follows that the languages which have head-final CPs will always be at least partially OV, with V bearing \(^{\wedge}\). As such, it is expected that clausal complements in these languages will surface preverbally, all else being equal, with postverbal CPs being assigned a special/marked interpretation of some sort. Bayer’s (1999, 2001) discussion of head-final CP placement suggests that this expectation accurately reflects what we find in languages permitting this option.

To recapitulate what we have established in this section, then: it has been argued that FOFC appears to play a role in CP extraposition in OV languages, contrary to

\(^{31}\) We have found two potential counterexamples to this: Harar Oromo (Cushitic) and Akkadian (Semitic). Further research is required into these languages.
what we would expect under the analysis of FOFC in section 9.3. As a first step towards remedying this shortcoming, we have highlighted the potential relevance of the layered derivation model to explaining the crucially different behavior of CPs and DPs. In terms of a model of this type, the difference between DPs and CPs follows from the fact that the former enter the derivation as syntactic atoms, whereas the latter do not. In the following section, we show in more detail how the interaction between these two considerations allows us to understand the CP-placement facts with which this chapter is principally concerned.

9.5.3 Complement stranding (CS) and the strict LCA

Assuming Kayne’s (1994) definition of c-command, asymmetric c-command will always linearize a given structure as spec>head>comp. Note that this is only true, however, insofar as dominance regulates the linearization of branching specifiers which contain more than one category. Consider the following example:

\[
\text{the picture is of Mary}
\]

Although, the maximal category the asymmetrically c-commands and therefore precedes is (according to Kayne’s 1994 definitions), the same is not true for the terminals the, picture, or of Mary. The terminals contained in a branching specifier fail to enter into an asymmetric c-command relation with categories in the clausal spine (again assuming Kayne’s definition of c-command). As long as we stipulate that the linear position of terminals (e.g. the, picture, of Mary) depends on that of the dominating maximal category (e.g. maximal the), all is well. Uriagereka’s (1999) proposal, though, is that this stipulation should be challenged, and that what we will term “a strict LCA” should be adopted instead:

\[
\text{Strict LCA}
\]

\[
\text{If X asymmetrically c-commands Y, then X precedes Y.}^{32}
\]

---

32 As an anonymous reviewer points out, there is a way in which (41) is not actually all that strict, given that it is not a bi-directional statement. This is a crucial difference between the “strict LCA” given here and the “very strict LCA” adopted by Nunes and Uriagereka (2000), which is an iff statement. The reviewer also points out that the inclusion of “asymmetrically” in the definition implies that symmetrical structures will not lead to a crash at PF. We do not see this as an immediate problem as, according to standard
The version of the LCA in (41) makes no mention of dominance. This means that branching specifiers cannot be linearized with respect to the main clausal spine; rather, Uriagereka proposes, complex specifiers must be atomized in order for a derivation to converge at PF. As Uriagereka notes, this has the advantage of simplifying the LCA and deriving Huang’s (1982) CED without stipulation: specifiers and adjuncts will necessarily be atomic (because of the strict LCA and the need for total linear order) and will therefore always be strong islands. 

Sheehan (2010a,b) builds on this approach, but shows that Uriagereka’s (1999) and Nunes and Uriagereka’s (2000) approaches are too restrictive in ruling out all subextraction from complex left branches. Even in English, derived specifiers behave differently from externally merged specifiers, and, furthermore, the prediction is very obviously problematic in head-final languages such as Japanese and Turkish, which quite readily allow extraction from what on an LCA-perspective must be complex (left-branching) left branches. She points out that, given a certain interpretation of labeling, there is a principled reason why head-final and derived specifiers should behave differently to externally merged head-initial specifiers.33

To see how this works, we will first need to present an outline of the main assumptions of Sheehan’s labeling proposal (cf. Sheehan, forthcoming a, for detailed motivation of the assumptions underlying this proposal). She proposes that a radical interpretation of Bare Phrase Structure might in fact take labels to be copies of projecting terminal nodes, so that all projection leads to segment formation and all arguments are formally adjuncts. This takes Kayne’s claim that there is no specifier/adjunct distinction to its logical conclusion, so that projection to category is banned and there are always the same number of categories as terminals as a result. The following definitions, from Sheehan (forthcoming a) serve to restate Kayne’s category-based definitions of dominance and c-command in terms compatible with the copy theory of labeling:

assumptions, the failure to specify a single unambiguous linear order of all PF-interpretable terminals will independently lead to a crash, or rather to either (i) an additional application of spell-out, (ii) a morphological symmetry-breaking operation, or (iii) if all else fails, a PF-crash. As long as the LCA and the conjunction of precedence pairs remain the only way to map hierarchical structure to linear order, symmetry will be ruled out (unless it involves null elements). Note that the version of the LCA in (41), like Nunes and Uriagereka’s, differs from Kayne’s (1994) version in that it fails to require what Uriagereka (1999) terms “the induction step,” which states in relation to (41) that “the terminals dominated by X precede the terminals dominated by Y.” The omission of dominance from the definition is crucial, and we return to this issue in more detail below.

33 In fact, Sheehan (2010b, forthcoming a,b) gives up the idea that all head-finality is derived via movement for a variety of reasons. She nonetheless maintains that FOFC provides empirical evidence for a weaker version of the LCA which regulates disharmonic word-order combinations. A full explanation of how this would work is beyond the scope of this chapter.
(42) **Complete Dominance:** A category X completely dominates a category Y iff \( X \neq Y \) and the sum of the paths from every terminal segment of Y to the rest of the tree includes every non-terminal segment of X exactly once.

(43) **Partial category dominance:** a category X partially dominates a category Y iff \( X \neq Y \), the shortest path from every copy of Y to the root of the tree includes a copy of X, but X does not completely dominate Y.

(44) **C-command:** A c-commands B iff A and B are categories, \( A \neq B \), A does not partially dominate B and any category which completely dominates A also completely dominates B.\(^{34}\)

The outcome of these definitions is, unsurprisingly, that a specifier asymmetrically c-commands a head and a head asymmetrically c-commands its complement. The approach makes more nuanced predictions, though, where branching specifiers are concerned. Whereas some kinds of specifiers are predicted to be ruled out as unlinearizable, as in Uriagereka’s approach, others are predicted to be permitted (cf. Sheehan 2010a, forthcoming b).

Sheehan shows that if we combine the copy theory of labeling and the strict LCA with the assumption that scattered deletion (cf. Fanselow and Cavar 2000, Bošković 2001) is available as a last resort, then the prediction is that non-atomic head-initial phrases which move from a complement to a specifier position will strand their complements. In basic terms, this is due to the fact that the complement ZP contained in a complex specifier cannot be ordered with respect to the head, X (as neither c-commands the other); it is thus “unlinearizable” in its derived position (indicated by the symbol !! in (45)):

![Diagram](image)

According to Kayne’s LCA, ZP would be required to precede X by virtue of being dominated by Y, which asymmetrically c-commands X. Under the copy theory of labeling, the motivation for this “induction step” (which orders terminals based on the c-command domain of nodes which dominate them; cf. Uriagereka 1999) falls away, and so some special strategy is required to order ZP with respect to X in (45). With externally merged specifiers and DPs more generally, renumeration facilitates the necessary strategy: the specifier is first atomized in the manner outlined in section

\(^{34}\) Note that under these definitions a category can both dominate and c-command another.
9.5.2, before being integrated with the clausal spine; this is illustrated in (46) (as before, outline font indicates an atomized structure):

(46) 

\[ \begin{array}{c}
  \text{DP} \\
  \text{v} \\
  \text{v} \\
  \text{VP} \\
\end{array} \]

With derived specifiers, which were externally merged as complements (and are not inherently strong islands; see the discussion in section 9.4.2 above and also in section 9.5.7 below), the most economical linearization strategy is argued to be scattered deletion, as in (47), as it allows linearization of all terminal nodes without the need for multiple spell-out:

(47) 

\[ \begin{array}{c}
  \text{T} \\
  \text{V} \\
  \text{Obj} \\
  \text{V} \\
  \text{V} \\
  \text{T} \\
  \text{V} \\
  \text{Obj} \\
\end{array} \]

The scattered deletion proposed here is highly constrained and emerges as an optimal linearization solution as a consequence of the way in which the labeling system outlined above interacts with (41). In the specific case of (47), the fact that projection involves copying of the projecting head as the label of a phrase leads to an important asymmetry between the head V and its complement, Obj: if the maximal projection V is a copy of the head V, it will be able to determine the linear position of the terminal V (since the maximal projection asymmetrically c-commands T), while simultaneously not determining that of Obj (since Obj is “too embedded” to c-command out of its containing VP). In the absence of the induction step, Obj will then not be orderable with respect to T: no asymmetric c-command obtains between (the derived copy of) Obj and T and, crucially, no ordering between Obj and T is defined by the concatenation of independently defined precedence pairs—these specify that V>Obj and V>T, giving rise to two potential orders: V>Obj>T and V>T>Obj. Assuming that PF requires a single unambiguous linear ordering of terminals, it is expected that underspecified orders will be avoided at all costs. For this reason, Sheehan proposes, “complement stranding/extraposition” applies in such contexts, i.e. PF is forced to target the first-merged copy of Obj, rather than the derived copy contained in SpecTP; since the lower copy is asymmetrically c-commanded by T, the total linear order V>T>Obj results, giving a linearizable structure.
As an anonymous reviewer points out, it is important to stress here why harmonically head-final derived specifiers differ from head-initial specifiers in such contexts, given that that they do not necessitate extraposition. In Japanese, for example, it has been shown that head-final non-nominalized CPs surface in the preverbal position unproblematically (cf. (27)). The crucial difference in such cases is that the concatenation of independent precedence pairs serves to define a single unambiguous order of terminals without either atomization or complement stranding being necessary. Consider (48) by way of illustration (as above, ^ signals the fact that the heads in question have triggered comp-to-spec movement, i.e. movement of their c-selected complements into their specifiers):

In (48), the derived maximal copy of C^ (located in Spec-V^) asymmetrically c-commands V^, and the derived maximal copy of T^ (located in Spec-C^), asymmetrically c-commands C^.

The relation of the above to FOFC should be evident. If we adopt BHR’s approach to head-finality, head-final phrases involve local comp-to-spec movement. Sheehan’s analysis therefore leads us to expect that a subset of these comp-to-spec movements—those involving right-branching phrases—will give rise to CS. In addition to the RM-imposed restrictions on the distribution of ^, a further class of FOFC-effects therefore emerges from the obligatory stranding of complements in contexts such as (45).
9.5.4 Complement CPs and complement stranding

Sheehan (2009) claims that the obligatory stranding in (45) can account for many of the FOFC effects reported by BHR. In fact, however, neither the CS nor the RM account is independently restrictive enough with respect to the word orders they predict. As we have seen, the RM account does not facilitate a straightforward account of why CP-extraposition occurs to prevent *[[C TP] V] structures. The CS account, on the other hand, predicts that adjacent heads should always and very readily have the possibility of inverting, something that careful work by Cinque (1999, 2005) has shown not to be the case. In other words, CS is restricted by the distribution of ^ features. In relation to CP placement, the relevant non-occurring structure (*[[C TP] V [C TP]]) is illustrated in (49):

(49)

The non-occurrence of these structures seems to be attributable to additional, as yet not fully understood, restrictions on scattered deletion, meaning that scattered deletion is blocked where the stranded complement is not a complete extended projection. Thus extraposition of this kind typically targets PP, CP and, in more restricted cases, DP.

Here, we suggest that a combination of the RM and CS approaches is necessary in order to provide an explanatory account of the observed FOFC-related effects.

Crucially, both approaches rely on the LCA as a fundamental principle of linearization and assume that (c-selection-related) ^ triggers local comp-to-spec movement (cf. also Holmberg 2000, Julien 2002). We therefore propose that c-selection-related ^ distribution is governed by RM in the manner described in section 9.3, with ^ necessarily constituting a narrow syntactic movement trigger. In combination with CS and the

35 BHR account for extraposition via “radical spell-out.” claiming that material that is transferred to the interfaces (spelled out) is radically removed from the computation, not only being opaque for further probing, but in fact being shipped wholesale to the interfaces, with the result that it is also not available for further movement operations. See Biberauer and Roberts (2005) for initial discussion of this proposal in relation to the absence of SVOAux orderings in earlier English.

36 Superficially, West Germanic infinitival structures in which only part of the infinitival complement surfaces postverbally (in “extraposed” position) would seem to instantiate counter examples to the generalization that scattered deletion can only target complete extended projections. Verb raising and verb-projection raising are two cases in point (cf. Evers 1975, Haegeman and van Riemsdijk 1986). Worth noting about these structures, however, is that they necessarily involve so-called restructuring verbs, which seem to involve matrix verbs which trigger leftward movement not as a consequence of the c-selection-related ^ associated with lexical verbs more generally, but instead via an Agree-related ^ (cf. Biberauer 2008, and BHR 2009, 2010 for discussion). We leave the specifics of this matter for future research.
renumeration approach to DPs outlined above, we may then understand the extraposition facts by appealing to Rosenbaum’s (1967) idea that embedded CPs can be concealed nominals in certain cases (cf. also Biberauer 2008, Biberauer and Roberts 2008). Let us consider how this works.

Where $V^\wedge$ selects an atomized DP, the result will clearly be an OV order, as shown in (46) above. Where $V^\wedge$ selects a non-atomic CP, however, the result should, as things stand, involve the stranding of TP, as illustrated in (49) above. This apparent prediction obviously fails to hold, suggesting that the structure of embedded CPs may in fact be more complex than has been assumed up to now (cf. also Biberauer 2008 and Biberauer and Roberts 2008).

Recent work on CP-complementation (cf. Kayne 2008 and Arsenijevic 2009 and references cited therein) independently argues that complement CPs are in fact disguised nominals (strikingly, the reasoning leading to this conclusion is based on entirely unrelated considerations, producing a remarkable convergence on Rosenbaum’s earlier ideas). Building on this work, we propose that CP arguments are optionally selected by a particular species of little $n$ when selected as arguments.37 It is worth noting that the type of $n$ we have in mind here is likely to differ from the n(s) heading the lower nominal phase (cf. Julien 2005). As Myler’s (2009b) discussion of morphological phenomena clearly shows, “verbalizing” $v$ cannot be the same entity as the various $v$s determining the argument structure of the lower phase, and the same may well be true of “nominalizing” $n$. We leave the specifics of this very interesting question to future research. Also worth noting in connection with the nature of the $n$ under which clauses may be embedded is that if strong islandhood is a direct result of atomization, it is predicted that nominalization of a clause will not necessarily be sufficient to render it opaque for subextraction.38 Thus, it is, for example, conceivable that clauses could be nominal without being strong islands as long as the trigger for atomization (non-defective D, by hypothesis) is absent. Indeed rich nominalizing languages show rather clearly that embedded nominalized clauses are not necessarily strong islands. Thus Myler (2009b), for example, shows that Quechua nominalized clauses permit extraction. Consider the following example from Imbabura Quechua:

(50) Ima- ta- taj ya- ngu [Juan ima randi -shka-ta]
    what-      -WH think-2     John buy-      NOM-ACC

‘What do you think that John bought?’ [Cole 1982: 21]

37 From this perspective, the oft-discussed “hybrid” qualities of CP-complements (cf. Holmberg 1986, Rizzi 1990b, and the references cited in Vikner 1995 for discussion) may be understood as an effect of the fact that CPs can be selected by a nominalizing head $n$. If this idea is on the right track, it may be that so-called extreme nominalizing languages are less different from so-called extreme finite languages than is generally thought (cf. Givon 2006 for recent discussion).

38 In section 9.5.6, we discuss examples of subextraction from nominals and show that their behavior can also be explained by this analysis.
The fact that extraposed CPs are not strong islands (cf. (22)–(23) above) is therefore not incompatible with them being concealed nominals. Furthermore, the proposal that CPs are in fact embedded under a nominal layer allows us to understand the dass-co-ordination facts noted in (30b) above: if these dass-clauses are in fact embedded in nominals, the fact that they may be co-ordinated with regular nominals is readily explicable.

Let us now consider how the presence of nominal superstructure correctly predicts the placement of complement CPs. According to Sheehan (2009a, 2010b), it is the entire complement of the projecting head of a non-atomized phrase that creates a linearization problem when this phrase occupies a specifier position. For this reason, the whole complement of the projecting head is predicted to be stranded in such contexts, i.e. TP in the case of a head-initial CP. If little \( n \), rather than C, is the projecting head in this case, then the CP extraposition facts are as expected, however. This is illustrated in (51) below:

\[
(51) \quad \begin{array}{c}
V^\wedge \\
\quad n \\
\quad \quad V^\wedge \\
\quad \quad \quad CP
\end{array}
\]

Example (51) indicates that the approach proposed in this section is able to account for the “extraposition” of head-initial CPs that we observe in OV languages with head-initial CPs. At first sight, it might, however, seem to run into the same empirical difficulty as the PF-based approach discussed in section 9.5.2 above (cf. also the discussion surrounding (25) in section 9.4.2): if the head-initial CP is spelled out in its first-merged position within the complement of \( V^\wedge \), the analysis outlined so far might lead us to expect unattested \([V [CP]]Aux\) structures in cases where the matrix clause features a compound tense. Consider (52) below by way of illustration:

\[
(52) \quad \begin{array}{c}
!! \quad \begin{array}{c}
Aux^\wedge \\
\quad V^\wedge \\
\quad \quad n \\
\quad \quad \quad \begin{array}{c}
\quad \quad V^\wedge \\
\quad \quad \quad \begin{array}{c}
\quad \quad \quad \begin{array}{c}
\quad \quad \quad \quad CP
\end{array}
\end{array}
\end{array}
\end{array}
\end{array}
\]

Expected order: \( n \ V \ CP \ Aux \)
As (52) shows, if CP is spelled out in its first-merge position within the moved VP in structures featuring an auxiliary, an unattested structure would result. Importantly, however, spelling out the first-merged CP within the specifier of Aux^ is not possible under the system proposed here since neither of the CP copies located within this specifier can be ordered with respect to Aux^\. Consequently, a lower copy, which does not pose a linearization problem, needs to be targeted for spell-out. Since none of the material within the complement of Aux^ will in fact be overtly realized (n) or spelled out (the lower copies of V^) in structures like (52), it may in fact be the case that the CP located in Spec-V^KP can be spelled out in this case; alternatively, the lowest CP-copy may become the ultimately realized copy. Regardless of which option is taken, what is clear is that the analysis proposed here crucially makes a different prediction from that discussed in section 9.4.2; specifically, the prediction is that head-initial CPs will always be spelled out finally, regardless of the size of the verbal cluster since the linearization problem highlighted by the CS approach arises with each ^-bearing verbal head.

An interesting implication of the analysis regards the connection between the headedness of n, C, N and V. In order for a CP to be extraposed, it is necessary that it be selected by a head-initial n (cf. Biberauer 2008, Biberauer and Roberts 2008). If, on the other hand, n bears ^, the prediction is the Japanese order TP-C-n-V, since the structure in question will be consistently head-final, as in (48) above. The important thing is that, given the logic of the RM-based ^-distribution system, n has functional status and so cannot start a new extended projection. It follows that n can only bear ^ in a language in which C (and all heads down to V) also bear ^. For this reason Bengali head-final CPs can occur in either preverbal or (more marginally) postverbal position (as they can be selected by either n^o or n), whereas the opposite is not true with head-initial CPs: a head-initial CP can only ever be selected by a head-initial n, thus giving rise to obligatory extraposition. A final interesting question regards the connection between the directionality of n and N. It might also be expected that n can only bear ^ in languages in which N bears ^. This would mean that extraposition would only occur in languages with head-initial NPs. While this prediction clearly holds in German and Dutch, it is not so clear that it holds for Hindi and Turkish. We leave this matter for future research.

The upshot of the preceding discussion, then, is that FOFC stems from two interacting factors. On the one hand, the distribution of c-selection-related ^ features within an extended projection is regulated by RM. On the other hand, layered derivations and the strict LCA explain why FOFC appears to hold intermittently across extended projections (e.g. between V and CP, but not between V and DP). The diachronic pathways discussed in section 9.2 are very clearly linked to the RM-imposed constraints on ^ feature distribution, whereas the CP extraposition facts result from the strict LCA and complement stranding. Crucially, the presence of the nominal which triggers CP-extraposition in languages like German and Hindi is forced to be present as a FOFC-compliance strategy (to avoid an unlinearizable structure, which cannot be rescued via CS).
9.5.5 The significance of deficiency

What we have proposed in the preceding section is that head-initial finite CP-complements in OV languages like German are in fact embedded within a nominal superstructure. Since the nominals in question are head-initial and are externally merged as branching structures, they trigger extraposition in accordance with Sheehan’s CS proposals. The question that now arises is how to account for the non-finite CP-placement facts highlighted in section 9.5.1 above.

Recall that the placement options available to clauses of this type appear to be regulated by headedness and “deficiency:” specifically, head-final structures may surface both pre- and postverbally if they are of sufficient “size.” The “size” question is also highlighted in Wurmbrand (2001: 294), in the following terms (cf. also Hinterhölzl 2005: 15):

the “bigger” a complement . . . , the more likely it is to extrapose; the “smaller” the complement . . . . , the more likely it is to occur in intraposed position.

In the context of the analysis we are proposing here, the “deficiency” aspect of clausal placement can be understood as the consequence of the fact that defective clauses do not constitute complete extended projections. If the latter is a condition for atomization (see also section 9.5.6 below), we would then expect that defective complements will always be merged with their selecting verbs as branching structures. Since these branching structures are, moreover, head-final in languages of the type we have been considering in this paper, we would expect them to be able to undergo comp-to-spec movement under the influence of their selecting verb’s ^-feature without creating any linearization difficulties: as we saw in section 9.5.3, left-branching specifiers do not pose a linearization challenge in the context of the kind of LCA-based linearization mechanism assumed here. The preverbal placement of reduced head-final clausal complements is therefore accounted for (cf. (29b) and (33a)).

Still to be accounted for are cases like (29a), which (superficially) involve the extraposition of a full control CP. Two possibilities suggest themselves here: first, that these clauses, like the finite CPs discussed in the preceding section, are embedded under a nominal. This is not implausible if one considers the fact that the verbs selecting complements of this type may also select for nominals (cf. Er hat mir diese Zeitung empfohlen and Er hat es mir empfohlen—literally: he has me this newspaper recommended, and he has it me recommended, i.e. “he recommended this newspaper/it to me”). Alternatively, it could be the case that these clauses differ from those featuring an overt complementizer—never an option in non-purpose control complements in German—in lacking a nominal superstructure and being selected by the restructuring verb, which, if note 36 is on the right track, may in fact be associated with an Agree-related ^ rather than a c-selection-related (i.e. comp-to-spec-triggering)
^ If this is correct, the difference between (29a) and (29b) may in fact not be that (29a) involves an extraposed CP, while (29b) does not, but, instead, that only the Agreed-with PRO has undergone raising in (29a), while the entire CP is pied-piped in the case of (29b). See Biberauer and Roberts (2008) for an account along these lines.

To conclude this section, then. We have seen that our proposal is able to account for the discrepancy between “full” and “deficient” clausal complements in systems like West Germanic: since the latter instantiate partial clausal projections which are, moreover, head-final, we expect them to pose no linearization difficulties in the context of the system proposed here, with the result that preverbal realization is expected in all cases. What superficially appears to be partial complement postposing of the type discussed in note 36 and also the optional postverbal realization of control complements may in fact involve structures in which only a part of the infinitival complement has undergone movement into the higher clause, a matter we leave for fuller future investigation.

9.5.6 Phase theory and atomization

We return now to two aspects of the analysis which have, until now, been represented in simplified form: the formal status (atomized or not) of phrases which are externally merged as specifiers and the structure of nominals. This section deals with the former of these matters, while the following section addresses the latter.

Up to now, we have simply been assuming that a specifier must be atomized prior to being externally merged. As we have seen, Uriagereka (1999) and Nunes and Uriagereka (2000) motivate this model on PF-interface grounds. An apparent problem here is, however, that it appears to involve look-ahead: a branching specifier does not actually become a problem until it reaches the PF-interface. From this perspective, the motivation to atomize it before it has even been externally merged emerges as contra the spirit of a derivational approach.

Sheehan (2010a) proposes a way to resolve this problem, which also creates the possibility of a much needed link between Chomsky’s notion of phase and the layered derivation perspective outlined in section 9.5.2. The idea would be that narrow syntactic structures have no order, but keep track of cumulative c-command relations, potentially for independent reasons (e.g. in order to regulate narrow syntactic operations such as agreement), on a Merge by Merge basis. As long as non-branching heads/specifiers, or derived specifiers are merged, each new derivational step will simply add one new c-command relation; when a branching specifier is externally merged, however, this gives rise to underspecification of c-command relations: all material except the projecting head fails to either c-command or be c-commanded by the portion of the clausal spine constructed before that point. The proposal we would
like to make is that this situation causes the narrow syntax to transfer the whole structure to PF/LF so that this issue can be resolved. What this means, in effect, is that phases are phrases containing a non-atomized externally merged branching specifier. Assuming that nominals always entail at least some functional structure and that pronouns are in fact complex entities of the type proposed in accounts like Cardinaletti and Starke (1999), Harley and Ritter (2002), and Neeleman and Szendröi (2007), among others, this means that vP will be a phase in transitive/unergative, but not unaccusative contexts. Furthermore, CP will not be a phase, unless it contains an externally merged phrase of some description (e.g. a non-movement derived topic), or it represents the portion of structure generated by merger of the final remaining elements in a given numeration. At the point of Transfer, PF establishes an order between all terminals. Where the externally merged specifier is head-initial, this necessitates atomization. Where the externally merged specifier is harmonically head-final, no such atomization is necessary. Once an order has been established, there is a final chance for material to escape to the phase edge via A-bar movement, before the material in the phase is removed from the derivation (following the system in Chomsky 2001). This means that at the point at which movement to the phase edge is permitted, head-initial specifiers have already been atomized, and hence no extraction from them will be possible. They can, however, be moved wholesale to the phase edge, subject to the restrictions outlined in Ko (2005). In the case of harmonically head-final externally merged specifiers, subextraction to the phase edge will be possible (cf. Sheehan 2010a for empirical support for this difference).

The attractive aspect of this approach is that it provides us with a testable, conceptually motivated definition of phasehood. We leave a thorough exploration of these predictions to future work. In these terms, though, as atomization of DP is not triggered by the presence of an externally merged specifier, DP is not a “phase.” The question remains, then, why DPs are so often atomized. Of course, externally merged specifier DPs are atomized as part of Transfer to the PF/LF interfaces for the reasons discussed above. Complement DPs, however, are atomized for independent reasons, plausibly connected to saturation or completeness. In this sense, derivations are layered and both atomization and cyclic Transfer (phases) are triggered where two independent derivations are combined.

39 If this proposal is correct, recent suggestions that the computational system may exhibit an “LF bias” (cf. Berwick and Chomsky 2011, Boeckx 2011, and Richards 2008) would need to be re-evaluated. Given the plausibility and, arguably, desirability from a minimalist perspective of viewing movement-triggering \( \wedge \) as additionally also serving a linearization-related role (cf. section 9.5.2 and note 28 above), it is not so clear that the view that PF must “make do,” while LF dictates should be accepted as the null hypothesis.
9.5.7 Subextraction from nominals in OV Germanic

In our comparison of DPs and CPs, we have abstracted away from the well-known fact that not all nominals are strong islands. Rather, it would appear that non-specific indefinite nominals embedded under certain predicates allow subextraction:

(53) Who have you got a picture of?

Crucially, this effect also holds in OV Germanic languages:

(54) Über Syntax hat Sarah sich ein Buch ausgeliehen
    about syntax has Sarah herself a book borrowed
    ‘About syntax Sarah borrowed herself a book’

[German, De Kuthy 2002: 2]

This appears, at first blush, problematic for the claim that DPs are atomic in OV Germanic. Note, however, that these non-specific indefinites are precisely the objects which permit complement extraposition:

(55) Sie hat öfters ein Buch ausgeliehen über Syntax
    she has frequently a book borrowed about syntax
    ‘She frequently borrowed a (non-specific) book about syntax.’

The fact that ein Buch follows the VP adverbial öfters clearly shows that this nominal is located in a non-scrambled position (cf. Diesing 1992 and much subsequent work). Strikingly, complement extraposition is impossible when the nominal undergoes scrambling, i.e. when it receives a specific interpretation:

(56) *Sie hat ein Buch öfters ausgeliehen über Syntax
    she has one book frequently borrowed about syntax

Furthermore, only (55)-type structures permit extraction:

(57) Über was hat Sarah ein Buch (*öfters) ausgeliehen?
    about what has Sarah a book (frequently) borrowed
    ‘What has Sarah borrowed a book about?’

These examples clearly show that we have to distinguish between specific and non-specific nominals and their extraction and extraposition possibilities in German. Referring again to the notion of saturation/completeness mentioned at the end of the previous section, we may propose that specific nominals are always merged as atoms, thereby precluding complement extraposition (and permitting scrambling in West Germanic), while non-specific nominals can be merged as either atoms or branching structures, with the latter therefore raising the linearization issues discussed in previous sections. Note that for the PP complement to be stranded, it follows that indefinite articles cannot be projecting heads in English/German/Afrikaans: if this
were the case it would lead us to expect preverbal spell-out of just the indefinite article, with the nominal it selects being spelled out postverbally. Rather, then, indefinite articles in these languages must be specifiers, and it emerges that Abney’s analysis of indefinite articles (cf. note 8) cannot hold universally. See Sheehan (2009) for further discussion of the implications of stranding for the structure of the noun phrase.

What we have seen in this section, then, is that nominals are not universally atomized, with the result that we expect subextraction to be possible from non-atomized structures. Both atomized and non-atomized nominals are, nevertheless, expected to exhibit systematically different behavior from full clausal complements in that at least a part of a nominal complement will be overtly spelled out preverbally.

9.6 Conclusion

The purpose of this chapter has been to argue that a conceptualization of linearization based on a version of the LCA is required to account for certain word-order asymmetries in natural language. More specifically, it has been proposed, following previous work by Biberauer, Holmberg, and Roberts, that head-finality is regulated by the presence of movement-triggering $^\wedge$ features associated with lexical heads. This feature can spread through extended projections, subject to RM, and must necessarily be understood to trigger narrow syntax-internal movement: the word-order asymmetries which arise as a consequence of the Final-over-Final Constraint (FOFC) cannot be accounted for if the ordering between heads is purely a PF matter. This obviously has important consequences for our perspective on the way in which syntactic structure is constructed.

Our discussion has also highlighted the fact that, as it stands, BHR’s account cannot readily explain the full range of observed FOFC-effects, in particular the discrepancy between the placement of nominals and full clausal complements in non-rigid OV languages: the fact that head-initial CPs are necessarily postposed when they have not undergone scrambling superficially appears to be a FOFC effect, but cannot be accounted for straightforwardly by the above-mentioned constraints on $^\wedge$ distribution. Adopting a revised version of the LCA, akin to that proposed by Uriagereka (1999), however, facilitates insight into this discrepancy. More specifically, we have shown that the relevant patterns can be understood if we adopt Sheehan’s proposal that the linearization difficulties arising from the revised LCA may be resolved via complement stranding at PF. Combining the feature distribution and complement stranding ideas, we therefore obtain a wider range of predictions regarding the structures that are expected to be ruled in and ruled out as a consequence of FOFC. More specifically, we are able to understand FOFC-violating orders not only within extended projections, but also across certain extended projections.
A particular view of how syntactic derivations proceed and of how narrow syntax interfaces with PF and LF has also emerged during the course of the discussion. We have also proposed that phases can be characterized as the points at which two branching phrases are merged. This proposal obviously has numerous consequences for the way in which notions like “lexical array,” “phase,” and “spell-out domain” are interpreted, which we cannot do justice to here.

More generally, numerous issues, of course, remain open. Some of the theoretical ones have been noted here, but there are also empirical questions which require more systematic consideration, notably reduced infinitival clauses in restructuring-type structures and PPs. What we hope to have achieved here, however, is to present and motivate a new approach to structure building and linearization which addresses both long-standing and more recently uncovered challenges in accounting for the distribution of disharmonic structures, and which opens up the possibility of understanding a range of further structures of this type.
10

Don’t Feed Your Movements When You Shift Your Objects*

ŽELJKO BOŠKOVIĆ

10.1 Introduction

Until recently, it has been standardly assumed that constructions like (1) involve A-movement of the wh-phrase from the object position to the subject position, followed by wh-movement to SpecCP.¹

(1) Who was arrested?

However, Chomsky (2008) proposes a new treatment of such constructions (see also Hiraiwa 2005). According to Chomsky, instead of A-movement feeding wh-movement, (1) involves two separate movements from the deep object position. Roughly, who moves to SpecTP from the object position, and it also moves to SpecCP from the object position, with the two movements proceeding in parallel and with only the highest copy pronounced. The parallel movement hypothesis has a significant impact on the way structure building proceeds. For one thing, the change in the timing of movement it introduces eliminates A-A’ movement feeding in examples like (1). In this chapter I provide additional evidence for the no feeding analysis and show that the analysis provides a tool for teasing apart different analyses of object shift in Icelandic. I will show this in section 10.3 of the chapter. In section 10.2 I go over several arguments for the no feeding analysis, showing that the analysis has considerable empirical motivation.²

* I thank anonymous reviewers for helpful comments and Halldór Sigurðsson for help with the data.

¹ The last step has been somewhat controversial. However, recent literature (see An 2007a, Boeckx 2003, and Pesetsky and Torrego 2001) provides very strong evidence for the existence of vacuous wh-movement in local subject questions (see the discussion below for another argument to this effect). At any rate, what is important for our purposes is that the example is standardly assumed to involve movement to SpecTP.

² Following Chomsky (2008), I will implement the no feeding analysis in terms of parallel movement. The reader is referred to Bošković (2008) for an alternative way of implementing the no feeding analysis as well as evidence in favor of the alternative. I will not be comparing the two analyses in this Chapter.
10.2 Don’t feed your movements

McCloskey (2000) shows that, in contrast to standard English, West Ulster English (WUE) allows quantifier float (Q-float) under wh-movement based on examples like (2).³

(2) What do you think [CP (all) that he’ll say [CP (all) that we should buy (all)]]?

Consider now the following examples from McCloskey (2000).

(3) Who i was arrested all ti in Duke Street?

(4) *They i were arrested all ti last night.

Although WUE allows (3) it behaves like Standard English in that it disallows (4). Notice first that the contrast between (3) and (4) provides evidence that local subject questions do involve wh-movement: if who in (3) were to remain in SpecTP, we could not make a distinction between this example and (4). However, this cannot be the end of the story. If who were to move to SpecTP prior to moving to SpecCP in (3) it seems that it would still be impossible to account for the grammaticality of the construction, given that (4) is unacceptable. When it comes to the floating of all, (3) and (4) would be identical: all would be stranded by movement from the object position to SpecTP in both examples. To make a difference between the two examples, McCloskey (2000) (see also Bošković 2004a) therefore suggests that the wh-phrase in (3) moves directly to SpecCP, the underlying assumption being that wh-movement, but not movement to SpecTP, can float all in the position in question (see Bošković 2004a and Fitzpatrick 2006 for different accounts of why this is the case, an issue that goes beyond the scope of this chapter). A question that arises under this analysis is how the standard requirement that the SpecTP position be filled in English is satisfied in (3) if who moves directly to SpecCP. Before discussing McCloskey’s answer to the question (for an alternative analysis see Bošković 2004a), let us see how he prevents who from moving to SpecTP in (3). McCloskey suggests that Q-float involves a step in which the NP the Q modifies moves to SpecDP, the Q being located in D. The movement yields the order NP Q within the DP. When the NP in SpecDP is a wh-phrase, D acquires the +wh-feature from it so that SpecDP counts as an A¹-position. The wh-phrase (who in (3)) then cannot move to SpecTP, since this would involve improper movement. Rather, it moves directly to SpecCP. How is the requirement that forces overt movement to SpecTP

³ In what follows, I will be assuming Sportiche’s (1988) stranding analysis of Q-float. The reader should bear this in mind. Under Sportiche’s analysis, what and all in (2) start as a constituent. What then moves away stranding all. This means that there is a trace of what next to all, with which all forms a constituent. Below, for ease of exposition I will often say that a floating quantifier modifies/cannot modify a trace in this scenario. However, the reader should not attach deep meaning to the term modify here. This simply means that movement that leaves behind the trace in question can/cannot strand a quantifier, whatever the reason for this is (the quantifier always forms a constituent with a trace under Sportiche’s account).
satisfied in (3)? McCloskey suggests that overt movement is preferable to Agree, the mechanism which allows feature-checking at a distance without actual movement. However, when a requirement cannot be satisfied without a violation through movement, satisfying it through Agree, i.e. without movement, becomes possible. In the case in question, features of T cannot be satisfied through movement since this would result in improper movement. Therefore, features of T can be satisfied without movement via Agree. (It is implied either that the EPP is a featural requirement or that there is no EPP. The analysis is inconsistent with Chomsky’s 2001 filled Spec requirement view of the EPP.)

It seems that under this analysis we should always be able to get around a violation caused by overt movement by doing Agree. For example, we should be able to get around the Left Branch Condition effect and the that-trace effect, where overt movement causes a violation (see (5)), by doing feature checking via Agree, i.e. without movement (which means leaving the relevant element in situ, as in (6), which is impossible).

(5) a. *Whose did you see t₁ books?  
   b. *Who₁ do you think that t₁ left?
(6) a. *You saw whose books?  
   b. *You think that who left?

As noted in Hiraiwa (2005), the parallel movement hypothesis allows us to preserve McCloskey’s direct movement to SpecCP analysis of (3), which is necessary to make a distinction between (3) and (4), and at the same time easily answers the question of how the standard filled SpecTP requirement is satisfied in (3) (which we saw above ended up raising a problem for McCloskey’s analysis). Under the parallel movement analysis, who in (3) moves directly to SpecCP, as desired, but it also moves to SpecTP, so that the filled SpecTP requirement is satisfied. Most importantly, since there is no feeding relation between the A and the A’ movement in question, all in (3) is not floated under movement to SpecTP, which must be disallowed given the ungrammaticality of (4). The major accomplishment of the parallel movement analysis is that it enables us to fill the lower A-position in spite of the absence of a feeding relation between the movement of the NP that fills this position and the movement of this NP to a higher A’-position.

Chomsky (2008) observes that there is a difference in the grammaticality status between extraction out of subjects that are generated as external arguments and subjects that are generated in object position, and shows that the difference can be accounted for under the parallel movement analysis of such examples. Under this analysis, wh-movement takes place directly from the θ-position of the relevant arguments. Chomsky then capitalizes on the fact that (7b), but not (7a), involves wh-movement from object position, which we independently know is allowed (7c).
(7)  a. *It was the car (not the truck) of which the driver caused a scandal.
   b. It was the car (not the truck) of which the driver was found.
   c. It was the car (not the truck) of which they found the driver. (Chomsky 2008)

Chomsky (2008) observes that certain Icelandic data discussed by Holmberg and Hróarsdóttir (2003) (see also Hiraiwa 2005) also provide evidence for the parallel movement analysis. Consider (8).

(8)  a. Það virðist/*virðast einhverjum manni [hestarnir vera seinir]
   EXPL seems/seem some man.DAT the-horses.NOM be slow
   ’It seems to some man that the horses are slow.’
   b. Mér virðast t NP [hestarnir vera seinir]
      me.DAT seem.PL the-horses.NOM be slow
   c. Hvaða manni veist þú að virðist/*virðast twh [hestarnir vera seinir]
      which man.DAT know you that seems/seem the-horses be slow
      ’To which man do you know that the horses seem to be slow?’
      (Holmberg and Hróarsdóttir 2003)
   d. Hverjum mundi/??mundu hafa virst twh [hestarnir vera seinir]
      who.DAT would.3SG/3PL would.3PL have seemed the-horses.NOM be slow
      ’To whom would it have seemed that the horses are slow?’
      (Nomura 2005)

(8a) shows that lexical experiencers block agreement with a lower nominative NP (the verb must have the default 3SG form). An NP-trace does not induce a blocking effect, as shown by (8b). Holmberg and Hróarsdóttir (2003) interpret examples like (8c–d) as indicating that a wh-trace does induce a blocking effect. Notice, however, that if the experiencer in (8c–d) were to move to SpecTP before undergoing wh-movement, the intervening element would be an NP-trace. (8c–d) should then pattern with (8b). To account for (8c–d), Holmberg and Hróarsdóttir (2003) suggest that the wh-phrase does not, in fact, cannot, undergo movement to SpecTP in (8c–d). Rather, it must move directly to SpecCP, hence the blocking effect. (The intervening trace is then a wh-trace.) As noted in Chomsky (2008), the parallel movement hypothesis can be straightforwardly applied to the Icelandic data under consideration. The wh-phrase moves to both SpecTP (satisfying the EPP) and SpecCP from its base position in (8c–d). As a result, the trace left in the base position must count as a wh- as well as an A-trace. It’s A’-property apparently suffices to induce a blocking effect.

4 Not all speakers share the judgments reported by Holmberg and Hróarsdóttir (2003); see Sigurðsson and Holmberg (in press) for relevant discussion.
5 The following discussion slightly modifies Chomsky’s analysis. What matters for Chomsky is that only a part of the experiencer A-chain intervenes between T and the nominative NP in (8b), while the complete (trivial) experiencer A-chain intervenes in (8c–d).
10.3 Icelandic object shift

I now turn to object shift in Icelandic. Consider the following data involving Q-

\[(9)\] a. Ég las bækurnar ekki allar.
I read the-books not all

\[\]

b. *bækurnar sem Jón keypti ekki allar
the-books that Jon bought not all

\[‘the books which Jon didn’t buy all of’ (Déprez 1989)\]

(9a) is an example involving object shift, which shows Q-

float is possible under object

shift. On the other hand, (9b) shows that, in contrast to WUE, Icelandic does not

allow Q-

float under movement to SpecCP (more precisely relativization in the case at

hand. Note that WUE allows Q-

float under relativization, see Fitzpatrick 2006.)

Consider now (9b) more closely. Nomura (2005) observes that examples like (9b)

involve an object shift context. We would then expect the relevant NP to be able to

undergo object shift prior to undergoing relativization. Given that object shift

licenses Q-

float, Q-

float should then be licensed in (9b). In other words, under the

standard analysis the ungrammaticality of (9b) is surprising since the quantifier is

floated under the object shift movement (which then feeds relativization), just as in

(9a). The data under consideration seem to be an obvious candidate for a parallel

movement analysis. Under this analysis, wh-movement and object shift in (9b) take

place from the same position, in particular, the position in which all is floated—there

is no feeding relation between the two. If object shift does not feed wh-movement, the

above problem can be resolved since the relevant trace is a trace of both wh-

movement and the object shift movement. We can then easily account for (9b) if,

as in most languages, a floating quantifier in Icelandic cannot modify a trace that

even ambiguously counts as a wh-trace (see here fn. 3; in other words, if a trace is

created by movement to SpecCP, it cannot be modified by a floating quantifier).7

---

6 For original discussion, see Diesing (1996), where object shift is roughly characterized by definiteness/

non-focus interpretation. An anonymous reviewer notes that it is actually not completely clear that (9b)

satisfies the semantic conditions on object shift under the Vergnaud (1974)/Kayne (1994) raising analysis of

relativization, which, however, I do not adopt here. At any rate, this potentially interfering factor does not

arise with respect to the topicalization example in (10) below.

Note that there is a debate in the literature regarding the final landing site of Icelandic object shift. At this

point I am using the terms object shift and object position neutrally, without committing myself to a

particular analysis. The issue will be discussed in detail below.

7 I leave open here what is responsible for the apparent cross-linguistic variation with respect to the

possibility of a floating quantifier modifying a trace left behind by wh-movement (i.e. movement to

SpecCP). In doing so, I follow McCloskey (2000) and other relevant literature, which also leaves the

issue open. (Another, possibly related question is why the modifying-a-wh-trace option is apparently very

rare cross-linguistically. From a cross-linguistic point of view, Icelandic is actually well behaved, WUE

being a rare exception.)
Notice also that under neutral intonation, topicalization patterns with relativization in disallowing Q-float, so that the point made above with respect to (9b) can be extended to (10).

(10) ‘Bækurnar keypti Jón ekki allar.
the-books bought Jon not all
‘All the books, Jon didn’t buy.’

Now, there is a controversy regarding the landing site of Icelandic object shift. The majority of the literature assumes that the final landing site of Icelandic object shift is the accusative Case position, namely SpecvP (SpecAgroP in a framework that assumes AGR Phrases). Under this analysis, the relevant part of (9a) has the structure in (11), with the quantifier floated in the θ-position of the object, and ekki adjoined to VP.

(11) \[vP bækurnar [vP ekki [vP[V' allar]]]]

On the other hand, Bošković (1997, 2004a,b), Chomsky (1999), Hiraiwa (2001), and Svenonius (2001, 2002), among others, argue that Icelandic object shift involves movement to a position above SpecvP/SpecAgroP. I will assume here Bošković’s (2004a,b) implementation of this analysis, where it is argued that the floating quantifier in (9) is located in SpecvP (position through which the relevant NP passes), with ekki adjoined to vP (see Bošković 2004a,b for relevant discussion; note that it is argued in Bošković 2004a that floating allar in a lower position would in fact lead to a violation of licensing conditions on Q-float\(^8\)). The relevant part of (9a) then has the structure in (12).\(^9\)

Notice that one could try to account for (9b) under the object shift-feeding-relativization analysis (i.e. without parallel movement) by assuming that in languages like Icelandic and Standard English, which do not allow floating quantifiers to modify a trace left by wh-movement, a floating quantifier cannot be c-commanded by an A’-trace of the host DP (the A’-trace would be the trace left by relativization from the object shift position). However, this would not work because of constructions like (i), where under the feeding movement analysis the wh-trace in SpecTP c-commands all. The same point can be made with respect to Icelandic (ii).

(i) Which books must have all been bought?
(ii) bækurnar sem hafa allar verið keyptar
the-books that have all been bought

Notice that under the no feeding analysis, both wh-movement/relativization and movement to SpecTP take place from the deep object position in (i–ii), with movement to SpecTP proceeding successive cyclically, stranding the quantifier in an intermediate position. (Following Bošković 2002a, 2007b, Boeckx 2003, and Chomsky 2008, I assume that there is no feature checking in intermediate positions. Anticipating the discussion below, notice that there are no phases between the deep object position and SpecCP.)

\(^8\) The licensing conditions are incompatible with the movement-to-SpecvP analysis from (11), but are fully compatible with the movement-above-vP analysis, under the structure in (12).

\(^9\) For ease of exposition, I will continue to use the term object shift although it is not really appropriate under the movement above SpecvP/SpecAgroP analysis. There are many arguments in the literature that English objects move to SpecAgroP/SpecvP overtly (see, for example, Boeckx and Hornstein 2005, Bošković...
I will now consider how the structures in (11) and (12) fare with respect to the parallel movement analysis, on which wh-movement and object shift both take place from the position in which allar is located. Before comparing the two accounts, let me emphasize that I take the data in (9)–(10) to provide evidence for the parallel movement analysis of object shift/wh-movement “interaction.” This means that even acceptable examples involving such interaction should be treated in terms of parallel movement. This, for example, holds for (13), given Diesing’s (1996) arguments (see also Bobaljik 1995b) that object shift is obligatory in object shift contexts (i.e. with definite NPs).

Let us now try to tease apart the structures (11)–(12) by using the parallel movement analysis of object shift/wh-movement “interaction.” (Recall that allar merely indicates the launching site of parallel movement, which, as discussed above, takes place in (9b)–(10) as well as (13).) It turns out that the analysis cannot be applied to the structure in (11). If we were to apply the analysis to this structure, the relevant NP would simultaneously undergo object shift and wh-movement from the deep object position. However, the problem is that wh-movement from the complement position of the verb is blocked by Chomsky’s (2000, 2001) Phase-Impenetrability Condition (PIC), which says that only the edge of a phase is accessible for movement outside of a phase. Since vP is a phase, C cannot target an object within the VP complement of the vP phase head.

Turning now to the structure in (12), the parallel movement hypothesis can be easily applied to this structure. Here, the object first moves to the Spec of the vP
phase. The relevant NP then simultaneously undergoes object shift and wh-movement from this position. Since the position is located at the edge of the $vP$ phase, wh-movement does not violate the PIC on this derivation. We then have here an argument that the analysis on which the final landing site of Icelandic object shift is higher than $vP/AgroP$ is superior to the analysis on which Icelandic object shift lands in $SpecvP/SpecAgroP$.

In conclusion, I have provided a new argument for the parallel movement hypothesis based on Icelandic object shift. I have also shown that the parallel movement analysis enables us to tease apart two different approaches to Icelandic object shift. In particular, it provides evidence that the landing site of Icelandic object shift is higher than $SpecvP/SpecAgroP$.

Since, given the PIC, it is not possible to move out of $vP$ without moving to $SpecvP$, parallel movement for wh-movement and object shift would take place only from the phasal edge position, $SpecvP$. Strictly speaking, it is then not quite true that there is never any feeding relation between movements—movement to the phasal edge, $SpecvP$, feeds both object shift and wh-movement. Making the relevant distinction (when there is a feeding relation, and when there isn’t) is rather straightforward, given the relevance of phases/phasal edge for the feeding movement case.
Structure Building That Can’t Be*

SAMUEL DAVID EPSTEIN, HISATSUGU KITAHARA, AND T. DANIEL SEELY

11.1 Introduction

This chapter proposes and explores the form and empirical consequences of a deduction of cyclic Transfer. The basic idea is this: First, we adopt Chomsky’s (2007, 2008) feature-inheritance analysis whereby the phase-heads C and v (inherently) bear phi-features and transmit these features to the heads of their complement, T and V, respectively. Second, T and V, once they bear such inherited phi-features, then raise, via their edge feature (EF), a local DP to their Spec. This raising, e.g. raising to Spec-T after C is merged and after T inherits phi-features from C, is, informally speaking, “counter cyclic” Internal Merge (IM); i.e. it is non-extending Merge that fails to apply to the “root” node, namely the C-projection. Chomsky (2000) formulates the No-Tampering Condition (NTC) so as to allow such “counter-cyclic” operations. Unlike Chomsky, however, we argue that such “counter-cyclic” IM cannot actually “infix” Spec-T inside the C-rooted object but rather yields a structure with (again, informally) “two peaks”—a tree with no single root (to merge to)—causing the derivation to (deducibly) halt. For the derivation to continue, one

* Versions of this material were presented at The “Ways of Structure Building” conference held in November 2008 at The University of The Basque Country. We are grateful to Vidal Valmala and Myriam Uribe-Etxebarria for their hospitality and for their continued interest in our work. Aspects of this research were also presented at a colloquium talk at the Michigan State University Department of Linguistics. For valuable feedback we are grateful to Kyle Grove, John Hale, and Alan Munn. This research was also presented at a University of Maryland Department of Linguistics colloquium in April 2011. We thank especially Norbert Hornstein, Terje Londahl, and Dave Kush for very helpful feedback. We are also indebted to students in our “triply-cotaught” 2009 syntax seminar for very valuable contributions, including Miki Obata, David Medeiros, Tim Chou, and Dan Parker. Finally, but by no means least, we thank Noam Chomsky for his interest in and immeasurable contributions to our research.

1 Below, we consider two analyses regarding the offensiveness of the “two-peaked” phenomenon: one is that there is no single root, causing the derivation to halt until Transfer can apply, removing one of the
peak (the phase-head-complement—for reasons that we will return to) must be removed by Transfer from the workspace. In this way, we seek to deduce the timing of Transfer: always and only when a phase-head is externally merged and phi-features are inherited by the head of the phase-head-complement. And we seek to deduce the syntactic object targeted by Transfer, namely the phase-head-complement—transforming the two-peaked object into a single-peaked object to which Merge can once again apply.

Our analysis is based on a number of fundamental assumptions. One is that (following Chomsky (2008: 137)) Merge “comes free:” some structure building operation is required of any system, and Merge \((X, Y) \Rightarrow \{X, Y\}\) is the simplest such operation. By the Strong Minimalist Thesis (SMT), then, we seek to maximize the explanatory effects of Merge: the only relations available to the narrow syntax (NS) are those that directly follow from the simple, single, necessary structure building operation, Merge \((X, Y) \Rightarrow \{X, Y\}\). In effect, this is the derivational approach to syntactic relations of Epstein et al. (1998), Epstein (1999), Epstein and Seely (2002, 2006). Another key assumption of our analysis is what we call the Label Accessibility Condition (LAC) (adapted from Chomsky (2000: 132: (53)):

\[(i)\] The Label Accessibility Condition (LAC)

Only the label of an entire syntactic object, the root, is accessible to NS.\(^2\)

Any system must access something and given 3rd factor considerations, access is made with least effort; hence LAC is itself deducible if the label of the root is available to NS that has accessed the root.

We argue that our system has both conceptual and empirical advantages compared to that of Chomsky (2007, 2008). For example, Chomsky argues that “counter cyclic” IM is an instance of Merge. However, we show that, in fact, his analysis incorporates an additional “replacement” process (see Freidin (1999) for compelling arguments in favor of a ban on replacement operations in an earlier framework). Replacement poses a problem for the SMT, a problem that does not arise with the analysis proposed here. Our system also derives the “invisibility” of Spec-T to C. Given the derivational approach to syntactic relations, C is not in a relation to Spec-T since at the point in the derivation when C was merged to T, the Spec-T position was not yet created. This allows us to deduce a series of empirical effects noted but unexplained in Chomsky (2007, 2008). We argue that there are further empirical peaks. The other is that the two-peaked structure will never result in a single semantic value at the CI interface. Consequently, one of the peaks must be removed by Transfer in the syntax; i.e. each of the syntactic objects in the “two-peaked” structure is given over to CI individually.

\(^2\) As we will see in a moment, this formulation does not prohibit IM; however, IM takes place only after all EM (given some Lexical Array) has been exhausted. Notice also that under the simplified representation generated by Merge (i.e. one lacking “labels” in the sense of Chomsky (1995b)), an element of \([X, Y]\) is still designated as “the label.”
differences between the “traditional” Phase-Impenetrability Condition (PIC) and our two-peak-effect Transfer, and the latter gives apparently correct results for object agreement in Icelandic that are problematic for the timing of Transfer as formulated in the Chomsky (2007, 2008) system.

To illustrate our approach in more detail, consider a typical instance of what is referred to informally as “counter-cyclic” IM at the CP phase for a simple sentence like:

(2) Bill ate rice

Under Chomsky’s analysis, once the phase-head C is externally merged (EM) in,

(3) \[ CP \{ C [TP [T [vP Bill [v [VP ate rice]]]]]] \]

C transmits its lexically inherent phi-features to T. T now acts as a probe, locates the external argument Bill, and raises it to Spec of T, resulting in:

(4) \[ CP \{ C [TP Bill [T [vP Bill [v [VP ate rice]]]]]] \]

The claim in Chomsky (2007, 2008) is that “counter cyclic” IM as in (4) (along with IM at the root, e.g. wh-movement) is an instance of Merge. Recall there is one and only one structure building operation, namely, Merge (X, Y) \(\Rightarrow\) \{X, Y\}. But notice that, in fact, a form of “replacement” is required to get the (desired) result illustrated in (4). Informally speaking, to derive (4) from (3), the TP complement of C must be “disconnected” from this C; then IM merges Bill to this “existing” (specless) TP, creating a “new” (specful) TP; and then this “new” TP is “reconnected” to C, creating a “new” CP, one with a different TP (i.e. a TP containing Spec) now appearing as sister to C and the previous (specless) TP sister to C is no longer C’s sister. In set-theoretic terms, IM of Bill to the “existing” TP takes (5) as input:

(5) \{C, T^1\} (where \(T^1 = [T [vP Bill [v [VP ate rice]]]]\))

and produces a new syntactic object \(T^2\) as in (6):

(6) \(T^2 = [Bill [T [vP Bill [v [VP ate rice]]]]]\)

\(T^2\) (with Bill as Spec-T) then replaces \(T^1\) in (5): such replacement removes \(T^1\) from (5), and merges \(T^2\) with C. Note furthermore that a “new” CP now results, namely,

(7) \{C, T^2\}

It is this “new” CP in (7), and not the “old” one in (5), that enters into further derivational processes.

What we suggest is that given SMT, there is Merge, but Merge alone cannot replace existing terms, i.e. it cannot map (5) to (7). Accepting this, one might then argue that

\(^3\) For expository convenience we continue to use the terms “External Merge” and “Internal Merge,” but there is, in fact, just a single operation, Merge (X, Y) = \{X, Y\}. 

the application of two operations, namely Delete (= “disconnect”) and Merge (= “reconnect”) might be capable of mapping (5) to (7). However, we will argue that this possibility is banned by a Law of Conservation of Relations (LCR) which we propose as a generalization of the Extension Condition (EC) (which subsumes NTC):

(8) The Law of Conservation of Relations (LCR)

In NS, syntactic relations (among existing terms) cannot be altered throughout a derivation.

This is by hypothesis an “economy” condition, whereby it is arguably computationally natural that no relations are created only to be destroyed. LCR in effect forces NS to preserve all existing relations.\(^4,5\) Thus, “counter cyclic” IM, e.g. subject raising in (4), will necessarily result in (9), presented in informal graph-theoretic terms:

\[
\text{T2} \\
\text{CP} \\
\text{Bill} \\
\text{C} \\
\text{T1} \\
\text{T} \\
\text{vP} \\
\text{Bill} \\
\text{v}
\]

In effect, a “two-peaked” object is created. More formally, “counter cyclic” IM, generating (9), will produce two distinct but intersecting set-theoretic syntactic objects, SOs, which happen to share a term, namely, \(T^1 = \{T, vP\}\). That is, as a result of this “counter cyclic” IM, NS yields two intersecting sets: \(\{\text{Bill, } T^1\}\) and \(\{C, T^1\}\). We argue that these intersecting set-theoretic SOs do not have a single root; there is no single category that (reflexively) contains all terms. Since there is no root, then given LAC, NS cannot access either of these sets’ labels. These intersecting SOs thus cannot undergo Merge after the series of phase-level operations that generated them, i.e. the derivation halts.\(^6\) We propose further that in order to continue the derivation, one

\(^4\) LCR is more general (hence more restrictive); so, if LCR derives the empirically desirable aspects of EC and NTC, then LCR is preferable.

\(^5\) Note that if Chomsky’s (2004) theory of adjunction is extended to head-movement, head-adjunction does not alter existing “head-complement” relations. Whether head-movement exists in NS is an open question, but it is not clear whether LCR eliminates “syntactic” head-movement.

\(^6\) Alternatively, under the law of semantic composition, these intersecting set-theoretic SOs will not yield a single semantic value. Thus, there must be a way to overcome the intersecting situation in NS. One possible remedy is to send them to the semantic component separately, and we propose that Transfer dissolves this intersecting situation by removing one intersecting set from the workspace of NS. Informally
“peak” (i.e. one of the two intersecting sets) must be eliminated, precisely the job of Transfer. Thus, cyclic Transfer is deduced, and the positive empirical consequences that we consider then follow.

11.2 Valuation-induced Transfer: a problem

In this section, we consider the valuation-induced deduction of cyclic Transfer of Chomsky (2007, 2008) and Richards (2007). This is important since it might be asked: If a deduction of cyclic Transfer already exists, why propose an alternative? Our answer is that a viable deduction does not exist since, in fact, valuation-induced Transfer leads to massive under-generation. Reviewing Epstein, Kitahara, and Seely (EKS) (2010), this section details the empirical problem with the valuation-induced approach, thus setting the stage for our alternative deduction of cyclic Transfer, as outlined above.

According to Chomsky (2000, 2001), syntactically determined features, like phi-features of C/T and Case of N, are lexically unvalued,7 becoming valued in the course of a derivation if the appropriate conditions for the application of Agree (i.e. valuation) are met. Such syntactically valued features must be removed by Transfer from an object bound for the Conceptual-Intentional (CI) interface, since they are uninterpretable at CI and would induce CI crash.

A central question for Chomsky (2007, 2008) is: When does Transfer (more specifically, its “spell-out” function) remove such CI-offending features? Before valuation is too early. Unvalued features won’t be “implementable” at the Sensorimotor (SM) interface; in effect, they would induce SM crash. But, after valuation is too late. Once valued, the distinction between valued and unvalued is lost; thus Transfer won’t “know” to spell out, say, the valued phi-features of C/T, since after valuation they are identical to the (inherently) valued phi-features of N; thus these CI-offending features would not be removed from the CI-bound object, and CI crash would result. As detailed in Epstein and Seely (2002), Transfer must, in effect, operate during valuation, i.e. Transfer must take place internal to a valuation operation, or else internal to the (“all-at-once” created) phase (see Chomsky 2004, 2007, 2008). A crucial consideration for present concerns is that valuation-induced speaking, one “peak” (i.e. one of the two intersecting sets) must be sent to the semantic component by Transfer. Which of the peaks is transferred will be considered below, as we attempt to explain why it is the PHC that is transferred (which is currently stipulated, though empirically motivated).

7 Crucially (as argued in Epstein et al. (1998), Epstein and Seely (2002)), Transfer cannot “know” which features will be interpretable at the not-yet-reached interfaces, since this requires lookahead and Transfer being sensitive to Interpretability at the interface. This is precisely why feature-valuation was incorporated into the theory. But then, all Transfer can possibly be sensitive to is valued vs. unvalued, and so it then follows that once a feature is valued it is too late to Transfer it since it is indistinguishable from an inherently valued (+interpretable) feature.
Transfer requires at least some derivational history, minimized as much as possible; it must “see” what goes from unvalued to valued in order to “know” to spell out (or remove) from the CI-bound object just those features whose values it witnessed changing from unvalued to valued.8

The logic of valuation (attractively) induces cyclic Transfer in Chomsky’s analyses. The logic is also explicitly appealed to in Chomsky (2007, 2008) to derive CP, and not TP, as a phase, and to derive feature-inheritance (from C to T, and v to V). Adopting Richards (2007), Chomsky (2007: 13) states:

…it follows that the PIC entails that TP cannot be a phase, with operations of valuation and A-movement driven by properties of T.

Suppose TP were a phase. Then its interior will be transferred by PIC, but the head T will retain its valued uninterpretable features. The derivation will therefore crash at the next phase.

The logic of valuation-induced Transfer is clear: Syntactically valued features at the phase-edge induce CI crash (thus Richards (2007) deduces that phi-features must not remain on the phase-head and be valued there, by direct agreement between the phase-head (functioning as a) probe and a goal; this would yield valued features at the edge, causing CI crash at the next phase). But, as revealed in EKS (2010), this very same logic leads to a pervasive empirical problem with the valuation-induced cyclic Transfer system: all instances of movement across a phase boundary are disallowed and this leads to massive under-generation. If T cannot be a phase-head because syntactically valued features at the edge of T will induce crash at the next phase cycle, then C and v cannot be either since syntactically valued features do in fact occur at the edge of C and v.

Consider a specific example of this problem for a simple sentence like Whom did they like? The relevant vP is

\[
(10) \quad [vP \text{ they } [v' v [vP \text{ like whom}]])]
\]

The phase-head v transmits its phi-features to V. V probes and agrees with whom (valuing Case). To escape the phase-head-complement, VP, whom is raised (by the EF of v) to Spec-v, and like adjoins to v.9 As a result, NS yields:10

\[
(11) \quad [vP \text{ whom } [v' v [v' v+like [vP \text{ like whom}]])]]
\]

+Acc +Phi

8 Note that Transfer also removes phonological features, which do not undergo valuation. In section 11.3, we present an analysis under which syntactically valued features and phonological features form a natural class.

9 Note that the motivation for the adjunction of V to v is not entirely clear, but it is a step required to restore the original VO order.

10 We leave aside the raising of whom also to Spec-V, since it is not relevant here.
By the PIC, the phase-head-complement VP is transferred and hence inaccessible. But note that both whom bearing syntactically valued Case, and like bearing syntactically valued phi-features, appear at the edge of the phase-head v (i.e. v and its specifiers). Indeed, all syntactically valued features are at the phase-edge v.(11) The derivational history of the valuation of these features is then “lost” by the PIC, and Transfer has no way to “know” to remove these features at the next higher phase from the CI-bound object. In effect, these features are stranded at the phase-edge and will induce CI crash. Thus, under the valuation-induced Transfer system, there is massive under-generation.

If EKS (2010) is on the right track, then it would seem that valuation-induced Transfer (along with valuation-induced feature-inheritance) “does not work” since valued features must be allowed to (i.e. do) occur at phase-edges in convergent derivations. Thus the (elegant) valuation-based attempt to deduce cyclic spell-out apparently confronts serious empirical problems. It is thus important that we consider alternative approaches attempting to explain the timing and nature of cyclic Transfer.

11.3 Resolving the before/after problem

EKS (2010) provide a potential solution to the empirical problem of the valuation-induced Transfer system reviewed above. The short story is that syntactically valued features are [−CI] features; this is not stipulative but rather follows from a natural and attractively restrictive view of the nature of features as either interpretable to CI (= +CI) or not (= −CI). By definition, [−CI] features are those that are uninterpretable at CI (which include all phonological features). EKS propose that all and only [−CI] features are removed from the CI-bound object, and this is done by Transfer. Thus, after valuation is not “too late.” Since syntactically valued features are [−CI], their derivational history is no longer required for Transfer to operate effectively.

On what basis can EKS claim that syntactically valued features are [−CI]? EKS first assume, under SMT (with the primacy of CI), that features are either [+CI] or [−CI]. EKS then assume what can be referred to as a law of conservation of features, a generalization of the inclusiveness condition:

(12) The Law of Conservation of Features (LCF)
In NS, features cannot be created or destroyed throughout a derivation.

11 Chomsky (2007: 20) noted the relevance of examples such as (11) in which valued features seem to appear at the edge—which induces crash under the Chomsky/Richards’ analysis. But, Chomsky regards such examples as unproblematic, writing “[v]aluation of uninterpretable features clearly feeds A’-movement (e.g. in whom did you see?). Hence valuation is ‘abstract,’ functioning prior to Transfer to the SM interface . . .” It is not clear to us, however, how his “abstract valuation” would circumvent the appearance of syntactically valued features at the phase-edge in (11). Note that this problem is very general, both in English (e.g. Him, I like) and cross-linguistically.
According to LCF, the nature of a feature cannot be changed during the course of a
derivation. Given these two assumptions, in effect, \([-\text{CI}]\) features are always \([-\text{CI}]\)
features, likewise for \([+\text{CI}]\) features. It thus follows that since lexically unvalued
features are \([-\text{CI}]\), they remain \([-\text{CI}]\) features regardless of whether they are valued
in the course of a derivation. Thus, syntactically valued features are \([-\text{CI}]\) features.

The LCF thus expresses the generalization that syntactic feature valuation never
creates semantically interpretable features anew. Inclusiveness fails to express this
since valuation satisfies inclusiveness (no new feature is added) yet we need to
express the (apparent) fact that valuation of a given feature never creates anew a
semantic interpretation of that same feature.

With this much in place, the pervasive empirical problem revealed above is
eliminated. Consider again (11), the structure troublesome for Chomsky (2007,
2008). The syntactically valued features—Case of whom and phi-features of like—are at
the phase-edge. However, Transfer still knows to remove them from the
CI-bound object under the proposed framework, since they are \([-\text{CI}]\) features,
and Transfer removes all such features (including both syntactically valued phi-
features and phonological features) from the CI-bound object in the next phase
cycle.\footnote{Under the current assumptions, Chomsky’s (1998) definition of Transfer
will suffice. It essentially states: Transfer deletes all \([-\text{CI}]\) features from SO (constructed by NS), forming SO\(_{<+\text{CI}>}\) (bearing only
\([+\text{CI}]\) features); and it sends SO\(_{<+\text{CI}>}\) to the semantic component, while it sends SO to the phonological
component. Note, given this formulation of Transfer, there is no way for \([-\text{CI}]\) features to get to the CI
interface, i.e. CI is “crash-proof” with respect to CI features.} Transfer before valuation is still too early, since unvalued features will not be
“implementable” at SM. However, after valuation is no longer too late, and the
massive under-generation problem is therefore resolved.

So far, we argued that valuation-induced Transfer “doesn’t work,” since it is
internally inconsistent with the concurrently adopted assumption that syntactically
valued features do in fact appear at phase edges. Guided by SMT, we then reviewed a
solution to one aspect of this problem: after valuation is not “too late,” under a
natural, and attractively restrictive, appeal to the nature of features (as either \([+\text{CI}]\) or
\([-\text{CI}]\)), independently needed for Transfer to remove phonological features destined
for the CI interface. But, interestingly, we now must return to the questions we
started with: when, and to which SO, does Transfer apply; and why? The valuation-
induced Transfer system, we suggest, ultimately fails to answer these questions. Must
we then stipulate the timing of Transfer, or can it be deduced? If after Valuation is
not too late to Transfer, then when, after Valuation, does Transfer apply? As
mentioned at the outset, we argue that the timing and the target of Transfer can be
deduced; it follows from the simplest characterization of structure building, and from
key 3rd factor considerations.
11.4 Deducing cyclic Transfer

Recall our basic idea of Transfer outlined in section 11.1: applying “counter cyclic” IM, but without appeal to replacement, produces a “two-peaked” object, or more formally, two intersecting set-theoretic SOs:

\[
\begin{align*}
& T^2 \\
& CP \\
& C \\
& T^1 \\
& T \\
& vP \\
& \text{Bill} \\
& v
\end{align*}
\]

Since there is no single root that (reflexively) contains all terms, LAC prevents these “two-peaked” intersecting SOs from undergoing Merge (after the series of phase-level operations that generated them). For the derivation to continue, one peak (i.e. one of the two intersecting SOs) must be removed, and this task can be carried out, we suggest, by Transfer. Thus, “counter cyclic” IM creates intersecting SOs that are not accessible to further Merge (effectively halting the derivation) and this “forces” Transfer to apply—see also fn. 1 and 6 above. Let us now consider the details of this approach, motivating and articulating our central assumptions.

11.4.1 How Merge applies

A number of key assumptions enter into our approach. First, we assume SMT, according to which,

language is an optimal solution to interface conditions that FL must satisfy; that is, language is an optimal way to link sound and meaning… If SMT held fully, UG would be restricted to properties imposed by interface conditions. A primary task of the MP is to clarify the notions that enter into SMT and to determine how closely the ideal can be approached.

Chomsky (2008: 2)

The “goal” of NS is to optimally build SOs that are legible, and hence useable to, the interfaces SM and CI. This has been a standard assumption of the Minimalist Program since its inception.

We assume further that 3rd factor considerations play an important role in what counts as “optimal.” 3rd factor involves “[p]rinciples not specific to the language faculty,” in particular, “…principles of structural architecture and developmental constraints…including principles of efficient computation, which would be…of particular significance for [all] computational systems…” (Chomsky 2005: 6).
One such 3rd factor efficiency consideration is LAC: only the label of an entire syntactic object, the root, is accessible to NS. But what is the root? We assume the following definition, under which not-yet-merged lexical items and non-intersecting set-theoretic SOs form a natural class:

(14) K is the root iff:
     for any Z, Z a term of K, every object that Z is a term of is a term of K.

As for term, we adopt a simplified definition, adapted from Chomsky (1995b):

(15) For any structure K,
     (i) K is a term of K, and
     (ii) if L is a term of K, then the members of L are terms of K.

The central (standard) idea is that the root is the term that reflexively contains all relevant terms (as defined above).

Any system must access something, namely the units out of which other units are constructed. Given 3rd factor considerations, access is optimal if made with least effort. By accessing only the label of the root, search is minimized, and therefore efficient. Thus, we assume with Chomsky (1995b), EM applies at the root only.

Merge is an indispensable part of the human faculty of language. It comes, as Chomsky suggests, “for free.” But, how Merge operates is constrained by 3rd factor considerations. We thus take seriously the idea of minimal search. The basic intuition is that NS executes Merge with minimal search, so that computational complexity is greatly reduced. Thus far, minimal search has been invoked to constrain only IM. Here we would like to extend the 3rd factor notion of “minimal search” to EM (as we must if there is indeed only Merge). In principle, there are two ways of finding SOs: direct finding and finding through search. Regardless of whether SO is a lexical item or not, NS finds it directly if it is the root. If an SO is an internal term of the root, however, NS must search into the root to find that SO. Thus, regardless of whether the root is a lexical item or a set-theoretic object, merger of two roots (i.e. EM) is always preferable to merger of one root and its subpart (i.e. IM), because finding roots involves no search. Thus, it follows from this general 3rd factor consideration that all applications of EM necessarily precede any application of IM within the construction of each phase (this is reminiscent of “Merge over Move” of Chomsky (1995b, 2000)). And if the label of the root is available to NS that has accessed the root, then LAC, which states that only the label of the root is available to NS, also follows. NS can find the root and its label directly, but it must search into the root to find any internal terms.

We would like to point out one important consequence of this efficiency-based analysis, which involves EM. By LAC, NS has access to only the highest label of the root. Thus, only EM of two roots is possible. NS cannot “reach” inside the root and “pull out” its term for EM. Thus, in (16),
only the entire object \( Z \) can participate in EM; \( X \) and \( W \), contained within \( Z \), are inaccessible to EM. Take a simple case. Suppose EM merges \( V_1 \) and NP, forming \( \{V_1, \text{NP}\} \). Then, it follows from 3rd factor considerations that EM cannot merge \( V_2 \) and NP (the latter a term of \( \{V_1, \text{NP}\} \)). Such merger of \( V_2 \) and NP has been (in essence) adopted by the sideward movement analysis of Hornstein (2001) and Nunes (2001) and the multidominance analysis of Citko (2005). But if the analysis proposed here is on the right track, then such merger of \( V_2 \) and NP is prohibited on 3rd factor grounds (see also Chomsky (2007: 8), Kitahara (2007)). Of course, more research into the phenomena in question is required and we note here only the conditional and formal entailment regarding incompatibility of theories.

However, after NS exhausts EM (meaning all lexical items from a subarray are now terms of the single root), search into the root is permitted. And in accord with minimal search, operations other than EM can be initiated. Such operations include feature-inheritance, Agree, and IM. With Chomsky (2007, 2008), we assume that the phase heads \( C \) and \( v \) inherently bear phi-features (and EF). Thus, once \( C \) and \( v \) are accessed, feature-inheritance from \( C \) to T and from \( v \) to V may take place. Agree values phi-features and Case, which takes place under the (restrictive) probe-goal mechanism. An accessed phase-head itself has access only to “PIC visible” SOs within its derivational c-command domain (= terms of the category merged with the phase-head), and such SOs can undergo IM to the edge of the accessed head.13

Recall that Merge “comes free.” Given SMT reviewed above, we thus seek to maximize the explanatory effects of Merge. Thus, the only relations available to NS are those that directly follow from the simple, single, necessary structure building operation Merge.

With this much in place, “replacement” cannot be postulated as a new operation, and of course, Merge alone cannot replace existing terms. But might the application of two distinct non-replacement operations, namely Delete (= “disconnect”) and Merge (= “reconnect”) carry out this task? We argue that this possibility is banned by LCR, which in effect forces NS to preserve all existing relations. Thus, “counter cyclic” IM (e.g. subject raising) will necessarily result in (17):

---

13 The Phase Impenetrability Condition, Chomsky (2000) states, in effect, that the complement of a phase head is inaccessible to the syntax at the next phase level.
Recall that (17) is an informal representation of the two intersecting SOs: \{C, T^1\} and \{Subj, T^1\}, which happen to share the term T^1. Crucially, these “two-peaked” SOs have no root, given the definition of “root” above. Neither T^2 nor CP is a term that contains all other terms: T^2 fails since one of its terms, say, T, is a term of CP yet CP is not a term of T^2. CP fails since one of its terms, say, T, is a term of T^2 yet T^2 is not a term of CP. Since there is no root, LAC prevents both T^2 and CP (and all other terms in (17)) from undergoing Merge after the series of phase-level operations that formed the two intersecting SOs in (17).

Given these assumptions, the derivation halts, but to see just where and how Transfer comes into play, we must first explore other central mechanisms of the system, beginning with feature-inheritance.

11.4.2 Feature-inheritance

With Chomsky (2007, 2008), we assume that the phase-heads C and v inherently bear phi-features. Recall that Chomsky/Richards deduce feature-inheritance from C/v to T/V, given the “before and after” problem noted above. For Chomsky/Richards, C/v cannot keep their phi-features, since if they did, these features would be valued at the phase-edge. But then, given PIC, their former unvalued status would be “lost.” The syntactically valued features on the edge are then, in the eyes of Transfer, indistinguishable from inherently valued features and so are un-transferred, ultimately causing CI crash.

Recall that, for us, since syntactically valued features are [−CI], there is no “after” problem for Transfer; after valuation is not too late, since Transfer “knows” to spell out [−CI] features, and syntactically valued features are [−CI]. How, then, do we induce feature-inheritance?

We propose that phi-features must in fact be a necessary but not sufficient condition for Case-valuation. Case-valuation is achieved not by phi-features alone—which

---

14 An alternative analysis is that the two-peaked object would not yield a single semantic interpretation, and thus in the syntax one peak would be removed and sent to the CI interface alone.
would fail to distinguish the Case valued by V (ACC) from the Case valued by T (NOM)—but by the combination of phi-features plus the “transitive property” (of transitive active voice verbs) or phi-features plus the tense element, borne by T (see Chomsky (1995b, 2000, 2001)). Case-valuation is thus done through the feature complexes: (i) phi + tense ⇒ Nominative, and (ii) phi + transitive ⇒ Accusative. So, we resurrect and maintain the standard (and we suggest ineliminable) Case-theoretic assumption: (i) finite T “assigns” NOM, and (ii) transitive V “assigns” ACC.

Thus, phi-feature transmission from C/v to T/V is motivated when Case-valuation is required (and possible). Presumably, in languages like West Flemish, C exhibits phi-agreement but not Tense. Chomsky (2007: 20) continues:

If that is basically accurate, then there are two possibilities. One is that Tense is a property of C, and is inherited by T. The other is that Tense is a property of T, but receives only some residual interpretation unless selected by C (or in other configurations, e.g., in English-like modal constructions).

Note that the first option (i.e. C bears tense and T inherits it from C) is adopted by Richards (2007) (without discussion). Here notice, that if the second option (i.e. T inherently bears tense) is adopted, then feature-inheritance is, for us, necessary for Case-valuation (hence, a prerequisite for convergence). Keeping this in mind, consider the following passage continued from the above quote:

One advantage of the latter option [the second option above] is that T will then have at least some feature in the lexicon, and it is not clear what would be the status of an LI with no features (one of the problems with postulating AGR or other null elements). Another advantage would be an explanation for why C never manifests tense in the manner of phi-features (if that is correct).

With these advantages, we adopt the second option, which forces feature-inheritance to take place for Case-valuation.

11.4.3 EPP = “double” EF

The next question is: What motivates IM? Chomsky (2007, 2008) proposes that for a lexical item, LI, to permit Merge (regardless of whether EM or IM), it must have some “mergeability” property. This property is called “the edge feature” EF of LI. We suggest that EF is in fact a higher-order property of features; similarly the property “strength” is a higher-order property of features in Chomsky 1995b.15 All linguistic

15 This view sheds light on the otherwise curious characteristics of the EF as formulated in Chomsky; namely: (i) EF is a property of LI (Chomsky 2008: 139); (ii) EF is undeletable from LI (Chomsky 2007: 11); (iii) EF (unlike phi-features) does not involve feature-matching (Chomsky 2008: 161:fn.49); and (iv) EF can be inherited from the phase-head along with the phi-features (Chomsky 2008: 157). (i) follows from the natural assumption that if a LI is made up of features that have the property “mergeability,” then the LI itself will have that property (likewise in Chomsky (1995b) a category had the property “strong” by virtue of one of its features having this property). (ii) follows from recoverability: there is no edge feature, but rather only lexical features with the property “mergeability,” since a feature cannot be deleted under
features, and specifically for present concerns the phi features of C/v, have the higher-order property of "mergeability." There is transmission of phi features from the phase-head to the head that it selects. What this means is that C will transmit unvalued phi-features to T, as will v to V. After feature-inheritance, T and V will bear "double" EF: one inherent on T/V (by virtue of the features that make up T and V) and one from inheritance of features from C/v. We propose that this "double" EF requires merger to its bearer T/V (while "single" EF only permits merger). In effect, this "double" EF is (our version of) EPP (a property requiring merger to its bearer). But, for us, the EPP is a derived property, emerging only with feature-inheritance. We have not deduced the EPP; we have no more to add about why "double EF" requires a Spec than anyone else does about why the EPP requires a Spec. But, our derived EPP (emerging in the course of a derivation), combined with our overall approach, does have positive empirical consequences.

11.5 Consequences

The analysis considered above provides a deduction of cyclic Transfer. It also has other empirical consequences.

11.5.1 Deriving the "invisibility" of Spec-T to C

Chomsky (2008) suggests that the ban on the extraction of the PP complement from within Spec-T, as in (18),

(18) *[CP [PP of which car] did [TP [the driver tPP] cause a scandal]]?

follows in part from the stipulation that Spec-T is invisible (as a goal) to (the probe) C. Specifically, he proposes that Spec-T becomes invisible to further computation once its Case is valued generalizing the inactivity condition of earlier work (cf. Nevins (2005), Obata and Epstein (2008)). By contrast, however, Spec-C (unlike Spec-T) continues to be visible after the valuation of its Case, as is required to allow successive cyclic wh-movement (e.g. who do you think saw her?). Exempting the construction of A-bar chains from such a generalized inactivity condition is a stipulation, and this asymmetry remains to be explained. The alleged invisibility of Spec-T poses another problem as well. Take the indirect question (e.g. I wonder who saw her). Under Chomsky’s (2008) phase-based model, Spec-C and Spec-T (each occupied by who) must be created simultaneously, and Spec-T becomes invisible upon the valuation of its Case. But then, how can Spec-T, being invisible, count as a position lower than Spec-C?

Chomsky’s (2008) analysis thus confronts (at least) the following two questions: (Q1) How does Spec-T become invisible, while Spec-C continues to be visible? and

recoverability, then neither can the edge property. (iii) Since "mergeability" is a (higher-order) property of a feature and not a feature itself, then there is no feature matching associated with EF. Finally, (iv) follows since phi features can be inherited, and "mergeability" is a property of these phi features.
(Q2) How can the relative height of Spec-T be calculated when Spec-T itself is invisible? In this section, we seek to provide a principled answer to each of these questions, and argue that this can be done by deducing the invisible status of Spec-T as a property of the independently motivated derivational model (Epstein et al. (1998), Epstein (1999), Epstein and Seely (2002, 2006)).

Under the derivational approach, NS establishes syntactic relations derivationally through the application of the indispensable structure building operation Merge, and no syntactic relation can be arbitrarily defined on output representations (contra GB theory). Specifically, c-command is the relation that Merge establishes between $a$ and terms of $\beta$ at the exact point of merging $a$ with $\beta$. One unique property of this approach is the following: Suppose $a$ is merged counter-cyclically with an embedded category $\gamma$, where $\gamma$ is not the root but is a distinct term of $\beta$ (i.e. $\gamma$ is embedded within $\beta$). Representationally, all the terms of $\beta$ that appear above $\gamma$ would either c-command or dominate $a$. But derivationally, these higher terms of $\beta$ will neither c-command nor dominate $a$. Why? Because when these higher terms of $\beta$ underwent their birth-merger, they were not merged with a category containing $a$ now residing in the counter-cyclically merged position. Thus, these higher terms of $\beta$ bear no derivational relation to such a counter-cyclically merged $a$. (See Kawashima and Kitahara (1996), and Epstein et al. (1998).) Under the current assumptions, this conclusion is strengthened. That is, it follows (without stipulation) that neither C nor Spec-C c-commands (counter-cyclically created) Spec-T derivationally; notice also that in the two-peaked representation, (13), C does not representationally c-command Spec-T either.

To see why this is, recall the indirect wh-question (e.g. I wonder who saw her) and consider the following trees (indices are introduced only for expository purposes, and the linear order is irrelevant):
The question is: Does $\text{who}_3 (= \text{Spec-C})$ or $\text{C}$ c-command $\text{who}_2 (= \text{Spec-T})$? Within Chomsky’s (2008) feature-inheritance analysis reviewed above (see also Richards 2007), EM merging C with $T^1$ precedes the necessarily simultaneous applications of IM, which create Spec-T and Spec-C by merging $\text{who}$ with $T^1$ (forming $T^2$) and $\text{who}$ with $C^1$ (forming $C^2$), respectively. With this execution of merger operations, it naturally follows that $\text{who}_2 (= \text{Spec-T})$ c-commands every term of its merged sister $T^1$, but neither $\text{C}$ nor $\text{who}_3 (= \text{Spec-C})$ c-commands $\text{who}_2 (= \text{Spec-T})$, as represented in (19). This is the case because C was merged with $T^1$ (of which Spec-T was not at the time a term), $\text{who}_3 (= \text{Spec-C})$ was merged with $C^1$ (of which Spec-T was not at the time a term), and c-command relations are established derivationally, at the exact point of merger. The present analysis thus deduces the invisible status of Spec-T to both $\text{C}$ and Spec-C as a property of the derivational model, while allowing Spec-C to continue to be visible to any “higher” category merged to a category of which Spec-C is a term—thereby allowing successive cyclic wh-movement, as desired.

Relative to such “counter-cyclic” IM, Chomsky (2008: 150) notes:

EF of $\text{C}$ cannot extract the PP complement from within SPEC-T: if it could, the subject-condition effects would be obviated. It must be, then, that the SPEC-T position is impenetrable to EF, and a far more natural principle would be that it is simply invisible to EF.

No analysis of this optimal result is offered in Chomsky (2007, 2008). But, the present analysis explains the “invisible” status of the Spec-T position.

As for calculating the height of invisible Spec-T from its visible occurrence, consider again (19) above. Chomsky (2008) suggests that the simultaneous applications of IM form a relation between $\text{who}_2$ and $\text{who}_n$, and a relation between $\text{who}_3$ and $\text{who}_n$, but not a relation between $\text{who}_3$ and $\text{who}_2$. Why is there no such relation? Presumably because there is no application of IM involving the positions of $\text{who}_3$ and $\text{who}_n$, and if there is no such Merge application, there will be no relation between them (assuming a derivational theory of relations). We argue that the “invisible” status of Spec-T to both $\text{C}$ and Spec-C is deducible from how Merge has applied in accord with the 3rd factor principle. If this deduction is on the right track, then the height of Spec-T is not calculated by the position of Spec-T itself. Instead, we propose that the height of Spec-T should be calculated by the position of its occurrence. Chomsky (1995b) defines an occurrence of K as the category to which K is merged. An occurrence of Spec-T is then its merger-sister $T^1$. Under this occurrence-based calculation, the position of a category is uniquely determinable by reference to its merger-sister. Thus, Spec-T counts as a position lower than Spec-C, because $T^1$ (= the occurrence of Spec-T) is a distinct term of $C^1$ (= the occurrence of Spec-C), and this crucial “term-of” relation between $T^1$ and $C^1$ was established derivationally.
11.5 Consequences

11.5.2 Icelandic object agreement revisited

Chomsky (2000) postulates that the complement domain of the phase-head _v_ is not accessible to operations outside of the phase _vP_. Thus, as desired in the following case, T cannot agree with NP_OBJ inside the phase-head-complement VP:

(20) a.* John _like_ these socks
    John.NOM like.PL these socks.PL.ACC

b.* . . . T.PL . . . \[vP v [VP V NP.PL.ACC]\]

This follows from the stipulation that Transfer removes the phase-head-complement VP from the workspace of NS at the completion of the phase _vP_. So when T is introduced, the VP (containing NP_OBJ) is “gone.”

However, it has been widely discussed that in languages like Icelandic, T can agree with NP_OBJ, as illustrated in (21) (an example discussed in Jónsson (1996), also cited in Bobaljik (2008)):16

(21) a. OK Jóni _likuðu_ þessir sokkar
    Jon.DAT like.PL these socks.PL.NOM

b. OK . . . T.PL . . . \[vP v [VP V NP.PL.NOM]\]

To allow this, Chomsky (2001, 2004) delays Transfer-application (removing the VP) until the merger of the next “higher” phase-head (= C). T—inherently bearing lexical phi-features of its own in Chomsky (2001, 2004)—thus can agree with NP_OBJ, exactly when T is merged, and hence before the subsequent merger of C, which triggers Transfer of VP.

Crucially however, recall that Chomsky (2007, 2008) proposes that T does not inherently bear phi-features of its own, but instead inherits phi-features from C. But this feature-inheritance analysis creates a new problem: before the merger of C to TP, T (lacking phi-features of its own) cannot possibly phi-agree with NP_OBJ, but after the merger of C to TP, T (now bearing phi-features inherited from C) cannot find the NP_OBJ, because the phase-head-complement VP (containing the NP_OBJ) has been removed by Transfer.

Under the current assumptions, in particular, the new deduction of cyclic Transfer developed in the preceding sections, such T-NP_OBJ agreement receives a principled account without recourse to an independent agreement parameter.

16 We leave aside how Nominative Case (NOM) gets realized on NP_OBJ in (21), since NOM can appear on NP_OBJ even when an intervening NP clearly blocks agreement between T and NP_OBJ as in (i) (an example discussed in Schütze (1997), also cited in Bobaljik (2008)):

(i) Mér virðist \[Jóni vera taldir t líka hestarnir\]
    Me.DAT seemed.SG Jon.DAT be believed.PL like horses.PL.NOM

    “I perceive Jon to be believed to like horses.”
Recall that Transfer removes the phase-head-complement if, and only if, the phase-head transmits its phi-features to the head of the phase-head-complement, which in turn values Case (ultimately creating the two-peaked effect). Consequently, in (20), Transfer removes the phase-head-complement VP, whose head V values Case, thereby correctly blocking T-NPOBJ agreement. By contrast, in (21), Transfer does not remove the phase-head-complement VP, whose head V does not value Case, thereby allowing T-NPOBJ agreement. Thus, (20) vs. (21) follows, and the English-Icelandic “parameter” is reduced to independent morpho-lexical Case-variant properties of heads.

This analysis also explains why T-NPOBJ agreement occurs with passive/raising verbs, as illustrated in (22) (an example discussed in Zanen, Maling, and Thráinsson (1985), also cited in Bobaljik (2008)):

\[(22)\]

a. OK Um venturinn voru konunginum gefnar ambáttir
   “In the winter were.PL the king.DAT given slaves.PL.NOM”

b. OK ...T.PL... [vP [vP Vpassive NP.PL.NOM]]

Legate (2002) argues that a verbal phrase with a passive/raising verb is phasal. Chomsky (2001) stipulates that these are “weak” phases, which, unlike his original “strong” phases, escape from Transfer-application. We eliminate this ad hoc distinction, reducing it to independently motivated Case properties.

Finally, let us address, once again, why Transfer removes the phase-head-complement if, and only if, the head of the phase-head-complement values Case. Under the feature-inheritance analysis, in order for V/T to value Case, they must first inherit unvalued phi-features from later-merged v/C respectively. This merger of v/C in turn projects vP and CP. But then “EPP/double-EF” satisfaction (i.e. creation of Spec-V and Spec-T) is necessarily “counter-cyclic,” i.e. it is merger but not at the root. As argued above, such “counter-cyclic” merger necessarily yields a “two-peaked” object—formally two intersecting SOs with no single root (to merge to)—causing the derivation to (deducibly) halt. To continue, one peak (the phase-head-complement) must be removed from the workspace of NS by Transfer.

---

17 At the time this volume was going to press, we learned that Takahashi (2010), on entirely independently grounds and based on extensive empirical evidence, argues that Case determines phase-hood. We have attempted to deduce this conclusion.
Specificity-driven Syntactic Derivation*

ANTJE LAHNE

12.1 Introduction

Specificity is arguably one of the main basic concepts of morphological theory. It is typically invoked in the form of the Subset Principle (e.g. Halle 1997) to resolve competitions between the markers of a language which arise due to lexical underspecification of inflectional markers. The Subset Principle has the effect that whenever a more specific marker competes for a syntactic context with a less specific marker, then it is always the more specific marker that is inserted.

It seems to be a recurring pattern in syntactic derivations, too, that structure building operations are ordered in such a way that elements that are more “specific” in one sense or another are preferred by the probing element over less specific elements. This pattern emerges in a number of seemingly unrelated environments, such as long-distance agreement (e.g. Polinsky and Potsdam 2001, Bobaljik and Wurmbrand 2005, Bhatt 2005) and Mahajan’s Generalization (Mahajan 1994, 1997, Sauerland 1995, Bittner and Hale 1996), which were previously accounted for by independent principles.

The goal of this chapter is to propose a new approach to syntactic structure building that accounts for these data in a uniform way. The underlying hypothesis is that specificity is a much more powerful underlying syntactic principle than was previously thought. I propose that syntactic derivations are driven by a specificity...
principle on Merge: when more than one constituent is a candidate for an agreement relationship or displacement operation, the operation will involve the more specific constituent. The term specificity here is not used in its formal semantic sense, but in its morphological sense: an element X is more specific than an element Y iff X has more (matching) features than Y. How specific an item is, is sometimes not visible at first sight. In some cases, we need to look into the internal make-up of the feature set of an element, i.e. features must be decomposed.

The “state of the art” in locality research is that locality effects exhibiting locality by intervention (=relative locality) have been uniformly attributed to the Minimal Link Condition (Chomsky 2000: 123, Chomsky 2001: 27, Fitzpatrick 2002). However, the MLC partly overlaps with the Phase Impenetrability Condition, a principle of absolute locality, as has been argued e.g. by Rizzi (2006) and Müller (2006a). I will call this case the superiority case. Though it is arguably impossible to attribute all locality effects to just one of the two locality types and thus dispense with principles of the second type, the desideratum is a unifying approach with no explanatory overlaps between the two types of principles. Furthermore, the Minimal Link Condition does not make accurate predictions for a second syntactic constellation, which I will call the equidistance case.

The new principle I am arguing for in this chapter yields all relative locality effects excluding the superiority case, as it does not crucially draw on the availability of search space, but on competition between potential goals. In addition, the specificity principle also accounts for a number of anti-MLC effects, such as order-preserving movement and anti-superiority effects, which were previously accounted for by independent, partially overlapping principles. The resulting system is one which works without the concept of intervention by closeness.

This chapter is programmatic in nature. It offers a new perspective on the nature of intervention and aims at renewing the discussion about the nature of locality.

The chapter is structured in the following way: section 12.2 gives the theoretical background on which the new approaches are based and introduces the new General Specificity Principle that is argued for in this chapter. In section 12.3, I discuss how effects of the Minimal Link Condition can be derived by means of the General Specificity Principle. I then show, in section 12.4, how the new principle can be employed to derive anti-MLC effects; I focus on order-preserving movement (section 12.4.1) and superiority violations (section 12.4.2). In the Outlook section (12.5) I discuss the limits of the new system and its status within the theory of grammar.

12.2 Theoretical background

Following Manzini (1994), Epstein and Seely (2002), Müller (2004), and Lahne (2008c), I adopt the null hypothesis that each phrase is a locality domain; in other words, each phrase is a phase. Consequently, the search space of the current head X at
a stage of the derivation comprises the next lowest head Y and the edge of Y, but not the material embedded under Y. Furthermore, this analysis employs a clausal structure consisting of the core categories C – I – v – V.

The underlying argument structure is such that the internal argument is merged as the sister of V, the external argument as Spec, v (Murasugi 1992, among others); v assigns internal case, and I assigns external case. The morphology model adopted here is lexical-realizational, thus the feature bundles delivered by C_{nu} are post-syntactically correlated with phonological features (Halle and Marantz 1993, 1994, Harley and Noyer 1999, etc.). Syntax thus manipulates abstract heads without phonological content, whereas morphology interprets the output of syntax.

I propose that syntactic derivations are driven by a specificity principle on Merge. The underlying principle is given in (1).

(1) General Specificity Principle:
A probe undergoes a syntactic operation with the most specific matching goal. Specificity is determined by cardinality of morpho-syntactic features: a set Q is more specific than a set H iff |Q| > |H|.

This principle is an extension of the Maximize Matching Effects constraint.

(2) Maximize Matching Effects (MME; Chomsky 2001: 15):
If local (P,G) match and are active, their uninterpretable features must be eliminated at once, as fully as possible; partial elimination of features under Match, followed by elimination of the residue under Match, is not an option.

Principle (2) operates at the points when the currently active head could potentially deal with its unhandled features in several ways. It has the effect that an agreement relation between a probe and a goal involves handling the maximal number of matching features.3

This becomes crucial when there is more than one potential action that the current phase head can trigger. For example, in (3), the C head contains a movement-inducing [\cdot\textit{wh}^\cdot]-feature.4 Wh-movement can be decomposed into an Agree relation (Agree C_{\text{[wh]}}, what_{\text{[wh]}}) and a Move operation (Merge C, what). One possibility is now that C agrees with the wh-element what in the feature [wh], but then moves

2 Arguments in favor of late insertion are given e.g. by Bobaljik (2002), Sauerland (1997).
3 The MME Principle is a version of the Earliness Principle (Pesetsky 1989), which is an independently well-motivated assumption underlying minimalist derivations.

(i) Earliness Principle:
An uninterpretable feature must be marked for deletion as early in the derivation as possible (Pesetsky 1989, Pesetsky and Torrego 2001).

4 Features that trigger structure building by (Internal or External) Merge or Move are notated as bulleted throughout this chapter; this notation follows Heck and Müller (2006).
another constituent, for instance the subject, to its edge, while the wh-element stays
lower in the structure. This, however, results in an unacceptable sentence, as (3b)
shows. The desired result is that C agrees with what in [wh] and then attracts what to
its edge, as shown in (3a).

(3)  a. What had the boy seen?
    b. The boy (‘what) had (‘what) seen (‘what)?      (No echo question)

This result is possible if the concurrence of syntactic operations is forced by an
underlying specificity principle such as the MME.

A central assumption of this analysis is that syntactic heads are bundles of
unordered features, which has been widely assumed for morphological features
(Jakobson 1936, Bierwisch 1967, Kiparsky 2001), and for the operation Merge [a, β]
(where a and β are unordered; Chomsky 1995b). This notion of head also corre-
sponds to feature value matrices in Lexical–Functional Grammar (Kaplan and
Bresnan 1982). However, there is internal structure in that the individual features
that make up the head are decomposable into subsets (or feature geometries), and
the features contained in those subsets are recognizable as belonging to a specific
subset. The effect of this idea is that all features of an active head are accessible (as
opposed to e.g. Müller (2008b), where due to a feature hierarchy, only one feature of
an active head is accessible at a time). In addition, several features can be dealt with
simultaneously in Agree or Merge operations (this is equivalent to what happens
during vocabulary insertion in post-syntactic morphology models like Distributed
Morphology).

The effect of the GSP is identical to that of MME in constellations where a single
potential goal for an active probe is involved: Agree between P and G involves
handling of the maximal number of matching features. However, MME and the GSP
differ in one crucial respect: MME does not have an effect on the choice of goal if
there is more than one potential goal in the search space; due to the Minimal Link
Condition (MLC), it is invariably the closest goal that the probe agrees with. The
GSP, on the other hand, has the additional effect that, with more than one potential
goal being available in the search space of a probe, the probe agrees with the goal that
has the highest number of matching features. These two cases are best illustrated by
means of an abstract example. In a possible derivation, there is a goal γ in the search
space of a probe π, and {F₁*, F₂*} contained in π’s feature set P match γ’s features
{F₁, F₂}.

Following Sternefeld (2006), the stars are used here to indicate that a feature F is an unvalued probe,
no matter if it is an agreeing or attracting probe.
Example 1: Maximize Matching Effects Case

\[ \pi P \]

\[ \pi \quad \gamma P/x \]

\[ [P=\{F_1, F_2, F_3, F_4\}] \]

\[ \gamma \quad \ldots \]

\[ [G=\{F_1, F_2, F_5, F_6\}] \]

An agreement relation involving only \( *F_1 \) and \( F_1 \) is not an option in this constellation. Thus, (5) is an impossible derivational step; the agreement relation must involve both \( *F_1 \), \( F_1 \) and \( *F_2 \), \( F_2 \), as shown in (6).

\[ \pi P \]

\[ \pi \quad \gamma P/x \]

\[ [P=\{F_1, F_2, F_3, F_4\}] \]

\[ \gamma \quad \ldots \]

\[ [G=\{F_1, F_2, F_5, F_6\}] \]

In this example, the GSP and MME have the same effect. However, the two principles differ in their effect in contexts where more than one potential goal is available in the search space of a probe. In derivations obeying MME, it is always the closest goal that the probe agrees with, due to the Minimal Link Condition. The GSP, on the other hand, can overwrite the MLC: If the GSP is assumed to be at work, then the probe...
agrees with the goal that has the highest number of matching features, no matter if it is the closest available goal or not. This is demonstrated in a second abstract example: In a second possible derivation, there are two potential goals, $\gamma$ and $\alpha$, in the search space of $\pi$. The closer goal $\gamma$ has two matching features for $\pi$, and the more remote goal $\alpha$ has three matching features.

(7) Example II: The GSP Case

\[
\begin{align*}
\pi & \\
\text{(P=\{*F_1*,*F_2*,*F_3*,*F_4*\})} & \gamma & \alphaP/\piP \\
\gamma & \alphaP/\piP \\
\text{(G=\{F_1,F_2,F_5,F_6\})} & \text{(A=\{F_1,F_2,F_3,F_7\})}
\end{align*}
\]

In this constellation, feature handling between $\pi$ and the closer goal $\gamma$ is not an option (see (8)). Likewise, feature handling involving only $*F_1*$, $F_1$ and $*F_2*$, $F_2$ is not valid, as feature handling must involve the greatest possible number of features that can be dealt with (see (9)). Thus, in a derivation obeying the GSP, the agreement relation must involve $\{*F_1*,*F_2*,*F_3*\}$ of $\pi$ and $\{F_1,F_2,F_3\}$ of $\alpha$; this is shown in (10).

(8)

\[
\begin{align*}
\pi & \\
\text{(P=\{*F_1*,*F_2*,*F_3*,*F_4*\})} & \gamma & \alphaP/\piP \\
\gamma & \alphaP/\piP \\
\text{(G=\{F_1,F_2,F_5,F_6\})} & \text{(A=\{F_1,F_2,F_3,F_7\})}
\end{align*}
\]
This system is predicted to act in such a way that a probe discharges as many features as possible as soon as possible. This allows for cases of apparent indecisiveness, namely when there are two goals that the probe can agree with in the same number of features. In Lahne (2008b), I argue that this constellation leads to optionality. The phenomenon discussed there is defective intervention in Icelandic:

(11) Mér finnst/finnast tölvurnar ljótar
me.DAT find.sg/find.pl computers:DEF.NOM.PL ugly
‘I consider the computers ugly’

(Holmberg and Hróarsdóttir 2003)

In brief, the approach to these data in Lahne (2008b) is that \( \Gamma^o \) has three kinds of probing features: the EPP feature, the set of phi-features, and the case feature \( [\text{nom}^*] \). However, there is no constituent that \( \Gamma^o \) can handle all three features with—it cannot use the EPP feature on the lower DP \( tölvurnar \) (as it is an argument of the embedded predicate), and neither can it assign nominative case to the dative subject \( mér \)—but it is possible to use two of the three features on one goal, and the remaining feature on the other goal. There are two possible orders in which this can be done. The idea is that the order of
operations is optional, and both orders are attested: one order leads to the singular agreement option, the other leads to plural agreement. The features \([\bullet \text{D} \bullet]\) and \([\ast \text{nom}\ast]\) cannot be handled together, as the EPP feature cannot be used on the lower DP. This leaves two possible feature combinations: One possibility is to handle two of the three features, \([\bullet \text{D} \bullet]\) and \([\text{n} \bullet]\), with the dative argument, and subsequently use the remaining feature \([\ast \text{case} : \text{nom}\ast]\) on the lower DP, as shown in (12).

As a result, \(I\) shows singular agreement. The second possibility is to use \([\ast \text{case} : \text{nom}\ast]\) and \([\text{n} \bullet]\) on the lower DP, and then handle \([\bullet \text{D} \bullet]\) with the quirky subject, as shown in (13):

Now the verb shows plural agreement.

The morphology model adopted here is lexical-realizational, thus the feature bundles delivered by \(C_{\text{HL}}\) are post-syntactically correlated with phonological features.
Syntax thus manipulates abstract heads without phonological content, whereas morphology interprets the output of syntax.

### 12.3 Deriving Minimal Link Condition effects

The Minimal Link Condition (MLC; Chomsky 2000: 123, Chomsky 2001: 27, Fitzpatrick 2002) is arguably the only locality condition of the relative type which is widely agreed on today. Interestingly, its effects, which are discussed in detail in this section, can also be derived as effects of the General Specificity Principle. In what follows, I demonstrate that the syntactic constellations for which the MLC makes a statement (superiority case, equidistance case, A-over-A case, two-edges case) follow in a system in which not the closest available goal is chosen, but the most specific goal.

The strongest version of the Minimal Link Condition is given in (14).

(14) **Generalized Minimal Link Condition** (e.g. Fitzpatrick 2002):

In a structure $X_1[\bullet\bullet] \ldots [Y_{[F]} \ldots [Z_{[F]}]]$, movement to $[\bullet F \bullet]$ can only affect the category bearing the $[F]$-feature that is closer to $[\bullet F \bullet]$.

As Müller (2004, 2006a) notes, the effect of the MLC is limited by the Phase Impenetrability Condition, as the MLC presupposes search space, while one of the conceptual reasons for phases is to reduce the derivational complexity and thus relieve active memory by limiting the search space (Chomsky 2004).

In what follows, I will refer to subportions of a structure in which $X$ is a (phase) head, and $Y$ the next lowest (phase) head; $G_1$–$G_5$ are potential goals:

(15)

```
XP
  └── G₁ ─ X´
    └── (G₂) ─ X´
        └── X₀ ─ YP
            └── (G₃) ─ Y´
                └── G₄ ─ Y´
                    └── Y₀ ─ G₅ ...
```

Arguments in favour of late insertion are given e.g. by Bobaljik (2002), Sauerland (1997).
There are four possible constellations that the MLC is to derive:

(I) The superiority case:
A probe X has to decide between a goal at the edge of Y and a goal located in the complement of Y (G₄ vs. G₅).

(II) The equidistance case:
X has to decide between two goals from the edge of the lower head Y (G₃ vs. G₄).

(III) The A-over-A case:
X has to decide between Y(P) and an element from the edge of Y (Y vs. G₄).

(IV) The two-edges case:
X has to decide between two goals, the one at the edge of X, the other at the edge of Y (G₁ vs. G₄).

12.3.1 Superiority and equidistance effects
An MLC effect of type (I) is e.g. the well-known subject-object asymmetry with multiple wh-questions in English shown in (16).

(16) a. Who₁ tₐ saw what?
    b. *What did who see tₐ?

The relevant portion of the derivation is given in (17).

(17) \[
\begin{array}{c}
X' \\
/ \\
X^o \quad YP \\
/ \\
G₄ \quad Y' \\
/ \\
Y^o \quad ... \quad G₅ \quad ...
\end{array}
\]

Here the MLC has the desired effect that an operation cannot involve Xₒ and G₅ due to the presence of the higher possible goal G₄. However, for this case, even under the widest definition of the Phase Impenetrability Condition (Chomsky 2000: 108, Chomsky 2001: 13)⁷, there are MLC effects that are ruled out by it. In a minimal-static system, the observed effect follows automatically: G₅ is inaccessible to X as it is already spelled out when X is merged (Müller 2006b). Thus, if each phrase is a phase,

⁷ (i) Phase Impenetrability Condition:
The domain of a head X of a phase XP is not accessible to operations outside XP; only X and its edge are accessible to such operations.
then the MLC is not needed to derive this case. This overlap should be eliminated if possible.⁸

Müller (2006a) takes this as a conceptual argument against the MLC:

this could be taken to suggest that strictly derivational approaches should dispense with the MLC in toto since this constraint presupposes an articulated representation (the search space for the probe). Arguably, in a derivational approach, minimality effects should not be covered by a constraint that accesses a significant amount of syntactic structure, i.e., a representation (MLC); rather, they should emerge as epiphenomena of constraints that reduce the search space (PIC).

The only domain where the MLC could have an effect in such a system is thus the search space of the current head. One such case is the equidistance case, where a probe has to decide between a number goal which are all located at the edge of the next-lower phrase head. The relevant portion of structure is given in (18).

Here the MLC, taken strictly, makes the prediction that the presence of G₃ blocks operations involving X and G₄. This is, however, the wrong result: it is in exactly this constellation that anti-MLC effects like order-preserving movement occur (Richards 1997, Starke 2001, Williams 2003), that is, the probe does not seem to choose the closest goal, but actually the structurally deepest goal. This problem was solved by assuming that a strict application of the MLC is blocked by a prioritized equidistance principle (Chomsky 1993, 2001) according to which G₃ and G₄ are equally distant from X:

| (19) | **Equidistance** (Chomsky 2001: 27): Terms of the edge of HP are equidistant from probe P. |

⁸ See Rizzi (2006) and Chomsky (2000), who notes (p. 47, fn. 52): “The effect on the MLC is limited under the PIC, which bars ‘deep search’ by the probe.”
According to an alternative proposal, X attracts the goals in the empirically “wrong” order, that is, first $G_3$, then $G_4$, but this effect is repaired by a “Tucking-In” mechanism (Richards 1997), which, during the attraction of $G_4$, breaks up the structure just created and merges $G_4$ anti-cyclically below $G_3$.

If we look back at the two cases discussed so far, it seems that the MLC is not successful in deriving the correct result. In the superiority case, its effect is yielded by an independent core principle of grammar (= the Phase Impenetrability Condition); in the equidistance case, it makes wrong empirical predictions, which calls for either an additional constraint which makes the MLC unapplicable for this case in the first place, or a remedy so drastic that two basic underlying principles of grammar, the Strict Cycle Condition and the Extension Condition, are violated to correct the result. Müller (2006a) concludes from this that the MLC is not at work at all. If this is so, then the superiority case is a case of absolute locality (i.e. a PIC effect), and in the equidistance case, there is a priori no preference for either of the goals. The advantage of this solution is that it eliminates the explanatory overlap between relative and absolute locality (i.e. between the MLC and the PIC), and that it allows for order-preserving movement without invoking tucking-in or equidistance—note that the trigger for order-preserving movement is in any case an independent, additional factor.

The discussion has so far centered on problems occurring if the locality effects discussed are approached from an intervention-by-closeness perspective. How does the General Specifi city Principle perform in this respect? In the superiority case, the GSP does not overlap with the PIC, as it does not draw on the availability of search space (i.e. a representation); rather, it weighs the number of derivational actions a probe can undergo with one or several goals. With regard to the equidistance case, I argue that the GSP derives the findings correctly: probes choosing their goal on the basis of specificity do not automatically prefer the highest goal, but the most specific goal. This view leads to a number of interesting effects; see section 12.4.2 for a detailed discussion.

12.3.2 A-over-A and two-edges effects

However, in the two remaining cases (III) and (IV), an MLC-less system seems to make the wrong predictions. Let us examine at A-over-A effects first. The relevant structure is given in (20).

(20) \[
\begin{array}{c}
X' \\
\text{Xo} & \text{YP} \\
G_3 & Y' \\
Yo & \text{...}
\end{array}
\]
Let us have a look at four examples that exhibit this constellation. The first example involves Unambiguous Domination effects in German as analyzed by Takano (1994), Koizumi (1995), Kitahara (1997), Müller (1998), Sauerland (1999): Scrambling of a whole VP is possible, but scrambling of an NP inside the VP followed by remnant scrambling of the VP leads to an unacceptable result. This is illustrated in (21).

(21) a. dass [vP_2 [vp_2 zu lesen] keiner t_2 versucht hat]
   that the book.ACC to read no.one.NOM tried has
   “That no one tried to read the book”

   b. *dass [vP_2 [vp_2 zu lesen] [np_1 das Buch] keiner t_2 versucht hat]
   that to read the book.ACC no.one.NOM tried has

The second example comes from Breton agreement (Jouitteau and Rezac 2006). Breton shows a complementarity effect in that the $\phi$-features of a phonologically null NP are coded by $\phi$-agreement morphology on the verb (= “rich agreement”), whereas the $\phi$-features of a phonologically overt NP are not coded by $\phi$-agreement morphology on the target (“invariant agreement” [= frozen 3SG agreement or bare stem]). This is illustrated in (22).

(22) a. Gant o mamm e karf-ent /*karf-e pro bez-añ
   with their mother R would.love-3PL /*would.love-3SG 3PL be-INF
   “They would like to be with their mother”

   b. Gant o mamm e "karf-ent /*karf-e Azenor
   with their mother R "would.love-3PL /*would.love-3SG Azenor
   ha Iona bez-añ
   and Iona be-INF
   “Azenor and Iona would like to be with their mother”

Jouitteau and Rezac (2006) analyze the complementarity as a locality effect: $v$ in Breton has nominal properties. It is thus assumed to bear interpretable 3SG $\phi$-features. Consequently, when $v$ probes for $\phi$-features in its search space, then the node $v/vP$ intervenes between $I$ and the external argument, which is contained in the $vP$. I must therefore value its unvalued $\phi$-features with $\phi$-features of $v$, which results in 3SG (“frozen”) agreement on $I$. The rich agreement is merely an effect of incorporation: If the external argument is an affixal pro, then it incorporates into $T$ and thus contributes its $\phi$-features to the feature set of $T$ (i.e. it becomes a bound pronoun), which surfaces as rich agreement. The observation important to our discussion is that agreement of a higher probe $X$ with a lower head $Y$ seems to be preferred over agreement with an element at the edge of $Y$. 

12.3 Deriving Minimal Link Condition effects 283
A third example, also discussed in Jouitteau and Rezac, is phi-agreement in English (2006: 1927):

\[(23) \ \text{\textit{GerP us/"we arriving on time\}} seems/\textit{seem t}_1 \text{ (to be) unlikely.}}\]

Here the phi-probe I has two possible goals in its search space, the features \([3SG]\) of the gerund, and the features \([1PL]\) of the personal pronoun, which is located at the edge of the gerund phrase. At the time when I probes for phi-features, the gerund is located in its complement. Two observations are relevant. First, the probe invariably agrees with the \([3SG]\)-features of the gerund phrase, as can be seen from the ungrammaticality of "us/we arriving on time seem to be unlikely. Second, the first person plural pronoun does not appear in the nominative, but in a non-agreeing default case. This indicates that the probe I can neither agree in phi-features with the specifier of Ger, nor can it assign nominative case to it. Thus, agreement between I and Ger/GerP is preferred to agreement between I and material at the edge of the gerund phrase.

A very similar constellation can be found in the context, of Long-Distance Agreement (LDA), which is an agreement relation that seems to hold across the boundaries of locality domains. Basically, the configuration of LDA is such that the verb of a root clause agrees with the internal argument of an embedded clause. It occurs e.g. in Tsez, Khwarshi, Kutchi Gujarati, Hindi-Urdu, Blackfoot, Chukchee, Basque, and Itelmen (Polinsky and Potsdam 2001, Khalilova 2007, Grosz and Patel 2006, Bhatt 2005, Bobaljik and Wurmbrand 2005, Etxepare 2006, Bošković 2007a, among others). An example from Khwarshi is given in (24).

\[(24) \ \text{I\textquotesingle}šet\textquotesingle ul y-iq\textquotesingle-še goli [uža bataxu}
\text{cl5-know-PRS COP [boy(cl1):ERG bread(cl5)]}
\text{y-acc-u]}
\text{cl5-eat-PSTPT].CL4}
"Mother knows that the boy ate bread" \quad \text{(Khwarshi; Khalilova 2007)}

Butt (1993, 1995), Chomsky (2001), Legate (2005), Lahne (2008b) analyze LDA as cyclic Agree; i.e. the matrix verb agrees with the embedded verb, which has previously agreed with the embedded object. From this point of view, Tsez and Khwarshi LDA works in the following way: the embedded verb agrees with its internal argument in noun class, and then contributes this feature ([cl5] in the example above) to the features of embedded I by \(v\)-to-I movement (Lahne 2008b). The matrix verb, which, too, agrees with its object in noun class, now finds this feature and agrees with it.\(^9\) The relevant portion of structure is shown in (25).

\(^9\) The preference for \(\text{[cl:5]}\) over \(\text{[cl:4]}\), too, can be derived as a specificity effect (Lahne 2008b). The optionality of LDA is derived by optional \(V\)-to-I movement: if \(v\)-to-I movement does not take place, then the matrix verb, which probes for the class feature of its complement, only finds embedded I’s inherent
The important point for our discussion is that the external (ergative) argument, located at the edge of I, never intervenes in the class agreement between the matrix verb and its complement (see Bobaljik 2006: 29). Thus again, the data suggest that a probe prefers agreeing with the head Y of its complement to agreeing with an element at the edge of Y.

At first sight, these data could be taken as an argument for the view that although the Minimal Link Condition in its strongest version may be abandoned, a locality condition based on intervention by dominance must be kept to derive these cases. This route is taken by Müller (2004), who, on these grounds, argues for a weaker version of the MLC (which is actually an original “ingredient” of it): the classic A-over-A condition.10

\[ \text{(25)} \]

\[
\begin{array}{c}
\text{V} \\
\text{\[\text{\bullet cat: I, cl: 5\]}} \\
\text{\textit{knows}} \\
\text{\textit{NP}_{\text{ext}}} \\
\text{\[\text{case:erg, cl: 1\]}} \\
\text{\textit{boy}} \\
\text{\[\text{cl: 4, cl: 5\]}} \\
\text{\textit{I'}} \\
\text{IP} \\
\text{\textit{\langle vP\rangle}} \\
\end{array}
\]

\text{ate}

\text{The A-over-A condition correctly derives A-over-A cases, and does not say anything for the superiority and equidistance cases, which is the desired result. However, even this last residue of a closeness-based locality condition can be derived by the Generalized Specificity Principle. In all the cases we looked at—Unambiguous Domination effects, Breton agreement, phi-agreement in English, and long-distance agreement—a head X merges with a category YP due to a selectional feature [\text{\bullet cat: Y\bullet}] (see the abstraction in (27)). On the background assumption that the features of a head percolate to its projections (see e.g. Rezac 2004, Sternefeld 2006), X does not only find categorial information in the node YP, but also all head features of Y. Crucially, due to the GSP, agree between X and Y/YP must involve handling of the maximal number of matching features. Thus, if X has another feature [\text{\bullet F\bullet}] that it can value with a feature of Y/YP, then it must value it with Y/YP.}
How can that approach be applied to the cases we examined? Let us take LDA in Tsez and Khwarshi as an example. The merging of the embedded clause with the matrix verb is driven by a selectional feature (here: \[C15\]) of the matrix verb, which is checked against the categorial feature of \(I\). In addition, the matrix verb also agrees with its complement in noun class; that is, it has an unvalued probing feature [class:\(\square\)]. The specificity principle demands that if matrix \(V\) can value another of its features with a feature of \(I\), then it must do so. We already saw that there is always a noun class feature contained in \(I\): \(I\)'s own feature [cl:4], plus, under certain conditions, the object agreement feature [cl:5]). Thus, matrix \(V\) always chooses one of these features for agreement; it is therefore never led to probe on for the class features of the embedded ergative argument. The embedded external argument therefore neither takes part nor intervenes in long-distance agreement.11

\[\begin{align*}
\text{(a)} \quad & X' \quad X^0 \\
& \quad \downarrow Y_P \quad \uparrow Y' \\
& \quad \downarrow Y^0 \quad \uparrow G_3 \\
\text{(b)} \quad & X' \quad X^0 \\
& \quad \downarrow Y_P \quad \uparrow Y' \\
& \quad \downarrow Y^0 \quad \uparrow G_3 \\
\end{align*}\]

[Substandard] Basque long-distance agreement differs from the type presented here in that although the target of Basque LDA must be the embedded absolutive object, the matrix verb does not agree with the embedded object in all the categories that are involved in object agreement: the finite matrix verb agrees with the embedded object in number, but with the whole object clause in case (the seeming LDA in person, on the other hand, seems to be a case of restructuring (Etxepare 2006)). This seems to be contradictory to specificity-based derivations: the matrix probe seems to spurn the number feature of embedded \(I\), preferring that of the embedded object. However, Etxepare (2006) argues that the number feature of the closest embedded object is in fact present on the embedded, nominalized \(I\) (by agreement). Basque LDA can therefore be analyzed along the same lines as in Tsez and Khwarshi: The matrix verb regularly probes for case and number features of its object, and checks case against the embedded infinitive predicate. The expectation is now that the matrix verb also probes for number, and the specificity principle demands that if it can value its number feature with the same element, then it must do so. The crucial point is now that the embedded predicate is a nominalization, and, according to Etxepare (2006), embedded \(I\) has an uninterpretable number feature valued by the closest DP. In addition, by virtue of being of category N, the nominalized predicate must also bear the default phi-features [3SG]. Thus, the matrix probe finds two number features, one of which is always [3SG]. If the second feature is [3SG], too, then it is impossible to decide with which of them the matrix verb agrees—this is a case of “invisible optionality.” If, however, the second feature is [PL], then the plural feature is the more specific match, as it contains more internal structure (Harley and Ritter 2002), and the matrix probe prefers it over the [3SG] feature. Basque LDA is optional; the variation must be located in the presence or absence of an uninterpretable number feature on embedded \(I\).
Let us now consider the remaining two-edges case, recall (IV). The relevant structure is shown in (28).

\[(28)\]

\[
\begin{array}{c}
\text{XP} \\
\text{G}_1 \quad \text{X'} \\
\text{X}^o \quad \text{YP} \\
\text{G}_3 \quad \text{Y'} \\
\text{Y}^o \quad \ldots
\end{array}
\]

It was already argued above that the original MLC need not be retained. Crucially, if only the A-over-A condition is retained, then \(G_1\) does not block agreement between \(X\) and \(G_3\). There are, however, data that suggest that the presence of a higher specifier does indeed block a value operation between a head and an element at a lower phase edge.

One such example is the emergence of the ergative case pattern as proposed in Müller (2008a).\(^{12}\) The original analysis is based on intervention by closeness, more specifically, on a definition of closeness according to which the specifier of a head \(X\) is closer to \(X\) than an element contained in the complement of \(X\) (Pesetsky 1982, Collins 1994). The basic idea of Müller’s analysis is the logical possibility that \(v\) merges the external argument before it assigns case. This creates a two-edges case: the external argument is located at the edge of \(v\), and the internal argument at the edge of \(V\). In the next step of the derivation, \(v\) must assign its case feature \([\text{case:internal}]\) (where internal case refers to accusative and ergative, which are taken to be identical). Because of the notion of closeness according to which the specifier of a head \(X\) is closer to \(X\) than an element contained in the complement of \(X\), \(v\) is now forced to assign its internal case to the external argument. This gives rise to an ergative case encoding pattern. This is illustrated in (29).

\(^{12}\) Another example is presented in Lahne (2008c), where verb inversion in the context of successive-cyclic movement (e.g. Henry 1995, Kayne and Pollock 1978, Torregro 1984, Rizzi and Roberts 1989, Rizzi 1990a) is analyzed as a result of the early timing of Phase Balance (a constraint that synchronizes the current make-up of the workspace with the shape of the current phase): wh-movement to the edge of \(I\) happens before \(I\) can deal with its own EPP and \(\phi\)-features. The wh-element then blocks Value/Agree between \(I\) and elements at the edge of \(v\). As a result, \(I\) must move to \(C\) by a last-resort operation in order to satisfy these features.
The example seems to suggest that in structures of type (IV), intervention by closeness does play a role after all. We have, however, already reduced the MLC to the A-over-A condition in the previous discussion. The A-over-A condition does not make a prediction for this case—it is only defined for competition between the complement of a probe and material at the edge of the complement. Can the A-over-A condition be widened again in such way that it derives the two-edges case, or is it possible to formulate a closeness-based condition that holds only for case (IV)? The answer is yes, but only with considerable effort. I will eventually reject this option, but let me briefly sketch the difficulties it faces. A closeness-based solution would be to define an intervention condition that jumps in only in constellations of type (IV), and that has the effect that specifiers of X are closer to X than specifiers of Y, while the specifiers “within” an edge do not block each other. This constraint needed to be defined in such a way that it only holds for Value/Agree relations, and not for structure building, as derivation by phase unquestionably involves successive-cyclic movement from the edge of Y to the edge of X even if X already has one or more specifiers.\footnote{This is indeed the solution proposed in Müller (2008a) and Lahne (2008a), where intervention is defined by closeness (path): a specifier of X is closer to X than an element contained in the complement of X.} I will, however, not follow this path, as it is based upon a very specific stipulated definition of what counts as the shortest path, which is arguably an unsatisfying result.

A second, and possibly more interesting, solution would be to explain the apparent MLC-effect in case (IV) by means of an independent principle. There is indeed such a constraint: the General Specificity Principle (GSP). In the two examples discussed, the intervening specifier is merged due to a selectional feature \([\bullet N\bullet]\). The GSP has the effect that Agree between a probe and a goal must involve the maximal number of matching features. Hence, if X satisfies a selectional feature by merging a specifier \(G_2\), and if it can satisfy more features with \(G_2\), then it must satisfy these features with \(G_2\). Hence, Value/Agree operations involving X and \(G_3\) are blocked.
12.4 Deriving anti-MLC effects

This section continues the discussion in 12.3.1 above: There are a number of syntactic effects that seem to run counter to the Minimal Link Condition. In this section I show that these effects can be derived in a framework in which relative locality is derived by specificity.

12.4.1 Order-preserving movement

Order-preserving movement is a phenomenon that occurs in some languages when more than one constituent is moved within one derivation. It constitutes itself in the observation that the final order of the moved constituents corresponds to the order in which they were base-merged. This poses a challenge for intervention-by-closeness approaches, as the final order cannot be reliably predicted if a probe invariably chooses to attract the structurally higher goal first (see Richards 1997, Starke 2001, Williams 2003).

One language in which order-preserving movement manifests itself is Bulgarian. In Bulgarian multiple wh-constructions, the order of the fronted wh-phrases is exactly the same as the base-merge order: a fronted wh_{nom}-phrase precedes a wh_{acc}-phrase, and when three verbal arguments are wh-fronted, then the order is invariably “NP_{nom} < NP_{acc} < NP_{prep}” (Rudin 1988):14

(30) a. Koj_{i} kogo_{j} vižda t_{i} t_{j}? (only possible order of wh-expressions)
   who whom sees
   ‘Who sees whom?’

   b. Koj_{i} kogo_{j} na kogo_{k} e pokazal t_{i} t_{j} t_{k}? (only possible order)
      who whom to whom has pointed.out
      ‘Who pointed out whom to whom?’

The same order of wh-elements occurs under long extraction:

(31) Koj kude misliš če e otišul?
   who where think.2s that has gone
   ‘Who do you think has gone where?’

I would like to propose that this is a specificity effect. The new analysis is based on the following case feature decomposition:15

---

14 In this section, I will limit myself to ordering effects with regard to case features; I am therefore ignoring, for the time being, cases of extraction from secondary predicates and equatives.
15 The decomposition is based on Caha (2006), who proposes a similar decomposition of case into feature sets varying in size. Though detecting a possible universal approach to case decomposition is a desirable result, nothing definite can be said about this; I will therefore restrain myself from speculating about the possible universality of this decomposition.
Case Decomposition

NOMINATIVE [case]
ACCUSATIVE [case object]
PREPOSITIONAL [case object oblique]

In ditransitive contexts, V thus contains two selectional features [N], and two case features to be assigned, [case object] and [case object oblique]. The GSP has the effect that the argument which V merges with first is assigned the more specific case feature (= [case object oblique]). V’s second argument then receives the less specific case feature (= [case object]). The prediction is thus that structurally lower arguments always receive a more specific case than higher arguments.

The same argumentation holds for the data in (30): C has three structure building [wh]-features, and three potential goals in its search space. The difference to the specificity effects in section 12.3 is that all potential goals have the same number of matching features, namely [wh]. The probe thus has three operation options, none of which yields a more specific match than the other. We saw a similar (but not identical) case in section 12.2: defective intervention in Icelandic, where a probe uses two of three probing features on one goal, and the remaining feature on the other goal. The order of application—which goal is targeted for the more specific match—is undefined; this situation is resolved by allowing for optionality, that is, both orders of operation application are acceptable. However, the derivation is still strictly specificity-based: the operation involving two features has priority over the operation involving one feature. The situation here is different: all potential goals have the same number of matching features, namely [wh]. There is therefore a priori no specific order of operation application forced by the system; in other words, the probe cannot decide which element to merge first. The system must therefore find a repair strategy that allows it to continue with the derivation.

I propose that the probing head C makes a decision on the basis of the featural make-up of its potential goals: C now probes further into the feature sets of the goals, being sensitive to syntactically relevant features. Following e.g. Wiese (2003), I assume that wh-elements are underspecified with regard to person and number; they are fully specified only for category ([cat:D]), quantification ([+wh]), and case. The wh-elements are identical in [cat:D, wh]; they differ only in their case feature: køj bears [case] kogo [case object], and na kogo [case object oblique]. Thus, even though case does not take part in the Agree operation between C and the wh-items, it is now the decisive factor for the order of merge: C attracts the goal that is most specific in terms of case features, which is na kogo. C’s second goal is the second most specific wh-item in terms of case, kogo. The least specific wh-element, køj, is merged last. The prediction is thus that elements that are less specific with regard to morpho-syntactic features appear higher in the structure than more specific elements.
The system thus employs a specificity-driven repair strategy in cases where no decision can be made on the basis of specificity-driven matching alone. That we are dealing with a repair strategy here becomes particularly evident by the observation that, unlike the “strict” specificity effects in section 12.3, it can be overridden by other factors such as animacy constraints. This is the case in Bulgarian, where the specificity effect interacts with an animacy effect: animate arguments appear higher in the left periphery than inanimate arguments, regardless of their case marking. This is illustrated in (32) and (33).

((32) a. ?(?) Kakvo kogo e spoletijalo?
what whom is stricken
‘What struck whom?’
(Billing and Rudin 1996: 38)

b. Kogo kakvo e spoletjalo?
whom what is stricken
‘What struck whom?’

((33) a. ? Kakvo na kogo mu xaresva?
what to whom to-him appeals
‘What appeals to whom?’
(Billing and Rudin 1996: 40)

b. Na kogo kakvo mu xaresva?
to whom what to-him appeals
‘What appeals to whom?’

A similar effect can be found in English:

((34) Whom did what upset? (Simpson and Bhattacharya 2007)

As (34) shows, in English, too, the MLC effect can be overwritten if the object wh-phrase crosses an inanimate wh-subject.

I would like to propose that these are independent effects interfering with the specificity-driven derivation in such a way that C’s specificity-based repair strategy can be overwritten if it runs counter to a more important constraint, e.g. linearization preferences linked to animacy according to which animate elements must be merged last (see Simpson and Bhattacharya 2007).

12.4.2 Anti-superiority effects

Cinque (1999) and Rizzi (2002), among others, show that an intervening adverb blocks displacement of a deeper adverb to the C domain, but not if the deeper adverb is contrastively focused. Some German data illustrating this effect are shown in (35); similar effects can be found in French and Italian.
I would like to propose that this is a specificity effect. Consider the following structure, which shows the derivation shortly before C attracts an adverb:

```
(36)  C'
    /    \
   /     \      
  C       IP
    hat
      /   \  
     /     \ 
    NP     I'
      Karl
          /   \ 
         /     \ 
        AdvP   I'
          wahr
             /\ 
            /  \ 
           vP   <I>
               /   \ 
              /     \ 
             AdvP   v'
               sehr
                  /\ 
                 /  \ 
                <NP_Karl> v'

```

In a neutral sentence, both adverbs are equally specific in terms of relevant syntactic features. However, remember that in line with the theoretical background assumed here, the vP is already spelled out at the point when C comes into the structure, as I is a phase head. Thus, example (35c) is a true superiority effect; an unmarked vP adverb is not a prefield competitor for the sentence adverb in the first place. The situation is different for example (35d). Here the lower adverbal bears an additional focus feature. In this case, C, in addition to the verb-second requirement, bears a
structure building feature \[\bullet\text{focus}\] (or \[\bullet\text{contrast}\]; cf. e.g. Rizzi (1997). Now the lower adverb is successive-cyclically moved to the edge of I. Consequently, both the sentence adverb and the vP adverb are in the search space of C; they as well as the subject are therefore competing goals. Now specificity jumps in: C could satisfy its verb-second requirement with any of the three goals, but it can satisfy its \[\bullet\text{focus}\]-feature only with the vP adverb \textit{SEHR OFT}. Consequently, it chooses to satisfy both its verb-second requirement and its \[\bullet\text{focus}\]-feature with the vP adverb. 16

A similar effect can be found in German and Spanish, where superiority violations lead to a grammatical result:

(37) a. Wer hat was gesehen?
   who has what seen
   ‘Who saw what?’

   b. Was hat wer gesehen?
   what saw who

(38) a. quién ha comprado qué?
   who has bought what
   ‘Who bought what?’

   b. qué ha comprado quién?
   what has bought who

It has been argued by Wiltschko (1997, 1998), Fanselow (2004b), and Müller (2004) among others, that the underlying structure of (37) is not an “unmarked” structure, but a scrambled structure. There is thus an additional feature involved in the derivation, which is satisfied by Agree with the internal argument. The internal argument is therefore moved on, and the external argument is left behind. The same is arguably true for Spanish, although Spanish does not exhibit scrambling. Spanish has, however, clitic doubling in the context of left dislocation, which is the same type of semantic or information structure-related movement. This approach

16 I deliberately remain unspecific as to what the “verb-second requirement” actually is. It can of course be a feature \([\bullet X]\); then we have a simple case of specificity in terms of feature cardinality. It could, however, also be the trigger for what Frey (2004) calls “formal movement:” in the original proposal, this mechanism blindly moves the highest constituent of the middle field into the prefield, while preserving the semantic or pragmatic properties of the constituent, and without assigning it any more properties. This mechanism was originally proposed by Bhatt (1999) for Kashmiri; Frey calls it \textit{formal movement}, as the mechanism “does not seem to be related to any semantic or pragmatic property but seems to be a purely formal one.” I think that this mechanism is entirely compatible with the specificity approach if it is modified in such a way that it is not automatically the highest element that is moved to the prefield; rather, if there is an element X that has “additional business to do” with C (i.e. C has a structure building feature that attracts X), then X is preferred over other elements in the middle field. The two operations need not even be seen as distinct operations: Formal movement is not feature-driven and therefore a last-resort operation. But if the requirement that the prefield be filled can be met by a legitimate movement operation which is going to happen anyway (focus fronting in our case), then the focus constituent is the preferred goal for C.
allows for the empirical generalization that languages allowing for superiority violations also allow for information structure-related movement such as scrambling or clitic doubling.

12.5 Outlook

It seems to be a recurring pattern in syntactic derivations that structure building operations are ordered in such a way that elements that are more “specific” in one sense or another are preferred by probes to less specific elements. The new specificity principle has been shown to uniformly derive MLC and anti-MLC effects. But the principle has a much wider application potential; see e.g. Lahne (2008b), where specificity-driven structure building is used to derive long-distance agreement.

It now remains to show that the proposal does not overgenerate. In this context, a reviewer brought my attention to extraposition in French, as illustrated in example (39).

(39) Il leur a donné raison, aux présidentes
    He DAT.3PL has given right, to.the presidents.3PL.F
    ‘He said that they were right, the presidents’

Here a full NP (aux présidentes) is right-extraposed, while a co-referring third person plural clitic is moved between the subject and the finite verb. It seems, at a first glance, that the new approach wrongly predicts that the full NP aux présidentes is preferred to the structurally reduced clitic leur for the movement operation. Does specificity thus overgenerate?

The most serious problem that this kind of data poses is that there is virtually no agreement as to how extraposition or clitic doubling or Romance clitics in general are best analyzed. It can be argued on the basis of several different existing analyses that the clitic and the full noun phrase are not potential competitors in the first place: it has been argued, for example, that a full noun phrase is an NP (see e.g. Georgi and Müller 2010), so that the clitic and the full noun phrase are categorically distinct. From another viewpoint, clitics are not verbal arguments in the first place—they are base-generated outside the vP (Sportiche 1995); there is thus no competition for phi-feature agreement in extraposition structures. Furthermore, even if the clitic and the full NP are taken to be competitors,17 there is no approach to extraposition which is

17 If the full NP and the clitic are taken to be potentially competing goals for the phi probe, it is clear that this competition is not based on the presence or absence of particular φ or case features, as there is no structural variation along these lines. I would therefore like to propose that the full NP and the clitic are equally specific in terms of φ-features. The fact that leur does not indicate gender is a sign that the morphological marker /leur/ is underspecified; the syntactic context into which it is inserted does contain a gender feature. All features that appear in a derivation are thus relevant for specificity, even though they need not be spelled out. Elements that are perceived as more specific as they contain more structure or are realized by more specific vocabulary items are, therefore, not necessarily more specific in the syntax.
not controversial (see Van der Linden and Sleeman 2007 for an overview). Moreover, the existing analyses differ in the crucial question of the hierarchical relations between the clitic and the full NP, that is, if the clitic is structurally higher than the full NP, or vice versa. An entirely legitimate position would therefore be that we cannot say anything about these structures in terms of specificity-driven derivation until we have a proper understanding of these phenomena.

Nevertheless, I would like to put forth a proposal as to how examples like (39) can be derived in a specificity-based derivation. The basic idea is that at the point when the phi probe becomes active, the full NP is not a potential goal anymore. This approach reflects the level on which, in my view, specificity applies in the overall grammar system: specificity is a principle of core grammar which is restrained by peripheral conditions. In the case discussed here, a version of the Activity Condition (Chomsky 2000, 2001) seems to be active, namely the requirement that elements cannot be moved on once they are in their final landing site. For example, the wh-elements in multi-wh questions are potential goals for a head C because of their wh-feature. If, in a complex clause, one of the wh-elements is selected by a lower C by means of its [wh]-feature, then it cannot be selected by a higher C by means of this very same feature again; it is not recognized as a potential goal for the higher C in the first place.

With all this in place, I would like to propose the following analysis: Extraposed elements are “low topics” in the sense that they are moved from their base-merge position to the edge of a head c-commanded by I (Cecchetto 1999, Belletti 2005); if there are no functional categories between IP and vP, then this position is the edge of v (Lopez 2009). This movement operation is driven by information structural requirements. The discourse-anaphoric status of the full NP is reflected by a feature [anaphoric] (or [topic]), which is contained in the NP aux présidentes, and which makes it the only possible goal for a head probing for the feature [anaphoric]. For the sake of concreteness, let the probing head be v. The full NP is moved from its base position to the edge of v, thus satisfying v’s structure building feature [v]. The crucial point now is that the full NP has now reached its final position. It can therefore not be moved on (it is “out of the way”). When a higher head now probes for phi-features, it can only attract the clitic, not the full NP; the full NP is thus not a competitor for the agreement operation between I and the clitic.

To sum up, my view on the role of specificity is that the General Specificity Principle is, just like the Minimal Link Condition, a core grammar principle. It can thus be overruled by a number of peripheral constraints (i.e. interface conditions) like the animacy condition briefly discussed above, specificity of the goal (as in order-

---

18 Still, there are some points that are generally agreed on: the extraposed element is merged in an argument position, and it has the properties of a discourse anaphor, which is usually represented by an additional information-structural feature (e.g. Vallduví 1992, Lopez 2009).
preserving multi-wh movement), markedness, or a version of the activity condition. Thus, I assume that the specificity-based decisions made during syntactic derivations can be overwritten by other principles of grammar. I do not take this point as a shortcoming of the new approach; on the contrary. Any theory of grammar must accommodate factors that do not seem to be part of narrow syntax and which seem to disturb the simplicity of the model. Silverstein-hierarchy effects are one such factor, and there is wide agreement that they are independent factors interacting with basic syntactic principles (see e.g. Aissen 1999, 2003). In my view, the best strategy for modeling a grammar is not to try to derive the complex and partly irregular system as a whole (which is virtually impossible), but to carve out the regularities and derive as large a set of them as possible as one natural class of phenomena. Crucially, these core principles may overgenerate if that allows us to cover an overall larger set of data by one and the same principle; the seeming irregularities and counter examples can then be derived as effects of peripheral principles partly overwriting the effects of the core principles. This is where the advantage of the new proposal lies: the General Specificity Principle derives both MLC and anti-MLC effects; in comparison to the Minimal Link Condition, it thus singles out a different set of phenomena as one natural class, and crucially, it also derives a larger set. The principles that interact with the core principle of specificity are, unlike the Equidistance Principle, not principles of CHL, but what one might call interface or peripheral constraints. In my view, this makes it a very attractive alternative to the standard closeness-based set of locality principles.
Structure Building From Below: More on Survive and Covert Movement\(^1\)

WINFRIED LECHNER

13.1 Introduction

In the standard derivational model (e.g. Chomsky 1995b, 2000), movement of a category \(a\) is contingent upon two conditions. On the one hand, \(a\) needs to be endowed with a set of features \(F_a\) that can be identified by the syntactic derivation. On the other, the tree must contain a head \(\gamma\) which bears a feature set \(F_\gamma\) compatible with the features on \(a\), thereby facilitating checking or elimination of \(F_a\) by \(F_\gamma\). This constellation, on which features induce dislocation, in turn provides the basis for two possible interpretations of the movement operation, depending on the nature of the forces that are taken to act upon \(F_a\). If it is assumed that the syntactic relation between \(F_a\) and the trigger for movement is defined in terms of affinity between compatible features, one arrives at the standard Attract model of displacement schematically depicted by (1), in which affinity between \(F_a\) and \(F_\gamma\) causes \(\gamma\) to attract \(a\):

\[(1) \quad \gamma \text{ attracts } a: \quad \gamma F_\gamma \cdots a F_a \Rightarrow a F_a \quad [\gamma F_\gamma \cdots t_a] \quad \text{(where } F_a \text{ and } F_\gamma \text{ are compatible)}\]

But movement can also be described as a consequence of repulsion of \(F_a\) by an incompatible set of features in a designated graph theoretic relation to \(F_a\). On this alternative view, represented by (2), nodes that move are repelled from their local environment and pushed into higher positions of the tree in order to avoid a feature mismatch. In (2), repulsion between \(F_a\) and \(F_\beta\) forces \(a\) to leave the local domain of \(\beta\) (\(\boxtimes\) marks incompatibility):

\[^{1}\text{I am indebted to Klaus Abels, Elena Anagnostopoulou, Mike Putnam, Henk van Riemsdijk, Uli Sauerland, two anonymous reviewers, as well as the audiences of GLOW XXX in Nantes, the Roots 2009 Workshop in Stuttgart, and the members of the Athens Reading Group in Linguistics for helpful discussion and comments.}\]
The most recent incarnation of this concept, which transposes the notion of a push chain familiar from phonology into natural language syntax, has gained prominence under the signature of The Survive Principle (TSP; Stroik 1999, 2009, Putnam 2009).  

The present contribution aims at exploring some aspects and consequences of TSP and the repulsion-based approach towards structure building by movement. Central for these purposes, serving as a cynosure guiding the discussion, will be the search for two desiderata: (i) the proper formulation of TSP; and (ii) criteria which aid in distinguishing TSP from orthodox theories in which movement is modeled in terms of feature attraction. As for the first objective, specifically three questions need to be addressed in order to render the Survive model sufficiently precise for being submitted to empirical verification.

To begin with, the feature sets \( F_\alpha \) and \( F_\beta \) on two nodes \( \alpha \) and \( \beta \) may vary across different dimensions. Thus, it must be made explicit which types of incompatibility or mismatches trigger dislocation by TSP. In principle, TSP can be formulated in such a way that it either reacts to non-identity, or to the stronger condition that the feature sets do not intersect. These definitions differ inasmuch as only the former one predicts \( \alpha \) to be repelled from its local environment if the feature sets are non-identical yet overlap (i.e. if the values are e.g. \{F_1, F_2\} for \( F_\alpha \) and \{F_2, F_3\} for \( F_\beta \)). Moreover, the triggers for movement might be restricted to Case, also include F-features, or even be extended to semantic properties such as the logical type of an expression. While I have nothing to add on the former issue (see Stroik 2009), it will be argued below that an extension along the latter lines leads to desirable results.

Second, it has to be stated precisely which nodes qualify as possible landing sites for TSP-driven movement, determining the density of the movement paths (to borrow a term from Abels 2003). If only phase edges or cyclic nodes constitute legitimate hosts for intermediate copies, TSP e.g. generates chains that differ substantially from paths legitimated by alternative inceptions on which movement leaves a copy at every possible landing site. A particularly simple view will be defended in section 13.2.

Finally, the system must specify which concrete action to take if the Survive mechanism detects a feature incompatible constellation. For instance, in the environment \( [\beta_{F\beta} \alpha_{F\alpha}] \) depicted by (2) above, it must be decided which of the two incompatible nodes \( \alpha \) and \( \beta \) has to leave its position. Below, it will be seen that this choice actually does not have to be stipulated, but falls out from general principles of interpretability.

---

2 For previous implementations of push chains in syntax see Moro (2007) and van Riemsdijk (1997), among others. As a reviewer points out, the concept of push chains has also been employed by theories that interpret movement as a strategy to avoid local symmetry, e.g. prohibiting two DPs within a single domain (see e.g. Moro 2000, Richards 2001, Lechner 2004b).
Rendering explicit the assumptions in the three domains outlined above is essential for delineating the contours of a TSP-based theory of movement. But apart from these theory internal considerations, explorations into TSP also need to attain the second, larger objective of finding criteria which aid in distinguishing TSP from competing Attract-based models. This leads to another set of questions which are more broadly related to differences in strong generative capacity between TSP and theories that use Attract. For instance, it should be seen whether the two approaches map complex expressions that arguably involve movement to identical sets of trees. If such a study elicits non-trivial results, one might further ask to which extent there are invariant markers that discriminate between the structural representations produced by TSP and those predicted by Attract-based models. For instance, do the two systems differ in the number and position they postulate for intermediate landing sites? Another important, partially related issue is whether both theories of movement are equally consistent with representational models of the grammar, or if one of them displays a disposition towards a derivational implementation.

Since isolating sufficiently precise answers for these and related exigent questions depends on various complex factors, which include the choice of axioms for each system and an explicit separation between ancillary and core hypotheses, and since the interaction among these factors generates a large number of combinatorial options to consider, there is to date no comprehensive, systematic contrastive study of Attract vs. TSP. As a result, actual results in this area are sparse. While it has, for instance, been shown that some restrictions on movement can be reanalyzed in terms of TSP (see Stroik 2009, and contributions in Putnam 2009), evidence that there are phenomena which are only compatible with a TSP-based approach towards displacement has only been produced to a very limited extent so far.

In Lechner (2009), it was argued that criterial evidence of the type referred to above materializes in the shape of a particular set of restrictions on movement operations, first discussed in combination in Sauerland (2000), which are characterized by two properties. On the one hand, these dislocation operations are not driven by the need to establish a checking relation, but take place in order to avoid mismatches—in this particular case mismatches in the logical type of the expressions involved. Thus, the moved categories react exactly to the kind of incompatibility predicted by TSP to induce dislocation. On the other hand, the interaction between more than one movement operation results in configurations the generation of which requires the assumption of dense movement paths. This provides a strong argument for TSP, as only a TSP-based analysis forces movement to stop in each available intermediate landing site. In that paper, I also suggested a particular definition of the TSP which had the effect of providing a unified account of these properties.

The current contribution expands on Lechner (2009), extending the original account in four directions. First, I will propose a simpler and, arguably, more principled version of the TSP which renders various properties and consequences of the analysis more
transparent. Second, new TSP-based analyses will be included that have not been part of Lechner (2009). Third, the TSP will be contrasted more sharply with Attract-based theories of movement, focusing on the role of intermediate landing sites of movement. Finally, I will elaborate on various open ends and contentious issues, none of which were addressed in the original version, and try to clarify some probably less apparent aspects of the analysis. For some of these problems, I will suggest speculative answers, while a systematic investigation of others will have to await another occasion, pointing in the direction of future investigations into the nature of TSP.

13.2 Defining repulsion

Assume that two nodes are feature compatible just in case all their features match ((3a))$^3$, and that matching is understood as in (4):

(3) a. Two nodes $\alpha$ and $\beta$ are feature compatible iff the feature sets of $\alpha$ and $\beta$ match.
   b. Two nodes $\alpha$ and $\beta$ are feature incompatible otherwise.

(4) Two feature sets $F_\alpha$ and $F_\beta$ match iff they have the same$^4$ members.

What both Attract- and TSP-based theories have in common is that they use displacement as a strategy for obtaining feature compatibility between nodes in designated syntactic configurations (sisterhood or spec head relation). The main conceptual difference between the models resides in the way they resolve feature conflicts once a node $\alpha$ ends up in a position where it is feature incompatible with its local environment (sister or head). In Attract-based definitions of movement, this task falls to a system which searches for a c-commanding feature compatible attractor $\gamma$ (on most conceptions a head), as expressed by the familiar definition in (5):

(5) $\text{Movement}_{\text{Attract}}$
   For any nodes $\alpha$ and $\gamma$:
   a. If $\alpha$ is the closest node to $\gamma$ and $\alpha$ is feature compatible with $\gamma$ remerge $\alpha$ with a projection of $\gamma$.
   b. A node $\alpha$ is the closest node to $\gamma$ iff $\gamma$ c-commands $\alpha$ and there is no node $\beta$, s.t. $\gamma$ c-commands $\beta$ and $\beta$ c-commands $\alpha$.

$^3$ It might also turn out that compatibility is better formulated in terms of the proper subset relation. For instance, the discussion in Stroik (2009: 49–53) suggests that $\alpha$ is compatible with some head $\beta$ if the features of $\beta$ are a subset of the features of $\alpha$, but not v.v.

$^4$ Stroik (2009) argues that the sameness relation ignores whether the features have been checked. Thus, if $\alpha$ and $\beta$ both bear feature $F_\gamma$ and $F_\gamma$ has been checked on $\alpha$ but not $\beta$, the two nodes are nonetheless compatible. As far as I can see, nothing bears on that specific choice, though.
But local feature incompatibility can also be resolved by the alternative requirement that nodes with incompatible features leave their position in order to escape feature conflicts. On this view, movement is causally linked to the relation of feature incompatibility instead of compatibility. One implementation of this idea comes in the shape of The Survive Principle, which defines movement along the lines of (6) (adapted, with minor modifications, from Stroik 2009):

(6) Movement\textsubscript{Survive}  
(adapted from Stroik 2009: 45, (28))

For any nodes \(\alpha\), \(\beta\) and \(\gamma\):

(i) merge \(\beta\) with a new head \(\gamma\) and

(ii) remerge \(\alpha\) with a projection of \(\gamma\).

Whenever the context of feature incompatibility is met, (6) regulates movement in two steps. First, a new head \(\gamma\)—henceforth also referred to as trigger—is merged with a projection of \(\beta\) ((7b)). Then, \(\alpha\) is expelled from its original position, re-merging with a projection of the trigger \(\gamma\) ((7c)):

For theory-internal as well as empirical reasons, the definition in (6) was generalized in two directions in Lechner (2009). To begin with, restricting triggers to new heads is arbitrary, because not only the addition of higher heads but also first merger of the local head or specifiers potentially changes the feature composition of the newly created root node. Thus, it was suggested to consider any node that the root is combined with as a potential trigger for TSP-induced movement. Second, the notion of incompatibility was argued to include semantic type incompatibility in addition to feature mismatches. This change makes explicit a fundamental and arguably non-accidental similarity between TSP and the principle of type driven interpretation, paraphrased in (8), which expresses the widely held view that certain covert movements are induced by the need to repair type incompatibilities (Heim and Kratzer 1998). In both cases, a category is expelled from its local environment due to incompatibility with its sister node.
(8) **Type driven interpretation**
If a node $\alpha$ is type incompatible, 
move $\alpha$ to the next higher position type compatible with $\alpha$.

(9) a. A node $\alpha$ is **type compatible** iff the denotation of $\alpha$ and its sister can be 
combined by the principles of semantic composition (Function Application, 
Predicate Modification, etc . . . ).

b. $\alpha$ is **type incompatible** otherwise.

In Lechner (2009), it is therefore suggested to formulate TSP in such a way that it 
forces any feature or type incompatible node to be pushed up the tree in the smallest 
possible incremental steps. A definition which renders more precisely this 
generalized concept of Survive, and at the same improves on the version proposed 
in Lechner (2009), can be found in (10). Observe that, just like (8), the final definition 
of TSP in (10) uses an intransitive variant of feature (in)compatibility, as in (11):

(10) **The Survive Principle**
For any nodes $\alpha$, $\beta$ and $\gamma$, such that 
a. $\alpha$ is feature or type incompatible 
b. $\beta$ is the mother of $\gamma$ and 
c. $\gamma$ c-commands $\alpha$
remerge $\alpha$ with $\beta$.

(11) A node $\alpha$ is **feature incompatible** iff there is no node $\beta$, such that $\alpha$ and $\beta$ are in a 
sisterhood or specifier-head relation, and the features of $\alpha$ and $\beta$ match.

Extensionally, (10) divides instances of TSP driven movement into two subcases both 
of which have in common that addition of a trigger $\gamma$ induces dislocation of a locally 
incompatible $\alpha$. If $\alpha$ serves as the complement of a head $\gamma^\circ$, as in (12a), $\alpha$ will be 
pushed to a projection of the trigger $\gamma^\circ$. To see how this follows from (10), note that 
by assumption, $\alpha$ meets clause (10a). Furthermore, in accordance with (10b), $\beta$ is the 
mother of $\gamma^\circ$. Finally, $\gamma^\circ$ c-commands $\alpha$ in (12a), satisfying clause (10c). As a result, 
TSP dictates that $\alpha$ be recombined with $\beta$.

---

5 Whether remerger consists in a node traversing movement operation or copying into a buffer (Stroik 
2009) is immaterial for present purposes. (10) describes a well-formedness condition on trees that is—at 
least in isolation—also compatible with a representational framework. Derivational properties will enter 
with (21) below, though.
In the second possible scenario, depicted by (12b), $\alpha$ and the trigger $\gamma$ are not in a sisterhood relation, but $\gamma$ serves as a specifier or an adjunct that asymmetrically c-commands $\alpha$. This constellation falls under clause (10c). Moreover, since $\beta$ is the mother of $\gamma$, $\alpha$ moves and is internally merged with the root $\beta$, where it lands in an outer specifier or adjunct position.6

Definition (10) primarily differs from Stroik’s version (s. (6)) in four aspects. First, for Stroik, the trigger for movement can never be identical to the node that $\alpha$ is feature incompatible with. The new version admits such combinations in the form of (12a), where both functions are collapsed into the single position occupied by head $\gamma^0$. (The trigger and the incompatible node can also be dissociated, as in (12b).) Second, TSP as given in (10) reacts now to type as well as feature incompatibility. Third, triggers for movement are not restricted to heads but include now also specifiers and adjuncts, resulting in a less artificial system. Finally, (10) treats first merge and second merge (movement) alike in that nodes inserted by either operation can trigger movement. For Stroik, only first-merged heads (“new head” in (6)) possess this ability. Taken together, these are important simplifications which remove imbalances and asymmetries from TSP each of which would require independent motivation. In addition, the last triple of properties will be seen to considerably widen the empirical range, thereby increasing the explanatory strength of any TSP-based theory.

13.3 Quantifier scope restrictions

In Lechner (2009), I explored some empirical consequences of TSP in general and (10) more particularly, concluding that (10) receives support from the observation that it offers the first unified analysis of two complex scope restrictions in English. I will briefly describe these two basic paradigms in turn, and then expand on the treatment of the relevant contrasts in terms of (10). Further details of the analysis and reasons why alternative Attract accounts fail can be found in Lechner (2009).

6 Note that (10) does not need to include a minimality clause, limiting the upper bound of $\gamma$. Hypothetical cases in which $\alpha$ and $\gamma$ are separated by an intervening $\delta$, leading to non-local movement of $\alpha$ across $\delta$, do simply not arise, because each intervening $\delta$ also constitutes a trigger, and therefore forces local movement.
The first scope restriction comes in the shape of the well-known observation that in the English double object construction, the indirect object (IO) and direct object (DO) do not permute in scope:

(13) I gave a (different) child each doll. \[ \exists \succ \forall / \forall \succ \exists \] (Bruening 2001: (2a))

Bruening (2001) pointed out that the prohibition on QR across another quantifier for the DP cannot be expressed as an absolute constraint because DO may scope over the subject, as illustrated by (14). Rather, the condition must be formulated in such a way that it forces IO and DO to move in an order preserving fashion in (13).

(14) A (different) teacher gave me every book. \[ \forall \succ \exists \] (Bruening 2001: (28))

Furthermore, Sauerland (2000) observed that the subject may scopally interfere in between IO and DO ((15)). This demonstrates that a successful analysis cannot simply reduce order preservation to a requirement of string adjacency between IO and DO, but must be flexible enough to generate orders in which the subject is free to scope in intermediate positions.

(15) Two boys gave every girl a flower \[ \forall \succ 2 \succ \exists \] (Sauerland 2000: (49))

In (13), the two quantifiers that were seen to be frozen in scope stand in the relation of c-command. The second configuration that is diagnostic of the virtues of TSP is the Inverse Linking construction, exemplified by (16). Inverse Linking systematically differs from (13) in that it construes two quantifiers in a containment—instead of a c-command—relation. Moreover, as initially noted by Larson (1987), contexts such as (16) are special in that the embedded quantifier every city can obtain scope above the direct object it is embedded in ((16a, b)), but only on the condition that the subject does not scopally interfere between the inversely linked quantifier and its container ((16c); see Sauerland 2000, 2005, and Heim and Kratzer 1998: 234):

(16) \[ QP_1, \text{Two policemen spy on } [QP_2, \text{someone from } [QP_3, \text{every city}]] \]
   a. \[ 2 \succ \forall \succ \exists \] (inverse linking, wide scope for subject)
   b. \[ \forall \succ \exists \succ 2 \] (inverse linking, narrow scope for subject)
   c. \[ *\forall \succ 2 \succ \exists \] (inverse linking, intermediate scope for subject)

Descriptively, Quantifier Raising (QR) of the lowest object may skip one—but not more—quantifiers in the inverse linking construction, whereas in the double object frame (13), the direct object must not cross even a single quantifier. In what follows, it will be demonstrated that TSP derives these two scope restrictions from two ingredients: (i) a metric that regulates precedence among multiple movement operations; (ii) the existence of intermediate traces in non-edge positions of phrases.

13.4 TSP and movement

Before turning to a synopsis of the concrete TSP analysis, the current section will spell out in some detail how TSP builds up structure in derivations that involve scope fixing.
movement operations. To that end, it will be shown first that in a system based on (10), covert, type-driven QR results in the creation of dense chains with numerous intermediate landing sites. Subsequently, in section 13.4.2, a closer look at syntactic schemes in which more than one term has been displaced will reveal that (10) predicts order preservation effects for these contexts of multiple movement. Based on this finding, section 13.5 proceeds to the TSP account for the two scope restrictions of section 13.3.

13.4.1 TSP and single movement

TSP expels a category from its immediately dominating node if that category is not feature or type compatible in its local environment. The current subsection illustrates how TSP accounts for type-driven object QR in structures such as (17a). On standard assumptions (Heim and Kratzer 1998), the object quantifier (every book) of (17a) cannot combine with the denotation of its sister node (read) because the two meanings are not type compatible. As a result, TSP forces the object to move in small, incremental steps up to a (propositional) node it can combine with. (17b) details the TSP-driven derivation.

First, *every book* adjoins to VP, triggered by merger of the verb. Given that the generalized quantifier cannot be interpreted as the sister of VP, which denotes a two-place relation, it must move on, stranding the e-type trace t. Observe that while the object itself is not interpretable in this intermediate position, its e-type trace is, ensuring that the representation is compositionally interpretable. Next, \( \text{v} \) is merged with VP. This operation induces further movement of the object to \( \text{vP} \), because irrespective of the precise semantic contribution of \( \text{v} \), the complex \( [\text{vP}] \) does not result in a denotation that *every book* can apply to. Such a configuration is eventually created once the subject *John* is merged and *every book* has moved into its final landing site right below \( \text{vP} \).

Three important points of clarification are in order here. First, given that compatibility is a symmetric relation, one might wonder (as a reviewer did) why type conflict in representation (17) is resolved by movement of the object and not by expelling the verb. Notably, a condition with this effect does not have to be stipulated but follows from the fact that verb movement simply fails to provide a directly interpretable structural representation. There is no standardly sanctioned operation that would allow the verb to bind a trace of a type \( (<\text{e},\text{t}>, \text{e}, \text{t}>) \) which would be type compatible with the quantified object. Thus, TSP unambiguously and correctly predicts displacement of the object. This interpretability criterion generalizes to all other cases under consideration.

Second, just like all theories that use typed logical forms, the current system presupposes that type information is accessible to the syntactic computation. This claim is not inherently incompatible with the syntactic autonomy hypothesis if it is assumed that natural language syntax interacts with a Deductive System (Fox 2000) and that this system also specifies the logical type of the expressions it manipulates. As a result, type-driven TSP can apply in syntax, even though it is motivated by properties of the formal meta language.

Third, TSP does not work in isolation but is part of a larger group of conditions that regulate the logical syntax of scope relations. For instance, the representations generated by TSP are also evaluated by other, independent components of the interpretive system such as Scope Economy (Fox 2000), which admits scope shifting operations only if they produce truth conditionally distinct readings. Thus, TSP overgenerates by creating representations which are then weeded out by independent principles.

Finally, there are also contexts in which TSP undergenerates. On a commonly held view, the ambiguity of the Antecedent Contained Deletion in (18) is related to structural factors (Fiengo and May 1994). While QR of the object QP *every book*...
John did to the embedded clause produces the narrow ellipsis reading ((18a)), wide ellipsis correlates with long object shift into the matrix clause ((18b)):

(18) I wanted to read every book John did ⊗
   a. I wanted to \{every book John did ⊗\}_1 \text{read } t_1 \quad \circ = \{\text{VP read}\}
   b. I \{every book John did ⊗\}_1 \text{wanted to read } t_1 \quad \circ = \{\text{VP wanted to read}\}

Evidently, the TSP analysis, which only inspects the logical syntax of the object language, lacks the capacity to generate the wide ellipsis representation, indicating that there must be an additional mechanism (such as Scope Economy) at work which carries certain quantifiers into higher domains. I have to relegate a more thorough study of the interaction between TSP with independent components to another occasion.

13.4.2 TSP and multiple movement
The contexts which are of interest for present purposes involve scope restrictions on the relative order of more than one QP, and accordingly involve more than one application of QR. Given that syntax is an informationally encapsulated system, the theory of movement must contain all the information necessary to determine the order of these operations and the positions in which the quantifiers land.

In general, TSP generates two different types of multiple movement configurations and it will be convenient to keep these two types apart for expository reasons. In the first case, to be referred to as trigger movement, an incompatible node \( a \) moves. Then, a trigger \( \gamma \) is added which itself is not type or feature compatible. As a result, both \( a \) and \( \gamma \) have to undergo further movement. In the second scenario, which will be identified by the term multiple target movement, two nodes \( a_1 \) and \( a_2 \) move. In a second step, the tree is expanded by a trigger \( \gamma \). Finally, both \( a_1 \) and \( a_2 \) raise across \( \gamma \).

The two constellations crucially differ in that the trigger \( \gamma \) is displaced in contexts of trigger movement, but remains immobile with multiple target movement.

(19) details how the derivation of trigger movement unfolds. In (19a), \( a \) has moved and a trigger \( \gamma \) is merged above \( a \). Assuming that both \( a \) and \( \gamma \) are incompatible in their local environment, TSP induces two order-preserving movement steps. First, \( a \) crosses over the trigger \( \gamma \), as illustrated by (19b). Next, in (19c), \( \gamma \) is expelled from its base position, with \( a \) serving as the trigger for TSP-induced movement:

\( ^8 \) It is irrelevant whether the trigger is first-merged or has been moved into that position. In fact, an instance of the latter option will be encountered in (22d) below.
At this point, a problem emerges. Evidently, something must prevent configuration (19c) from feeding further movement of $\alpha$ across $\gamma$, followed by another application of TSP to $\gamma$, and so on, leading to endless regress. One plausible venue to pursue in order to avoid this halting problem consists in adopting the assumption that iterative application of TSP to one and the same pair of nodes in one and the same environment is banned. Intuitively, this prohibition can be seen as a more general consequence of (i) feature-based theories of movement, on which the derivation essentially tests hypotheses about featural environments and (ii) the conjecture that for every context and set of features, every hypothesis is only tested once; this second requirement presumably falls out from economy. For TSP, the above implies that if a (pair of) node(s) has moved into a particular context and if this context has been found incompatible with the specific featural requirements of the nodes involved in all possible structural constellations, indicating that further movement does not improve compatibility, the computation is halted. The derivation proceeds then to the next higher level, at which the tree is expanded by new nodes, distinct from $\alpha$ and $\gamma$, which potentially entail changes in the feature/type composition of the expression. On this conception, to be adopted here, the TSP does accordingly not induce further movement once it has reached (19c). As a result, endless regress can be avoided.

Trigger movement preserves the original order of the expressions involved, and it does so by moving the lower category—in this case $\alpha$ in (19b)—first, followed by movement of the higher one. Thus, this type of multiple dislocation displays what will also be referred to as a low - high signature. It contrasts in this respect with the order of movements in the second possible configuration of multiple movement, namely multiple target movement.

In the initial step of multiple target movement, schematically rendered by (20), two nodes $\alpha_1$ and $\alpha_2$ have moved, followed by insertion of an immobile trigger $\gamma$. Moreover, movement of $\alpha_2$ precedes movement of $\alpha_1$.9

9 Typically, the initial configuration (20) is provided by the output of trigger movement ((19c)); this property is immaterial for the further development of the derivation, though.
At this point, a decision has to be made concerning the future development of (20). Since both $\alpha_1$ and $\alpha_2$ are incompatible, and since both reside within the c-command domain of $\gamma$, either of them is eligible for movement across $\gamma$. Thus, the derivation must be guided as to which of the two nodes is targeted by the next application of TSP. I will assume that order of movement is determined by the simple, natural algorithm in (21):

(21) *Last in - First out Edict (Lafite)*

Move the category that was last manipulated by the derivation first.

In essence, the *last in - first out* principle (*Lafite*) (21) encodes the claim that movement is modeled in terms of a push down automaton, that is a computational device that can remember and manipulate the last input signal, but not reach into the derivational history of the tree. For present purposes, this entails that the rules which generate movement keep track of the last category they have applied to and retrieve this category, resolving potentially ambiguous contexts such as (20). With this instruction and given that $\alpha_1$ has been merged last, TSP forces displacement of $\alpha_1$ first, as illustrated by (22b). (For ease of readability, the node which has been manipulated last is marked by underlining.)

(22) *Multiple target movement (high - low signature)*

a. $\alpha_1$ and $\alpha_2$ land below $\gamma$, and $\alpha_1$ moved last

b. $\alpha_1$ moves first (due to (21))
c. $\alpha_2$ moves next across $t_1$ and $\gamma$

\[
\begin{array}{c}
\alpha_1 \\
\beta_3 \\
\alpha_2 \\
\beta_2 \\
\end{array}
\]

\[
\begin{array}{c}
\gamma \\
\beta_1 \\
\end{array}
\]

\[
\begin{array}{c}
t_2 \\
t_1 \\
t_2 \\
\end{array}
\]

In the next step (22c), $\alpha_2$ moves across $\gamma$, presumably after having crossed over the lower copy of $\alpha_1$. (The latter depends on whether copies are taken to be legitimate triggers of movement.) Note that the definition of TSP given in (10) forces this second application of movement to land below, and not above, the landing site of the first one, resulting in order preserving “tucking-in” (Richards 2001b). More precisely, tucking-in follows because (10) requires movement to proceed in the smallest possible steps and because at the point of the derivation depicted by (22c), the values for the trigger of movement for $\alpha_2$ are limited to $t_2$ and $\gamma$. Thus, the definition of TSP adopted here derives order preservation effects.

Finally, once $\alpha_1$ and $\alpha_2$ have reached their position above $\gamma$, they end up in a configuration that matches the profile of trigger movement (19) (where $\alpha_1$ substitutes $\gamma$ in (19), and $\alpha$ is rewritten as $\alpha_2$). Moreover, since $\alpha_1$ and $\alpha_2$ have crossed over $\gamma$, the nodes have been transported into a new syntactic environment, satisfying the restriction discussed in connection with the halting problem. The constellation consequently triggers one final round of TSP induced dislocation. Specifically, movement targets first $\alpha_2$ and then $\alpha_1$, as in (22d), yielding an output configuration that preserves the order and $c$-command relations of the initial input representation (22a).

Even though these last two steps might seem innocuous or vacuous at first sight, because they do not contribute to a change of the relative order between $\alpha_1$ and $\alpha_2$, they have an important consequence in conjunction with the Lafite ((21)). In (22c), it was the lower node $\alpha_2$ that was underlined and that had moved last. Without (22d) and all things being equal, Lafite would therefore lead one to expect that $\alpha_3$ is targeted by TSP next, as illustrated by the counterfactual derivation in (23). Thus, the lower node $\alpha_3$ would have to relocate prior to $\alpha_1$ (see (23a)), with subsequent movement of $\alpha_1$ in (23b):
This alternative derivation crucially differs from (22) in that it leads to an order-reversing configuration. It becomes obvious now why the two trigger movement steps in (22d) are essential. Trigger movement reassigns the underline mark to the higher node $\alpha_1$, since in (22d) it is $\alpha_1$ which has been last manipulated by the derivation. As a result, Lafite demands that any further displacement operation apply to $\alpha_1$ first. Merging a trigger with $\beta^2$ at a later step of the derivation accordingly sets in motion once again the battery of operations (22b) to (22d), producing an order-preserving output in which $\alpha_1$ attaches above $\alpha_2$. And this is, as will be seen shortly, the way it should be. Thus, the final application of trigger movement predicted by the TSP is instrumental in capturing order preservation effects.

To recapitulate, both trigger and multiple target movement preserve order, but they do so for slightly different reasons. In the former contexts, the lower node, that is the node that moves up to the trigger, moves first, followed by movement of the trigger itself to the root. In the latter case, in which two nodes characteristically move close to a trigger, the TSP derivation specifies that the higher node moves first, followed up by the lower one. A final pair of movements renders the higher node the one which was manipulated last, opening up the possibility of further order-preserving movement.

The section to follow applies the abstract results gained so far to the analysis of the two interpretive restrictions introduced in section 13.3. In addition to providing a more systematic and thorough exposition of the TSP account of Lechner (2009), section 13.5 expands the empirical scope of the analysis in two directions by (i) including the DP-PP frame and by (ii) addressing scope freezing with VP-fronting. The latter aspect of the analysis is of independent interest, as it renders possible speculations about the source for cross-linguistic variation in the mapping from structure to interpretation.
13.5 Analyzing scope restrictions with TSP

In what follows, I will adopt a single output model of the grammar (Bobaljik 1995b, Groat and O’Neil 1996, Pesetsky 2000), in which overt and covert movement operations do not differ in their relative timing (pre- vs. post-spell-out), but apply in a single cycle and are distinguished only by whether the higher or the lower movement copy is pronounced. QR accordingly proceeds in overt syntax, the only difference to regular movement being that it does not have any phonological consequences. For evidence in support of the phonological theory of QR see also Fox and Nissenbaum (1999).

13.5.1 The double object construction

All known accounts of scope freezing in the double object constructions (13), repeated from above, rest on the assumption that both internal arguments undergo QR to a position above the base position of the subject.

(13) I gave a child each doll.  ∃ > ∀/#∀ > ∃  (Bruening 2001: (2a))

As this requirement is most straightforwardly expressed by adopting an ordered argument approach towards verb denotations, it will be assumed that give denotes a three place relation. Situation and/or event variables will be ignored throughout.10 Furthermore, the predicate and its internal arguments are parsed into a tree that maps precedence to c-command in the familiar way (Larson 1988).

The goal of the further exposition consists in establishing that the full paradigm of possible and impossible scope readings falls out from the TSP-based algorithm presented in section 13.4. In short, it will be seen that the order-preserving property of TSP is essential for understanding why IO and DO do not permute in scope in the double object frame (13). Moreover, the observation that subjects may be construed with scope inbetween IO and DO ((15)) provides support for the core assumption underlying the TSP that all movements proceed in smallest possible steps.

(15) Two boys gave every girl a flower  ∀ > 2 > ∃  (Sauerland 2000, (49))

The derivation of the double object frame exemplified by (15) starts in (24a). V° and DO are merged, and since DO is type incompatible, TSP requires it to move to the mother of V°. (Bracketed Greek letters refer to the variables in definition (10).) Next, as illustrated by (24b), IO is added, resulting in a tree that contains now two type

10 In Lechner (2009), I assumed a decompositional analysis for double object constructions, following Beck and Johnson (2004), without providing details of the semantics, though.
incompatible nodes (IO and DO). This configuration has been shown to give rise to trigger movement (cf. (19)).

As detailed by the graphs in (24c) and (24d), trigger movement ejects the lower of the two nodes (DO) first, followed by movement of the higher one (IO). Thus, (24c) and (24d) display the low - high signature typical of trigger movement.

The next triple of representations in (25) captures the evolution of the tree subsequent to merger of $v^\circ$, but before the subject has joined the derivation. Adding $v^\circ$ in (25a) results in a constellation which triggers multiple target movement ((22)), because the tree contains an immobile node ($v^\circ$) above two mobile ones. Moreover, IO was the last node to be manipulated by the derivation in (24d). According to Lafite ((21)), IO therefore has to cross over $v^\circ$ first, followed by order-preserving movement of DO, as shown by (25b).
At this point, IO and DO are neighbors in a new local environment, above \( v^\circ \). But this time, DO is underlined. As a result, the TSP induces the two additional movement steps shown in (25c). These two operations are crucial in that they pass back the underline mark from DO ((25b)) to IO ((25c)). Hence, it will be IO that moves next.

Tracing the further development of the tree, (26) depicts the consequences of combining (25c) with the subject. Once the external argument is merged in (26a), IO and DO move by the now familiar pattern of multiple target movement. IO moves first, as it was merged last ((26b)), and DO tucks in ((26c)). What is of particular significance is that both quantifiers are now for the first time located in type compatible positions. Thus, TSP ceases to force further displacement for the internal arguments:

\[
\begin{align*}
(25) & \quad \text{a.} \quad vP \\
& \quad \uparrow \quad vP \\
& \quad \quad \uparrow \quad VP \\
& \quad \quad \quad \uparrow \quad IO_2 \\
& \quad \quad \quad \quad \uparrow \quad DO_3 \\
& \quad \quad \quad \quad \quad \ldots \\
(25) & \quad \text{b.} \quad vP \\
& \quad \uparrow \quad IO_2 \\
& \quad \quad \uparrow \quad vP \\
& \quad \quad \quad \uparrow \quad DO_3 \\
& \quad \quad \quad \quad \uparrow \quad v^\circ \\
& \quad \quad \quad \quad \quad \ldots \\
(25) & \quad \text{c.} \quad vP \\
& \quad \uparrow \quad IO_2 \\
& \quad \quad \uparrow \quad vP \\
& \quad \quad \quad \uparrow \quad DO_3 \\
& \quad \quad \quad \quad \uparrow \quad t_2 \\
& \quad \quad \quad \quad \quad \ldots \\
(26) & \quad \text{a.} \quad [vP \text{SUB}_1 ] [vP \text{IO}_2 ] [vP \text{DO}_3 ] v^\circ \ldots \\
& \quad \text{b.} \quad [vP \text{IO}_3 ] [vP \text{SUB}_1 ] [vP t_2 ] [vP \text{DO}_3 ] v^\circ \ldots \\
& \quad \text{c.} \quad [vP \text{IO}_2 ] [vP \text{DO}_3 ] [vP \text{SUB}_1 ] [vP t_2 ] [vP t_3 ] v^\circ \ldots \\
\Rightarrow \text{scope order } 2 > 3 > 1
\]
Since the lowest position in which quantifiers are compositionally interpretable is the position the subject is first merged in, and since no other mechanism (such as QR) licenses scope reversal in the present system, IO and DO can only be assigned scope in that order. It follows that TSP correctly derives scope freezing for the internal arguments of the double object construction.

But the derivation is not complete, yet. More specifically, the subject still bears Case and phi-features that cannot be checked in vP. In a final series of movements, the subject is therefore ejected from its base position due to feature incompatibility, moving in small steps up to T. First, it stops inbetween IO and DO, as shown by (26d). Then, in (26e), it raises up to T.

\[(26)\] d. \[vP \ IO_2 \SUB_1 \ DO_3 \ v^\phi \ldots \ 2 \succ 1 \succ 3\]

\[e. \ SUB_1 \ IO_2 \ t_1 \ DO_3 \ v^\phi \ldots \ 1 \succ 2 \succ 3\]

Crucially, the subject is interpretable in all positions that it has passed through, because all copies are type compatible. As a consequence, the subject can be construed with widest scope ((26e)), intermediate scope ((26d)) or narrowest scope ((26c)). Thus, TSP does not only account for scope freezing, but is also successful in deriving the flexible scope of subjects.

Two remarks on this corollary of the theory are in order here. First, on current conceptions, all scope ambiguities that are to be accounted for in structural terms are derived by optional reconstruction in syntax. This view resonates with ideas articulated, among others, in Hornstein (1995), Johnson and Tomioka (1998), and Lechner (1996).

Second, the ability of TSP to derive scope flexibility of subjects is contingent on the assumption of a single output model in which all operations apply in one cycle. In a conservative model which postpones all QR to LF, overt subject raising would at most strand a trace in a vP adjoined position, as shown in (27a).

\[(27)\] a. \[vP \SUB_1 \ t_1 \ vP \ t_1, \ base \ vP \ t_1, \ base \ v^\phi \]

b. \[vP \SUB_1 \ t_1, \ base \ IO_2 \ DO_3 \ v^\phi \]

c. \[vP \SUB_1 \ IO_2 \ DO_3 \ vP \ t_1, \ base \ v^\phi \]

(27b) schematically represents the point in the hypothetical derivation at which the two object quantifiers have moved across \(v^\phi\). Finally, in (27c), IO and DO land inbetween the base position of the subject and the vP-adjoined trace. However, since the subject has already reached its overt position by spell-out at this point, there is no way to generate the intermediate scope reading \((2 \succ 1 \succ 3; \text{cf. (26d)})\). Thus, the analysis is not compatible with the traditional T-model of the grammar. Conversely, the interaction between conditions on overt and covert movement, which was seen to be responsible for licensing (26d), can be taken to provide novel support for a single output architecture.
13.5.2 The prepositional frame

Unlike the IO DO frame, ditransitive predicates that are parsed into a prepositional construction are not subject to scope freezing (Aoun and Li 1993, and others)

(28) I showed a picture to every student \( \exists \forall / \forall \forall \exists \)

Following Aoun and Li (1993), it will be assumed that (28) obfuscates the base order of the two internal arguments, and that PP is generated above IO, as in (29a) (see also Pesetsky 1995). The surface order is then derived by movement of DO across PP and verb raising. (29b) and (29c) track the relevant parts in the evolution of (28):

(29) a. I to every student\(_3\) showed a picture\(_3\) (base order)
b. I a picture\(_3\) to every student\(_3\) showed t\(_3\) (move DO across IO-PP)
c. I showed\(_4\) a picture\(_3\) to every student\(_2\) t\(_4\) t\(_3\) (verb movement)

There is one specific property which is essential for a TSP analysis to be able to account for the inverse scope reading of (28): the reversal between DO and IO-PP has to take place above the subject, i.e. in a position in which quantifiers are interpretable. This requirement is e.g. compatible with the widely accepted view articulated by Johnson (1991) that there are functional projections inbetween \( vP \) and TP and that these additional nodes may host overtly moved categories, among them the PP in (29b). In what follows, I will remain agnostic about the details of the derivation and restrict myself to spelling out the steps in more detail which lead to the conclusion that inversion must take place high, i.e. above the base position of the subject.

In the PP-frame, DO is generated as a sister to \( V^\circ \), while PP originates in some higher c-command position. Since DO denotes a quantifier in (28), it climbs up the tree in search of a type compatible position. The tree including DO and PP accordingly looks as in (30a). (30a) also depicts a second, independent movement operation, which separates the quantificational IO from the preposition. Since this step, which is also motivated by type mismatch, takes place in a separate part of the derivation (“work space”) it is DO-movement, and not P-stranding that counts as the last operation performed on the tree.\(^{12}\) Lafite ((21)) therefore dictates that the trigger movement configuration is resolved as in (30b) by moving DO first, followed by IO-raising:

---

\(^{11}\) The question whether the IO DO frame and the DO PP construction are derivationally related is orthogonal to present purposes. Note incidentally that the two analytical options are not mutually exclusive: some IO DO (or DO PP) orders could be derived, while others could be base generated.

\(^{12}\) The relations between IO and DO are structurally identical to the ones which hold between the inversely linked object and the subject in (31b) and (32a).
Merging the subject followed by multiple target movement yields (30c) ($v^\circ$ and subject raising suppressed). In the crucial inversion step (30d), DO moves across the indirect object. Shifting the verb to the left of DO finally results in the surface order $V^\circ \wedge DO \wedge PP$.\footnote{Note that IO raises covertly. Movement of DO is presumably Case driven and lands in a higher functional projection XP.}

The final order-reversing movement of DO in (30d) entails an important consequence. Since inversion between IO and DO takes place in an area of the tree in which quantifiers are interpretable, DO may be assigned scope either above or below IO, depending on

\footnote{Note that IO raises covertly. Movement of DO is presumably Case driven and lands in a higher functional projection XP.}
whether DP reconstructs into its vP-adjointed trace (marked by \( \triangleleft \)) or not. Scope freezing effects are therefore correctly predicted not to be attested in the DO-PP frame.

To recapitulate so far, the TSP-based system correctly derives the distribution of scope restriction in the double object construction and the prepositional frame. Below, in section 13.5.4, it will be seen that in particular the ability of TSP to provide an explanation for scope freezing is critical, because at the moment, a comparable Attract-based account is missing. (The analysis of the prepositional frame was contingent on an assumption shared by competing Attract theories—namely an additional movement step—and therefore does not help in distinguishing between the competing theories.)

13.5.3 Inverse linking

The first scope restriction was seen to manifest itself in contexts with three quantifiers in asymmetric c-command relation. The second restriction operates on triples of quantifiers, two of which are in a containment relation. Specifically, the goal is to explain why string (16) lacks reading (16c), on which the subject scopally interferes between the inversely linked QP3 and what will be called the container, in this case QP2.

(16) \([\text{QP}_1, \text{Two policemen spy on } [\text{QP}_2 \text{ someone from } [\text{QP}_3 \text{ every city}]]]\)

a. \(2 \triangleright \forall \triangleright \exists\) (inverse linking, wide scope for subject)

b. \(\forall \triangleright \exists \triangleright 2\) (inverse linking, narrow scope for subject)

c. \(*\forall \triangleright 2 \triangleright \exists\) (inverse linking, intermediate scope for subject)

The discussion of the DO-PP frame already revealed that TSP driven movement can affect subparts of the tree that are later joined by a generalized transformation with the spine of the derivation (see (30a)). This type of displacement is also at work in the assembly of \([\text{QP}_2 \text{ someone from every city}]\). Relevant parts are made explicit in (31a). Upon insertion of the predicate (spy on), the whole object moves to the left of the base position of the subject, as detailed by (31b):

(31) a. every city adjoins to container someone

b. Merge two policemen, move container someone

\[
\text{vP}^2
\]
The output of (31b) is followed by two displacement operations. Since subsequent to container movement in (31b), the subject is the category manipulated last, *two policemen*, is ejected from its base position first, resulting in (32a):

(32) a. Move subject *two policemen*  

b. Move inversely linked *every city*

Next, the inversely linked QP *every city* of (32a) needs to relocate for reasons of type incompatibility. As (33) reveals, the structural relation between *every city* (α in (33a)) and the trigger *two policemen*, (γ in (33a)) is identical to the one which was seen to hold in standard instances of dislocation, schematically depicted by (33b). The two trees only differ in the ordering between γ and δ:

(33) a.  

b.  

Due to this structural similarity, the definition of TSP in (10), repeated from above, applies to the inversely linked QP *every city* in (32a) just as it does to regular object QPs.

(10) For any nodes α, β and γ, such that  
a. α is feature or type incompatible,  
b. β is the mother of γ, and  
c. γ c-commands α:  
remerge α with β.
Moreover, (10) also determines the landing site of every city. (10) universally quantifies over occurrences of c-commanding γs. One such γ comes in the shape of the node vP1 in (32a). Hence, every city has to remerge with the mother of vP1, yielding the output representation (32b). Thus, the particular configuration (32), in which a node that is to undergo further movement is transported into its launching site by a container, unveils another, qualitatively new extensional property of (10): TSP leads to tucking-in also if the two movements start from non-commanding positions.

The structural relations encoded in tree (32b) contain all information necessary to generate the attested readings, and all sufficient information to weed out the unattested ones. Concretely, the LF-fragment (32b) maps directly to the scope order subject ≻ inverse linked QP ≻ container (= (16a)). Reconstructing the subject into its base yields inverse linked QP ≻ container ≻ subject (= (16b)). But as TSP failed to instruct the subject to generate an intermediate trace in between the inverse linked QP and its container, it is not possible to map (32b) into the unavailable reading (16c) on which the subject takes intermediate scope.14

Inverse linking is not only consistent with TSP, but also exposes the limitations of any account which computes precedence of operations on the basis of some notion of closeness such as the Minimal Link Condition (Chomsky 1995b). As shown by (34a), the two quantifiers every city and two policemen, are equidistant to the root node—both are separated from the root by a single segment. Alternatively, distance can be measured in terms of complete containment within a category. On this conception, every city is even closer to the root than the subject, because the former is only dominated by one of a multi-segment category, whereas both vP segments dominate two policemen.

(34) a. Subsequent to movement of inversely linked QP (every city)

b. Move subject to closest landing site

14 I assume that the mapping from LF representations to meanings is injective (“into”) in order to provide for the possibility that there are scope orders that are not solely determined by the structure of LF. The usual suspects include wide scope indefinites, de se-like readings, and branching quantification.
Crucially, no matter which definition is chosen, it is not the subject which qualifies as the closest node to the root. But this entails that derivation (34b), which underlies the unattested scope order inverse linked QP $\triangleright$ subject $\triangleright$ container should be legitimized by the grammar. It follows that a decision procedure on the order of movement which relies on the concept of closeness—to be precise, closeness to the landing site—is bound to fail.

Before turning to further empirical extensions of the TSP analysis, it is instructive to stop and pursue the point just taken up further by comparing how TSP- and Attract-based models treat multiple movement configurations. Even though a systematic and complete evaluation of the fundamental differences between these two alternative strategies for inducing displacement will have to await a future study, the comments in the next subsection aid in eliminating some logically possible theoretical options.

### 13.5.4 TSP vs. Attract

The TSP analysis outlined above rests on two generalizations about the way contexts that involve multiple movement are organized. First, if the mobile nodes asymmetrically c-command one another, movement proceeds in an order-preserving way. Second, configurations in which one operator is embedded inside another one are subject to the condition that no third operator may interfere. I will briefly comment on some complications these generalizations pose for competing Attract-based models, addressing the concept of closeness and intermediate traces in c-command and containment configurations in turn.

Attract-based models delegate decisions about the order of movement and the location of the landing sites to a metric which requires features to attract the closest compatible target. In principle, it would also be possible to define such a condition for systems that employ feature incompatibility as a trigger for dislocation, transposing the closeness requirement into an Attract-based framework. Suppose more concretely that $\alpha_1$ is the complement of a verbal $\gamma$, and $\alpha_2$ serves as its specifier (in a conservative phrase structure), and both $\alpha_1$ and $\alpha_2$ are incompatible in their local environment, as in (35a). Since $\alpha_2$ is closer in tree geometric terms to $\gamma$ than $\alpha_1$ is, it is possible to define TSP in such a way that it ejects $\alpha_2$ prior to $\alpha_1$:

(a) $\alpha_1 \beta^2 \beta^1 \alpha_2 \gamma$

(b) $\alpha_1, <et,t> \alpha_2, <et,t> \beta^3 \beta^2 \beta^1$

While a closeness condition can at least in principle be formulated for feature mismatches, it is harder to see how the concept of closeness can be employed in a

15 For a (descriptive) generalization that extends beyond quantifiers see below and Lechner, to appear.
meaningful way when dealing with type incompatibilities. Imagine that $a_1$ and $a_2$ are both generalized quantifiers which are type incompatible with their sister nodes, as in (35b), and that movement of $a_1$ precedes movement of $a_2$. Is it not possible to express this relation in terms of relative closeness to a given node that induces type incompatibility? This is so because compositionality imposes the requirement that all semantic composition target sister nodes. Hence, the type theoretic well-formedness conditions only hold for the pair $a_1$ and $\beta^2$, and the pair $a_2$ and $\beta^1$. But these relations are—trivially—equally close. If TSP is to be given its most natural formulation, in which it subsumes type and feature mismatches under a single principle, a closeness metric is therefore, at most, a partially successful guide for the properties of movement.

Next, consider the different predictions which Attract-based models and TSP generate for the position of intermediate movement copies. More precisely, there are two structures to consider. The first context comes in the shape of the intermediate scope representation for the double object construction in (26d), repeated below:

(26) d. \[ \text{VP}_2 \text{IO} \quad \text{VP}_1 \text{SUB} \quad \text{VP}_3 \text{DO} \quad \text{VP}_1 \text{t} \quad \text{VP}_1 \text{v} \quad \text{VP}_1 \text{...} \]

Without additional assumptions, an Attract model leads one to expect that the subject, which originates in SpecvP, directly moves to the next functional projection in the tree. What is entirely unexpected is that the subject lands in between IO and DO as in (26d). Feature attraction approaches have at least not demonstrated yet that they are capable of replicating the results attained by TSP without stipulation.16

In a second relevant set of contexts, two nodes to be reigned by some version of a minimality principle dominate one over the other. As dominance relations cannot be translated into scope relations, the two nodes need to be unfolded into a structure that repositions them into a c-command relation. This is exactly what was seen to be at work in inverse linking. However, the discussion of inverse linking above also revealed that a closeness algorithm fails because it cannot predict the order of movement that results in the only acceptable interpretation. The challenge is therefore once again with Attract-based theories to provide a solution to these puzzles.

To recapitulate, the TSP model offers two advantages over orthodox systems which motivate dislocation by feature attraction: it offers a natural explanation for the observation that overtly moved subjects are flexible, while covertly moved objects are not; and it derives the generalization that movement of embedded, inversely linked quantifiers is strictly local.

13.5.5 VP-fronting

Returning to further empirical ramifications of TSP, there is yet another aspect of the system that has not been recognized in previous versions of the theory. The inverse

16 Sauerland (2000) demonstrated that it is possible to frame a common analysis of double object constructions and inverse linking. For a comparison see Lechner (2009).
linking construction related three nodes that all underwent silent movement. Interestingly, it is also possible to identify movement triads that obey the same rules as inverse linking, but which offer the additional benefit of making one of these operations overt. In English, this configuration manifests itself in the form of scope freezing with predicate fronting, first discussed in Barss (1986). Just as with inverse linking, VP-topicalization in (36) involves three components: a subject (*no one*), a container that moves (the VP), and a node that needs to scope out of the container in order to ensure type compatibility (*every student*; example from Huang 1993).

\[ (36) \quad \ldots \text{and } [\text{VP teach every student, no one}, \text{will } \neg \exists \gamma / \forall \neg \neg \exists \]

Furthermore, the QP embedded in the container cannot be assigned scope over the subject. This restriction is strongly reminiscent of the condition operative in contexts involving inverse linking, which limited the scope options of the quantifier embedded inside the container. Descriptively, these similarities between VP-fronting and inverse linking can be captured as in (37):

\[ (37) \quad \text{If } \beta \text{ contains } \alpha, \text{ and } \beta \text{ moves across } \gamma, \text{ then } \alpha \text{ and } \beta \text{ cannot be separated by } \gamma. \]

On current views, which include the assumption that all movement proceeds in overt syntax, the differences between the two constructions that fall under (37) reside solely in phonological properties. Scope freezing with predicate fronting is therefore a direct corollary of the theory presented so far. The single output model generates structures for (36) which look for all means and purposes just like the ones for inverse linking (cf. (32)). As a comparison between the fragmentary representation for (36) given in (38) and the tree in (32) reveals, the two derivations are identical up to the values of labels and lexical items:

\[ (38) \quad \text{a. Merge container VP, move QP1 to TP} \]

\[ (38) \quad \text{b. Move object OP2} \]
Just as in (32), a quantifier—in this case the object every student—embedded inside a node that itself undergoes movement—in this case VP—is instructed by TSP to shift to the left edge of the container in order to escape a type incompatible environment.\footnote{17} In the next step, represented by (38a), the container VP\(^2\) adjoins to vP. At this point, the derivation needs to decide whether the subject (no one) or the node adjoined to the container moves next. Once again, this procedure is familiar from inverse linking. Given that no one has been manipulated last, the subject raises locally across the VP, driven by feature incompatibility. In a final step, depicted by (38b), the object moves to a position below the subject. In semantics, representation (38b) can then be transparently mapped to the overt scope order \(\neg \exists \supset \forall\).

While similar in many respects, there is one important difference between inverse linking and VP-fronting, though. The subject of the former construction is flexible in scope, and may be interpreted in any position it has moved through. By contrast, no one, in (36) must not reconstruct into its base.\footnote{18} Even though this difference at first sight poses an obstacle to a common analysis, it arguably follows from another, independent restriction on predicate fronting. Concretely, in order to account for the absence of inverse scope readings for VP-topicalization and related constructions, two operations must be banned: wide QR of the object across the surface position of the subject, and subject reconstruction into its base, as stated by (39) (for discussion see Sauerland and Elbourne 2002, among others):

\[(39)\] In contexts of predicate fronting, the subject cannot reconstruct into its base position.

As was seen above, TSP provides a theory of the former restriction, while (39) must be attributed to an independent source. Although I cannot offer a complete analysis at the moment, the absence of subject reconstruction with predicate fronting fits into the broader typological generalization noted in Adger (1994) that (certain) optional movement operations result in representations that are bi-uniquely mapped to interpretation. Thus, if an operation is optional, interpretation is fixed, while obligatory processes typically lead to ambiguity. For instance, languages that admit optional rearrangement of the middle field by scrambling such as German, Korean, or Japanese are typically scope rigid. By contrast, even canonical word orders feed scope inversion in English, which is generally taken to lack scrambling. The same principle correctly predicts that

\footnote{17}{The analysis is also compatible with the view that predicate fronting applies to vP, and not VP. In this scenario, the derivation involves an additional step of short subject movement.\footnote{18}{In this particular example, the absence of a narrow scope reading for the subject might also be due to the general resistance of negative quantifiers to reconstruct. (Partee 1971, Lechner 2006, 2007, Iatridou and Sichel 2009). But scope freezing is also attested with indefinites:

(i) They promised that someone will answer every letter, and answer every letter, someone will.}}
optional operations such as topicalization in English yield unambiguous scope orders. In a way, then, VP-topicalization makes English look like German. The more general question to be answered in the future, accordingly, is why lower movement copies become unavailable for interpretation in German as well as in English once predicates are placed into topic positions.

Interestingly, this preliminary result dovetails with another, related property of the present account, which also links certain aspects of English and German syntax. In the current TSP system, all scope inversion is the result of the interaction between type- and feature-driven movement. For subjects, raising to T e.g. creates a scope position above the object. For the DO-PP frame, the additional movement step of the DO is instrumental in that it generates a representation that can be mapped to two distinct interpretations. Moreover, if feature-driven movement is taken to be always visible in the overt component, it also follows that all scope ambiguity is dependent upon the application of at least one overt movement operation—which can later be undone by reconstruction. English transitive clauses are ambiguous because the subject moves overtly, and the DO-PP frame displays scope permutation due to overt DO shift.

This conception has the intriguing consequence of making English resemble scope rigid languages like German, in which scope permutation is also known to be contingent upon overt inversion of the scope bearing categories. As a result, differences between these two languages in the mapping from syntax to interpretation can be minimized. In addition, it becomes possible to locate more precisely the exact factors responsible for variation. Concretely, what discriminates English from German is the degree to which DPs may undergo total reconstruction in each language. In English, movement is free to reconstruct into all interpretable positions. Subjects and objects in the DO-PP frame can, therefore, be assigned narrow scope also in configurations that observe canonical word order. By contrast, German limits total reconstruction to the left edge of the vP, such that overt movement across this boundary cannot be undone. Ample evidence for this generalization comes from the study of the interaction between scrambling and interpretation (Frey 1993, Lechner 1996, 1998). In what follows, I will synoptically summarize some findings from the literature which support this assumption.

To begin with, transitive subjects in German sentences with canonical word order do not reconstruct below objects for scope. Given that object quantifiers are parsed into a vP-adjoined position, this observation is compatible with the claim that quantifier lowering by reconstruction must not penetrate the left edge of vP. Furthermore, scrambling reconstructs for the computation of c-command sensitive principles (Binding Conditions, NPI-licensing, and others) into a position just to

19 Adger (1995) argues for the same point, based on a different set of observations, though.
the right of the subject (Frey 1993). This restriction manifests itself in two configurations, which are collapsed into the single tree (40). (Assuming IO and DO to be quantificational, they occupy VP-external positions; nothing bears on this issue, though.)

First, short (object-over-object) scrambling never reconstructs. Thus, if DO moves into XP, coreference and binding relations, among others, are evaluated in that position.20 Second, in structures involving medium scrambling of an object across the subject, only the step that places the object above the subject is undone. As documented by (40), both restrictions fall out from the generalization that reconstruction must not “reach into” vP.

The holy grail for research on cross-linguistic variation in strategies for computing meaning is accordingly defined by the question why English and German are subject to distinct conditions on reconstruction (see e.g. discussion in Adger 1995). A possible venue to pursue in future work may proceed along the lines spelled out in the concluding remarks of this section.

Reconstruction restores the descriptive content of categories into lower chain links. Moreover, categories obtain their descriptive content only by lexical insertion. One might therefore try to regulate the “depth” of reconstruction by restrictions on the specific location in the tree at which lexical insertion takes place. The higher lexical insertion takes place, the fewer positions for reconstruction are made available by the derivation. In fact, such a strategy of delayed insertion has already been exploited successfully at various places in the literature by Late Merge accounts of anti-reconstruction effects (Fox 2002, Lebeaux 1995).

20 The observation that in scope rigid languages, all overt inversion feeds ambiguity, is accounted for reconstruction in the semantic component (Lechner 1996, 1998).
The specific implementation best suited for present purposes is provided by Takahashi’s (2006) theory of Whole Sale Late Merge. On this account, the restrictor argument of a determiner can be inserted subsequent to the application of movement into derived positions, resulting in delayed lexical insertion. In such contexts of Whole Sale Late Merge, the determiner raises on its own, schematized in (41a), followed by merger of the restrictor argument in a higher chain link, as in (41b):

(41)  

\( \text{a. Move determiner: } [\text{determiner}_1 \ldots [\text{t}_1] \)  

\( \text{b. Insert restrictor: } [[\text{determiner}_1, \text{restrictor}] \ldots [\text{t}_1] \)

Suppose now that in German, all DPs that have left vP are—for some reason yet to be exposed\(^{21}\)—assembled by Whole Sale Late Merge. Then, there will be no lower movement copies of restrictors inside vP, accounting for the absence of reconstruction into vP (see (40)). In English, lexical insertion of restrictors at least optionally starts in the foot of the chain, providing an extensive repository of potential reconstruction sites inside vP. Thus, movement freely reconstructs in English.

While successful in deriving the basic facts, it is evident that this set of assumptions still requires a stronger theoretical foundation as well as further empirical confirmation in order to be propagated to an adequate theory of reconstruction. I will have to relegate this task to future investigations.

13.6 A note on the density of movement paths

On current assumptions, all movement proceeds in smallest possible steps, resulting in dense movement path. Density might be the property that most radically distinguishes TSP from Attract-based models. In this final section, I will briefly address one of the many consequences of this assumption that merit further inquiry by considering a piece of evidence against dense movement paths recently discussed in Abels and Bentzen (2008).

Abels (2003) observes that the contrast in (42) generates an argument for the view that movement paths are “punctuated,” and not dense.

(42)  

\( \text{a. Which picture of himself, did it seem to John, that Mary liked?} \)  

\( \text{b. *Which picture of himself, did Mary seem to John, to like?} \)

(42b) can be ruled out by Condition A of the Binding Theory if it is assumed that there in an intermediate landing site for \text{which picture of himself,} below John in (42a), but not in (42b). On this assumption, the pair is parsed into a structure as in (43):

\[ \]

\(^{21}\) Naturally, it might be tempting to relate this difference to head-finalness of German VPs and properties of the linearization algorithm.
(43)  a. Which picture of himself, did it seem to John, [CP which picture of himself, that Mary liked]  
     b. *Which picture of himself, did Mary seem [TP to John, [XP to like?]]

No intermediate landing site for which picture of himself.

As the string to the right of John in (43b) is mapped to a tree (XP in (43b)) that could in principle host movement copies, the analysis entails that movement does not pass through every position along the movement path. Specifically, (43b) does not contain a copy of which picture of himself, at the left edge of XP. From this, Abels infers that movement paths cannot be dense.

But as pointed out by Gereon Müller (quoted in Abels and Bentzen 2008), (42b) can also be accounted for by a condition which demands that an anaphor be bound by the closest possible binder. Crucially, this requirement derives the correct results also if movement passes through every intermediate landing site, as in (44), because in (44), the trace of Mary counts as the closest antecedent for the copy which picture of himself. This defuses the argument against density of movement paths.

(44)  *[Which picture of himself,]_3 did [TP Mary, seem [TP to John, [tMary, [which picture of himself,]_3, [TP tMary, to like?]]]

In a reply to Müller’s objection, Abels and Bentzen (2008) bring to attention the fact that anaphor binding is not always determined by closeness (Barss and Lasnik 1986). Thus, they contend, closeness can also not be at stake in (44), reinstating the original argument against density of movement paths:

(45)  a. Mary explained the man to himself.  
     b. Mary explained the man to herself.

However, although correct for the particular constellation (45), this generalization does not extend to the relevant context in (46), where the relation between the anaphor and its antecedent is interrupted by a subject trace, instead of an object. What (46) demonstrates is that closeness is relevant if raising is involved:

(46)  *[John, seems to Mary, [TP t, to like a picture of herself,]]

Moreover, the same condition that a reflexive not be separated from its binder by a subject trace is also violated in (44). It can therefore be concluded that Müller’s objection is still valid, and that the argument against density of movement paths is not conclusive. Needless to say, the fact that one particular argument against a corollary of TSP—namely dense movement paths—apparently fails does, of course, not entail that others will, too.
13.7 Conclusion

In Lechner (2009), a particular version of a theory of movement was presented in which dislocation is motivated by incompatibility with a local syntactic context (TSP; Stroik 2009). The present contribution expands on TSP in various directions. First, a new definition of TSP was provided in (10), which is both simpler and more natural than the version of Lechner (2009). A second central objective consisted in the search for criteria that distinguish between TSP from competing Attract modes. Such diagnostics could be identified in the shape of scope restrictions in two contexts: double object constructions and inverse linking. In both cases, the analysis was contingent upon the interaction of the specific definition of TSP and the assumption of a simple algorithm which informs the derivation about the order operations if TSP is met by more than a single context in the tree. Third, the discussion included new empirical paradigms (the DO-PP frame and VP-fronting) which were seen to lend themselves to a natural analysis in terms of TSP. Finally, I followed up a number of ramifications and consequences of TSP. Among others, it was seen that just like any theory of movement, TSP needs to be supplied by an independent theory of reconstruction in order to guard against generating unattested interpretations. Such considerations led to some speculations about the relation between movement more generally and lexical insertion.
On Transparent Adjuncts in Japanese*

YOICHI MIYAMOTO

14.1 Introduction

In his seminal work, Huang (1982) accounts for “Adjunct Condition effects (ACEs)” under his Condition on Extraction Domain (CED), which prohibits a phrase from moving out of a constituent which is not a complement. In (i) for example, which class is extracted out of the adjunct headed by the preposition during:

(i) ?*Which class did you fall asleep [during t$\_j$]?

(Huang 1982: 499)

With the bracketed phrase being a non-complement, the CED correctly rules out the extraction in point. Under Huang’s proposal, the distinction between complements and non-complements plays an important role in defining what the possible domains for extraction are, and which constituents constitutes a barrier for extraction from within.

In contrast with this view, Demonte (1988) and Borgonovo and Neeleman (2000) show that there are cases where extraction out of an adjunct does not result in deviance, and conclude that adjuncts are not inherent barriers for extraction. This chapter shows that there are also cases in Japanese in which extraction out of an adjunct is permitted, and provides further cross-linguistic evidence for their conclusion.

* I am indebted to Jon Clenton, Yasuhiro Iida, Nobuhiro Miyoshi, Keiko Murasugi, Masashi Nomura, Masao Ochi, Mamoru Saito, Koji Sugisaki, and Yuji Takano, as well as to two anonymous reviewers and to the editors of this volume, Myriam Uribe-Etxebarria and Vidal Valmala for their valuable comments, suggestions, and criticism. I am also grateful to the participants of the 25th Linguistic Colloquium at Nanzan University (July 2008) and the conference “Ways of Structure Building” at University of the Basque Country (November 2008) for their valuable comments and suggestions. The usual disclaimers apply. The research presented here is in part supported by the Grant-in-Aid for Scientific Research (C) (No. 19520341), Japan Society for the Promotion of Science.
The case that this chapter examines is comparative deletion in Japanese. This construction is illustrated in (2):

(2) \[ \text{TP Osaka-de [TP Taroo-ga tomodachi-o [QP [CP Hanako-ga e in friend -ACC -NOM mikaketa-yori(-mo) oozei] mikaketa]](-koto) saw -than(-also) many saw (-fact)\]

‘In Osaka, Taroo saw more friends than Hanako saw e.’

In (2) the Japanese counterpart of the English \textit{than}-clause, the \textit{yori}-clause, is adjoined to the QP \textit{oozei} which corresponds to \textit{many} in English (Ishii 1991). In this chapter, I call the QP \textit{oozei} a “floating quantifier (FQ).” What is relevant for our discussion is that this FQ is predicative in nature, which allows us to duplicate in Japanese the contexts Demonte (1988) and Borgonovo and Neeleman (2000) examine in Spanish, English, and Dutch. I show that in certain well-defined contexts, the Japanese \textit{than}-clause, the \textit{yori}-clause, can be scrambled out of this complex QP (formed by the \textit{yori}-clause and \textit{oozei}) despite the adjunct status that this complex QP has. On the basis of data from Japanese comparative deletion, I conclude that the complement/non-complement distinction is not the sole factor causing ACEs. Following MacDonald (2006, 2008a,b, 2009), I propose an alternative account for the examples involving Japanese comparative deletion, based on Agree with Asp(ect) during the course of structure building. In essence, if a non-complement phrase establishes an Agree relationship with the matrix Asp, the phrase in question does not constitute a barrier and allows extraction from within.

The chapter is organized into the following sections. Following this introduction, section 14.2 presents Demonte’s (1988) and Borgonovo and Neeleman’s (2000) paradigms and the asymmetries in extraction out of adjuncts they discuss based on data from Spanish and English. The asymmetries in extraction exhibited by the Spanish and English examples cannot be accounted for under the traditional approach to barriers, which crucially relies on the complement/non-complement distinction. In section 14.3, I provide new data from Japanese comparative deletion which support the conclusion reached by the works summarized in section 14.2. In section 14.4, I adopt MacDonald’s (2006, 2008a,b, 2009) proposal on inner aspect, and argue that Agree with Asp during the course of structure building is the key for the absence of ACEs in Japanese comparative deletion. Finally, section 14.5 concludes this chapter.

14.2 Adjunct Condition effects

14.2.1 On the inherent barrierhood of adjuncts

Under the pre-minimalist framework, Chomsky (1986), adopting the essence of Huang (1982), proposes that non-complements are barriers, in the sense that if a
phrase is a non-complement, it necessarily constitutes a barrier for extraction from within. Accordingly, since adjuncts are non-complements, they become barriers. Consequently, no movement is permitted out of adjuncts. In the early minimalist framework, this type of approach is also adopted in Chomsky and Lasnik (1993).

This proposal predicts that adjuncts always block extraction because they are inherent barriers. However, this prediction is not borne out: based on the analysis of extraction out of depictive secondary predicates, which have the status of an adjunct, Demonte (1988) and Borgonovo and Neeleman (2000) show that extraction out of an adjunct is possible in some well-defined contexts, which I summarize in the next section.

14.2.2 Surprising facts regarding Adjunct Condition effects

Demonte (1988: 22) shows that subject-oriented and object-oriented secondary predicates do not behave the same with regard to extraction. Consider (3a–c):1,2

\begin{enumerate}
\item a. *Con qué novio, no sabes [si Lola compró el coche [enfadada t]]
\item b. ?Con qué novia, no sabes [si María encontró a Pepe [enfadado t]]
\item c. De qué novia, no sabes [si Pepe volvió [harto t]]
\end{enumerate}

‘At which boyfriend, don’t you know [whether Lola bought the car [angry t]]?’

‘At which girlfriend, don’t you know [whether Maria found Pepe [angry t]]?’

‘From which girlfriend, don’t you know [whether Pepe came back [fed up t]]?’

\footnote{The abbreviations used in this chapter are as follows: \textit{ACC} = accusative, \textit{CL} = classifier, \textit{DAT} = dative, \textit{GEN} = genitive, \textit{NEG} = negation, \textit{NOM} = nominative, \textit{PL} = plural marker, \textit{V} = verb.}

In addition, all nominal phrases are labeled as NP since a distinction between NPs and DPs is not crucial for the purpose of this chapter.

\footnote{An anonymous reviewer raises the possibility that (3b), for example, may involve a small clause consisting of \textit{Pepe} as the subject and \textit{enfadado con qué novia} as the predicate. If this is correct, the extraction of \textit{con qué novia} would take place from a complement position, instead of from an adjunct, as I am assuming in the text. If so, the grammaticality of (3b) would not be surprising at all. Notice, however, that this small clause-based analysis does not easily extend to all the cases discussed by Demonte or Borgonovo and Neeleman in their works. Further, it is not likely that in (4b), for example, \textit{arrive} can be analyzed as taking a small clause complement consisting of \textit{John} and \textit{whistling what}.}
The ungrammaticality of (3a) is just as expected, but the grammaticality of (3b, c) is surprising. In (3b), the wh-phrase is extracted out of the object-oriented secondary predicate. In (3c), the wh-phrase is raised out of the secondary predicate modifying the subject of the unaccusative verb volvió “come.” Given the assumption that the subject of an unaccusative verb is base-generated in an object position, we can group (3b) and (3c) together, leading to the generalization that in Spanish extraction is possible out of an object-oriented depictive secondary predicate.

This surprising behavior of depictive secondary predicates is not peculiar to Spanish, but is attested cross-linguistically. Borgonovo and Neeleman (2000: 199–200) provide further evidence that extraction out of an adjunct is possible in certain well-defined contexts. Consider the contrast between (4a) and (4b) in English:

\[(4) \begin{align*}
a. & \quad \text{*What, did John dance [dressed as t₁]?,} \\
   b. & \quad \text{What, did John arrive [whistling t₁]?}
\end{align*}\]

If secondary predicates are adjuncts, this contrast between unergatives and unaccusatives illustrated in (4) cannot be accounted for in a principled way under an approach based on the inherent barrierhood of adjuncts.

To recapitulate, we have seen that extraction out of an adjunct is not uniformly excluded. The generalization seems to be that extraction out of a depictive secondary predicate is allowed when the adjunct in question is object-oriented. In section 14.3, I provide additional support for this generalization by examining new data from Japanese.

### 14.3 Comparative deletion and floating quantifiers in Japanese

This section discusses comparative deletion in Japanese. Comparative deletion allows us to reproduce in Japanese cases similar to the ones Demonte (1988) and Borgonovo and Neeleman (2000) discuss, and provide a good testing ground for extraction of an element out of an adjunct in this language.

First, in section 14.3.1, I give arguments in favor of the predicative nature of oozei. In section 14.3.2, I introduce the structure under analysis, Japanese comparative deletion, and show that the Japanese counterpart of the than-clause, the yori-clause, forms a unit with the QP oozei: a complex QP. In section 14.3.3, I show that this complex QP has the status of an adjunct. If this line of reasoning is correct, it implies that the examples analyzed in this section where the yori-clause is extracted out of the complex QP formed by the yori-clause and oozei, involve extraction out of an adjunct.

#### 14.3.1 The predicative nature of the FQ oozei in Japanese

I first lay out basic properties of the Japanese counterpart of the QP many which is used in Japanese comparative deletion. One of the forms to express many in Japanese is oozei, which is used exclusively for humans. Oozei can surface either as a FQ,
following the NP it modifies, or in a pre-nominal position, preceding this NP. This is illustrated in (5) and (6). In (5a) and (6a), the QP oozei surfaces as a FQ, while in (5b) and (6b), it is located in a pre-nominal position.

(5) a. gakusei-ga [QP oozei] oyoida(-koto)  
   student-NOM many swam (-fact)  
   'Many students swam.'

   b. [NP [QP oozei]-no gakusei]-ga oyoida(-koto)  
      many-GEN student -NOM swam (-fact)  
      'Many students swam.'

(6) a. Taroo-ga gakusei-o [QP oozei] hometa(-koto)  
      -NOM student-ACC many praised(-fact)  
      'Taroo praised many students.'

   b. Taroo-ga [NP [QP oozei]-no gakusei]-o hometa(-koto)  
      -NOM many-GEN student-ACC praised (-fact)  
      'Taroo praised many students.'

What type of element is oozei? I adopt the assumption that the QP oozei is a predicate. This assumption is supported by the fact that oozei shows the same distribution as typical predicates like adjectives have. In this regard, Japanese oozei differs from English many. Notice that in English, many cannot occupy a position after the copula, which is a typical predicative position in this language, leading to the contrast between (7a) and (7b):

(7) a. The students are quiet.

   b. *The students are many.

In contrast, oozei can appear in copulative constructions in Japanese, as shown in (8). Thus, there is no Japanese counterpart to the contrast in (7); such a contrast is absent in this language, as shown in (8a, b):

(8) a. gakusei-ga shizuka da.  
      student-NOM quiet be  
      'The students are quiet.'

   b. gakusei-ga oozei da.  
      student-NOM many be  
      'The students are many in number.'

The lack of any significant grammatical contrast between (8a) and (8b) therefore supports the hypothesis that the QP oozei is a predicate. I will assume that when oozei surfaces in a pre-nominal position, as in (5b) and (6b), it is adjoined to NP. If oozei is a predicate, as I have just argued, it is also natural to assume that this QP can also be a secondary predicate, in the same way as adjectives can also be secondary predicates in English, as illustrated in (9):
(9) John studied linguistics quiet.

Accordingly, I assume that in (5a) and (6a), the QP oozei is a secondary predicate: in (5a) this secondary predicate is subject-oriented and occupies the same position that subject-oriented predicates occupy; in (5b) it is an object-oriented secondary predicate, and is located in the position normally assigned to this type of element.3

As is well known, extraction of an adjunct out of an island always results in severe deviance. Consider (10a) for instance; it involves a complex NP, where the noun uwasa “rumor” takes a complement clause. In (10b), we have extracted the adjunct isshookenmei-ni “very diligently” out of this complex NP, and the sentence is severely ill formed. In contrast, extraction of the argument kuruma-o “the car” in the same context is not as strongly deviant, as shown by the contrast between (10b) and (10c):

3 The two anonymous reviewers question the status of the QP oozei as a secondary predicate in (5a) and (6a). One of the reviewers points out that s/he finds a grammatical contrast between (i) and (ii):

(i) *gakusei-ga sono hon-o san-nin yonda(-koto)
   student-NOM that book-ACC three-CL read (-fact)
   ‘Three students read that book.’
(ii) gakusei-ga sono hon-o oozei yonda(-koto)
   student-NOM that book-ACC many read (-fact)
   ‘Many students read that book.’

The ungrammaticality of (i) shows that no element can intervene between the FQ san-nin and its modifier gakusei “student.” In contrast, such intervention does not cause any problem in the example which involves oozei, as shown in (ii). If oozei is an adverb, the grammaticality of (ii) is naturally expected since we know that an adverb can intervene between the object and the verb, as illustrated in (iii):

(iii) gakusei-ga sono hon-o isoide yonda(-koto)
    student-NOM that book-ACC quickly read (-fact)
    ‘The students read that book quickly.’

However, some of my informants (including myself) find the example in (ii) above degraded. Further, there is evidence that oozei does not behave as an adverb in the examples under analysis. Consider the contrast between (iv) and (v):

(iv) *issyookenmei-ni (-no) gakusei
    very diligently (-GEN) student
    ‘the student working very hard’
(v) gakusei-ga issyookenmei-ni benkyoo-shita(-koto)
    student-NOM very hard study -did (-fact)
    ‘The students studied very hard.’

(iv) shows that unlike oozei, adverbials like issyookenmei-ni “very hard” cannot appear within a nominal. If oozei were an adverbial, it would not be immediately clear why this element can appear within a nominal, as shown in (5b) and (6b) in the text.

Yet, if the contrast between (i) and (ii) is real, and on the basis of the evidence I have discussed regarding the parallel behavior between the QP oozei and adjectival phrases in Japanese, I am led to conclude that the QP in point may have a dual status for some speakers/dialects, having both predicative and adverbial properties. For the purpose of this chapter, as far as the QP oozei occupies a position higher than VP when it is subject-oriented whereas it is located within VP when it is object-oriented (see also Nakanishi 2007), the main claim of this chapter is not affected. See Fukushima (1991), Gunji and Hashida (1998), among others for discussion on FQs being adverbs.
(10) a. \[TP_vP Kazuko-ga [\text{NP}_{CP[TP_vP Akira-ga isshookenmei-ni} \text{kuruma-o naoshita]}-toyuu] uwasao-kiita)](-koto)\]
\[\text{car -ACC fixed -that rumor-ACC heard (-fact)}\]
'Kazuko heard [the rumor that Akira fixed the car very diligently].'

b. \[TP isshookenmei-ni_1 [TP_vP Kazuko-ga [\text{NP}\text{CP[TP_vP Akira-ga t}_1 \text{kuruma-o naoshita]}-toyuu] uwasao-kiita)](-koto)\]
'\[Very diligently\_1, Kazuko heard [the rumor that Akira fixed the car t}_1\].'

c. \[TP kuruma-o_1 [TP_vP Kazuko-ga [\text{NP}\text{CP[TP_vP Akira-ga isshookenmei-ni t}_1 \text{naoshita]}-toyuu] uwasao-kiita)](-koto)\]
'The car\_1, Kazuko heard [the rumor that Akira fixed t}_1 very diligently].'

Given that the FQ oozei is an instance of a secondary predicate like the AP quiet in (9), parallel to (10b) we expect scrambling of the FQ oozei out of an island to be severely deviant. This prediction is borne out. First, consider (11a, b) in which no island is involved:\footnote{Miyagawa (1989: 64) judges (i) as ungrammatical:}

(11) a. \[\text{Satoshi-ga [CP[TP_vP Yamada-sensei-ga kanja-o oozei shinsatsu-shita]}-to] kiita(-koto)\]
\[\text{Dr. Yamada -NOM patient-ACC many examination-did-that heard(-fact)}\]
'Satoshi heard that Dr. Yamada saw many patients.'

b. \[\text{Satoshi-ga [CP[TP_vP Yamada-sensei-ga kanja-o t}_t \text{shinsatsu-shita]}-to] kiita(-koto)\]
\[\text{many -NOM Dr. Yamada -NOM patient-ACC examination-did-that heard(-fact)}\]
'(lit.) Many\_t, Satoshi heard that Dr. Yamada examined t}_t patients.'

\footnote{Miyagawa (1989: 64) judges (i) as ungrammatical:}

(i) \[\text{san-satsu}_1 \text{Taroo-ga [Hanako-ga hon-o t}_t \text{katta-to]} omotte-iru(-koto)\]
\[\text{three-CL -NOM book-ACC bought-that think (-fact)}\]
'Taroo thinks that Hanako bought three books.'

Parallel to the discussion regarding (11b) in the text, my informants (as well as myself) find (i) acceptable although it may be slightly degraded. I leave issues regarding the deviance of (i) for future research.
My informants (as well as myself) find (11b) acceptable, although it may be slightly degraded. This shows that the FQ oozei can undergo long-distance scrambling. In contrast, when oozei is scrambled out of an island, the sentence becomes ungrammatical, as shown by the contrast between (12a) and (12b) and the one between (13a) and (13b):

(12) Complex NP Island

a. Satoshi-ga [NP\[TP\[v\[P Yamada-sensei-ga kanja-o oozei [-NOM Dr. Yamada -NOM patient-ACC many shinsatsu-shita]-toyuu] uwasa]-o kiita(-koto) examination-did-that rumor-ACC heard(-fact)

‘Satoshi heard the rumor that Dr. Yamada examined many patients.’

b. *oozei Satoshi-ga [NP\[TP\[v\[P Yamada-sensei-ga kanja-o many [-NOM Dr. Yamada -NOM patient-ACC shinsatsu-shita]-toyuu] uwasa]-o kiita(-koto) examination-did-that rumor-ACC heard(-fact)

(lit.) Many, Satoshi heard the rumor that Dr. Yamada examined t1 patients.

(13) Adjunct Island

a. Tanaka-sensei-ga [CP\[TP\[v\[P Yamada-sensei-ga kanja-o oozei Dr. Tanaka -NOM Dr. Yamada -NOM patient-ACC many shinsatsu-shita]-node] yotei-o henkoo-shita(-koto) examination-did-because plan-ACC change-did (-fact)

‘Because Dr. Yamada examined many patients, Dr. Tanaka changed the plan.’

b. *oozei Tanaka-sensei-ga [CP\[TP\[v\[P Yamada-sensei-ga kanja-o many Dr. Tanaka -NOM Dr. Yamada -NOM patient-ACC t1 shinsatsu-shita]-node] yotei-o henkoo-shita(-koto) examination-did-because plan-ACC change-did (-fact)

(lit.) Many, Dr. Tanaka changed the plan because Dr. Yamada examined t1 patients.

The strong deviance observed in (12b) and (13b) therefore indicates that the FQ oozei holds an adjunct status, which is not surprising if oozei is an adjunct secondary predicate.

Once I have shown that oozei is a predicate and has the status of an adjunct, let us analyze which position this FQ occupies in the structure. Here I assume that subject-oriented FQs are merged at the vP level whereas object-oriented ones are merged at the VP level. Evidence for this comes from the Japanese counterpart of the do so replacement test (Nakau 1973). This test clearly shows that subject-oriented and object-oriented FQs occupy different positions (Koizumi 1994). Assuming that soo-suru “do so” corresponds to VP, we incorrectly predict that example (14c) can mean the same as (14b) if the subject-oriented FQs san-nin/oozei are located within VP.
(14) a. sensei-ga san-nin/oozei hon-o katta.  
   teacher-NOM three-CL/many book-ACC bought  
   ‘Three/Many teachers bought books.’  

   b. gakusei-mo san-nin/oozei hon-o katta.  
   student-also three-CL/many book-ACC bought  
   ‘Three/Many students bought books, too.’  

   c. gakusei-mo [VP soo-shita].  
   student-also so -did  
   ‘Students did so, too.’  

However, (14c) means that students also bought books, but not that three/many students also bought books. If the latter interpretation is intended, we need to express it as in (15):  

(15) gakusei-mo san-nin/oozei [VP soo-shita]  
   student-also three-CL/many so -did  
   ‘Three/Many students did so, too.’  

In contrast, when the FQs are object-oriented, (16c) can have the same meaning as (16b) when we apply the do so replacement test:  

   -NOM student-ACC three-CL/many praised  
   ‘Hanako praised three/many students.’  

   b. Taroo-mo gakusei-o san-nin/oozei hometa.  
   -also student-ACC three-CL/many praised  
   ‘Taroo praised three/many students, too.’  

   c. Taroo-mo [VP soo-shita].  
   -also so -did  
   ‘Taroo did so, too.’  

Actually, if the FQs san-nin/oozei are not elided in (16c), the examples become ungrammatical, as shown in (17):  

(17) *Taroo-mo san-nin/oozei [VP soo-shita].  
   -also three-CL/many so -did  

To summarize, the examples in (14)–(17) show that object-oriented FQs are necessarily within VP whereas subject-oriented ones are outside VP. Assuming a VP-shell approach to the structure of VP, we can interpret this result as an indication that object-oriented FQs are within VP while subject-oriented ones are in vP (or a higher projection). In this chapter, I assume that subject-oriented FQs are adjoined at the vP level, while object-oriented FQs are adjoined to VP, as shown in (18) (see also Nakanishi 2007):
With the structure in (18) in mind, let us now turn to Japanese comparative deletion.

Some illustrative examples are given in (19). In (19a, b) the FQ *oozei* modifies the object and the subject, respectively.

(19) a. \[ TP[vP Tanaka-sensei-ga [VP [NP [QP [CP Yamada-sensei-ga kanja-o [QP [CP Dr. Tanaka -NOM patient-ACC Dr. Yamada -NOM]](-kota)]](-koto)]](-kota) \]

Dr. Tanaka examined more patients than Dr. Yamada examined e.

(ii) \[ TP kinoo [VP [NP [QP [CP e Taroo-o [shinsatsu-shita-yori(-mo)] shinsatsu-shita]](-koto)]](-kota) \]

More doctors examined Hanako than e examined Taroo yesterday.

Since the QP *oozei* itself can appear in a pre-nominal position, it is not surprising that the QP with a *yori*-clause can also appear in the same position. Since (i) and (ii) are not the focus of the discussion, I leave these examples aside in the remainder of this chapter.

Note, in passing, that the fact that the *yori*-clause and *oozei* form part of the –no-modifier in (i) and (ii) is further evidence that they form a constituent, as defended in section 14.3.2.

7 See also Kikuchi (1989) for an analysis of Japanese comparative deletion within a pre-minimalist framework.
In these examples, the comparison is established between the two amounts/quantities expressed by the object NPs, in (19a), and the subject NPs, in (19b). In (19a), the QP oozei modifies kanja “patient,” and the comparison is made between the number of patients whom Dr. Tanaka examined and the number of those Dr. Yamada examined. In (19b), the comparison is established between the number of doctors who examined Taroo and the number of those who examined Hanako. Under this intended interpretation, (19b) results in slight marginality.8

Let us now consider what the structure of this type of construction is. Ishii (1991: 132) suggests that in examples like (19a, b) the Japanese counterpart of English than-clause, the yori-clause, is adjoined to the QP oozei, as illustrated in (20). As this structure shows, the yori-clause and the QP oozei form a complex QP:9

8 Before proceeding, we should note that the word-order sequence of (19b) also has another interpretation irrelevant for our purpose here, describing a situation in which many doctors examined Hanako rather than examined Taroo. Under this interpretation, there is no comparison of two numbers. I ignore such an irrelevant reading throughout this chapter.

9 One anonymous reviewer suggests that yori-clauses may be selected by an implicit comparative marker associated with the QP oozei, projecting a Deg(ree)P, in a way parallel to English than-clauses, which are selected by the morpheme -er or more. Even if this were correct, as far as I can see, it would not affect the main proposal of this chapter. What is important for us is the fact that regardless of whether it projects QP or DegP, the complex QP formed by oozei and the yori-clause, is not a barrier when it modifies a complement. See Lechner (2004a) and Hayashishita (2009) for recent proposals on comparatives in English, German, and Japanese.
Supporting evidence for this structure comes from the fact that no element can intervene between the CP yori-clause and the QP oozei (see also Ishii 1991). Miyagawa (1989) observes that temporal adverbials like kinoo “yesterday” can intervene between a subject and the FQ that modifies it, as shown in (21):

(21) \[ TP[vP \text{gakusei-ga} \text{kinoo san-nin soba-o} \text{tsukutta}](\text{-koto}) \text{made (-fact)} \]

‘Three students made Japanese noodles yesterday.’

As illustrated in (22), kinoo can also intervene between the subject and the QP oozei when this QP modifies the subject:

(22) \[ TP[vP \text{gakusei-ga} \text{kinoo oozei soba-o} \text{tsukutta}](\text{-koto}) \text{made (-fact)} \]

‘Many students made Japanese noodles yesterday.’

However, as shown in (23), although temporal adverbials like kinoo are relatively free to intervene between an NP and the QP that modifies this NP, this type of adverbial cannot intervene between the yori-clause and the QP oozei in the structures under analysis.

(23) \[ TP[vP \text{gakusei-ga} \text{oozei soba-o tsukutta-yori(-mo)} \text{oozei soba-o tsukutta}](\text{-koto}) \text{made (-fact)} \]

‘More students made Japanese noodles than e made sushi (yesterday).’

The unavailability of the temporal adverbial kinoo to intervene between the yori-clause and the QP oozei strongly suggests that these two elements are generated as a constituent. In what follows I will assume that the yori-clause and the QP oozei form a constituent, a complex QP. The structure I assume for Japanese comparative deletion is given in (24):

---

14.3 Comparative deletion and floating quantifiers in Japanese 341

---

10 See footnote 6 for further evidence that oozei and the yori-clause form a constituent.
Assuming that the structure of Japanese comparative deletion is as in (24), I now turn to examine whether movement (scrambling) of the yori-clause out of the complex QP formed by the yori-clause and oozei is possible.

14.3.3 Scrambling of the yori-clause out of the complex FQ headed by oozei

The main goal of this section is to show that scrambling of the yori-clause out of this complex QP (FQ) is not always prohibited, even if this QP is an adjunct.11 Consider the examples in (19a, b), repeated below as (25a, b):

(i) Satoshi-ga [CP Dr. Tanaka -NOM patient-ACC
kanja-o
[TVP
Tanaka-sensei-ga
\-NOM
Dr. Tanaka
\-NOM
shinsatsu-shita\-yori(-mo)] oozei
\[\text{Dr. Yamada -NOM examination-did-than(-also) many
\text{shinsatsu-shita\-to\ kiita\(-koto\)
\text{examination-did-that heard (-fact)}
\text{‘Satoshi heard [that Dr. Tanaka examined [more patients than Dr. Yamada examined c]].’}}
\]

The structure of (i) is roughly given in (ii):

(ii) [TP Satoshi-ga [CP Dr. Tanaka -NOM patient-ACC
kanja-o
[TVP
Tanaka-sensei-ga
\-NOM
Dr. Tanaka
\-NOM
shinsatsu-shita\-yori(-mo)] oozei
\[\text{Dr. Yamada -NOM examination-did-than(-also) many
\text{shinsatsu-shita\-to\ kiita\(-koto\)
\text{examination-did-that heard (-fact)}
\text{‘Satoshi heard [that Dr. Tanaka examined [more patients than Dr. Yamada examined c]].’}}
\]

The grammaticality of (i) is not surprising since no extraction takes place in this example: the FQ with the yori-clause stays in situ, as shown in (ii).

Let us now analyze what happens if the whole complex QP (the yori-clause and oozei) moves. As shown in (iii), this entire QP can scramble long-distance:
The structure of (vi) is roughly as in (vii):

(vii) [TP ... [NP[CP[TNP Satoshi-ga ] Tanaka-sensei-ga ] kanja-o ] Dr. Tanaka -NOM patient-ACC ]

\[\text{Dr. Tanaka - NOM} \quad \text{patient-ACC} \]

\[\text{shinsatsu-shita]-toyuu\]} \text{uwasa]-o} \text{kiita]}[-(koto)]\]

\[\text{Satoshi heard [the rumor [that Dr. Tanaka examined [more patients than Dr. Yamada examined e]]].}\]

The structure of (vii) is roughly as in (vi):

(vii) [TP ... [NP[CP[TNP Satoshi-ga ] Tanaka-sensei-ga ] kanja-o ] Dr. Tanaka -NOM patient-ACC ]

\[\text{Dr. Tanaka - NOM} \quad \text{patient-ACC} \]

\[\text{shinsatsu-shita]-toyuu\]} \text{uwasa]-o} \text{kiita]}[-(koto)]\]

\[\text{Satoshi heard [the rumor [that Dr. Tanaka examined [more patients than Dr. Yamada examined e]]].}\]

The fact that (viii) is severely degraded suggests that the object-oriented complex QP formed by the yori-clause and the FQ oozei has the status of an adjunct.

The raising of the complex QP can be illustrated as in (ix):

(ix) [TP [QP yori-clause [QP oozei]]1 [TP ... [NP[CP ... t1 V]] rumor] V]]

\[\text{The fact that (viii) is severely degraded suggests that the object-oriented complex QP formed by the yori-clause and the FQ oozei has the status of an adjunct.}\]
Dr. Tanaka examined more patients than Dr. Yamada examined.

More doctors examined Hanako than examined Taroo yesterday.

If the yori-clause is moved out of the complex QP formed by the yori-clause and oozei and scrambles to sentence-initial position (leaving oozei stranded), a clear asymmetry emerges between these two examples, as shown in (26a, b). 12

(26) a. (?)[TP[CP Yamada-sensei-ga e shinsatsu-shita-yori(-mo)]]

[TP[VP Tanaka-sensei-ga kanja-o [QP t1 oozei]]]

shinsatsu-shita]](-koto)

b. ?*[TP[CP[TP e Taroo-o shinsatsu-shita]-yori(-mo)]1 [TP kinoo

[VP isha-ga [QP t1 oozei][VPHanako-o shinsatsu-shita]]]](-koto)

(26a), where the yori-clause is extracted from the object-oriented secondary predicate formed by the yori-clause and oozei, remains grammatical whereas (26b), where the yori-clause is extracted from the subject-oriented secondary predicate formed by the yori-clause and oozei, is highly deviant. The ungrammaticality of the latter

12 One anonymous reviewer points out that the contrast between (25b) and (26b) is not clear. My informants, however, find (25b) clearly better than (26b). This difference in grammaticality judgment might be due to some dialectal variation. At this point, I have no account as to why (25b) is slightly degraded and leave this issue for future research.
example comes as no surprise since it would be considered a classical instance of ACEs. Yet, the grammaticality of the former example is not expected since, just as (26b), it involves extraction out of an adjunct (the secondary predicate); however, it shows no ACEs. Since both examples involve extraction out of an adjunct, I have to conclude that what makes extraction out of (26a) possible is that it involves extraction out of a secondary predicate that modifies an object. This is reminiscent of the asymmetries in extraction out of secondary predicates discussed by Demonte (1988) and Borgonovo and Neeleman (2000), introduced in section 14.2.2.

Given that, as I have shown in section 14.2.2 (see examples (3) and (4)), the nature of the predicates and arguments involved plays a role in accounting for the possibility of extracting out of a secondary predicate in English and Spanish, it would not be a surprise if we found that in contrast with examples like (26b), where the matrix verb is transitive and the yori-clause moves out of the subject-oriented secondary predicate (the complex QP), extraction out of a subject-oriented secondary predicate is possible when the matrix predicate is an unaccusative verb. As the grammaticality of (27b) shows, extraction of the yori-clause is allowed in this case: crucially, the matrix predicate, the verb toochaku-suru “arrive,” is unaccusative.

(27) a. [TP kesa [TP gakusei-ga [QP[CP e kinoo toochaku-shita-yori(-mo)] oozei] toochaku-shita](koto) arrival -did -than(-also) many arrival -did (-fact)
   "This morning more students arrived than e arrived yesterday."

b. (?)[TP[CP e kinoo toochaku-shita-yori(-mo)]1[TP kesa

   [TP gakusei-ga [QP t1 oozei] toochaku-shita]](koto)

In section 14.4, I will present additional data, based on the unergative/unaccusative alternation, which support the relevance that the matrix predicate and the aspectual properties of this predicate have in allowing extraction of the yori-clause out of the secondary predicate formed by the yori-clause and oozei.

To sum up, the fact that Spanish, English, and Japanese behave in a very similar way regarding the possibility of extracting an element out of an adjunct (the secondary predicate) indicates that the lack of ACEs is not a language-particular phenomenon. Consequently, we have to explain why such an exception is allowed in universal grammar.13

---

13 Extraction out of an adjunct is also permitted under an approach based on sideward movement (see, among others, Nunes and Uriagereka 2000, Hornstein and Nunes 2002, Nunes 2004, this volume, and references therein). Under the sideward-based approach, the extraction in point is allowed only when the
14.4 The proposal

This section starts with an introduction to MacDonald’s (2006, 2008a,b, 2009) Agree-based proposal on inner aspect, which will be the basis for my account of the extraction facts analyzed in section 14.3.

MacDonald (2008b) proposes two event features of predicates: \(<ie>\) and \(<fe>\) features. The presence of an \(<ie>\) feature, which is assumed to be on Asp, indicates that the event that the predicate describes has a beginning. If an \(<fe>\) feature is present, the event has an end. MacDonald’s classification of the different aspectual types of predicates can be summarized as follows with some relevant examples:

\[(28)\]

a. Accomplishment/Achievement \(<ie>\) and \(<fe>\)

b. Activity \(<ie>\) only

c. Stative No features

\[(29)\]

a. Activity

\[\begin{array}{c}
[vP Taroo [AspP Asp [VP carried the piano]]] \\
<ie> <ie>
\end{array}\]

b. Accomplishment

\[\begin{array}{c}
[vP Taroo [AspP Asp [VP drank a beer]]] \\
<ie> <ie> <fe> <fe>
\end{array}\]

c. Achievement

\[\begin{array}{c}
[vP Taroo [AspP Asp [VP caught the bug]]] \\
<ie> <ie> <ie> <ie>
\end{array}\]

d. Stative

\[\begin{array}{c}
[vP Taroo [VP knows two foreign languages]] \\
<fe> <ie>
\end{array}\]

In (29a–c), the \(<ie>\) feature projects to AspP, as indicated by the arrow. Only telic predicates, namely accomplishments and achievements, have an \(<fe>\) feature. However, in order to capture the punctual nature of the event described by achievement predicates, and the different interpretations these two types of predicates allow with modifiers like almost (see MacDonald (2008b) for detailed discussion on this point), MacDonald proposes that the \(<fe>\) feature is on V, projecting up to VP, for the element that is extracted (i.e. sideward-moved) from the adjunct clause merges in an argument position in the matrix clause. Notice, however, that the cases that the present chapter discusses are different in that the element extracted out of the adjunct arguably merges in a non-argument position. Consequently, it does not seem possible to extend a sideward movement analysis to the cases analyzed in this chapter.
accomplishment, whereas it is on Asp for the achievement. As for the stative, no AspP is projected, and thus, neither $<$ie$>$ nor $<$fe$>$ features are present.

In order to understand how these event features are interpreted, and also to clarify the effect that the internal argument has on the aspectual interpretation of the event described by the predicate, consider the contrast illustrated in (30):

(30) a. John drank a beer in ten minutes/*for ten minutes.
   b. John drank beer *in ten minutes/for ten minutes.

(MacDonald 2008b: 4)

According to MacDonald (2006, 2008a,b, 2009), in (30a), a beer refers to a specific quantity of beer, and thus, it is [+q] (where “q” stands for specified quantity of X in Verkuyl’s (1972) sense). In contrast, beer in (30b) is [−q] since it does not describe any specific quantity of the material under consideration. Significantly, this very difference makes the interpretation of the predicate telic in (30a) and atelic in (30b); only the PP in ten minutes is acceptable in (30a), while only for ten minutes can be chosen in (30b).

Under the clausal architecture given in (31), MacDonald (2006, 2008a,b, 2009) argues that the correlation between the [+−q] distinction of the NPs and the telic/atelic interpretation, exemplified in (30a, b), is a consequence of an Agree relation established between Asp and the “closest” NP: the closest NP values the unvalued [q] feature of Asp via Agree; as a result, if the NP is [+q], Asp gets specified as [+q], and if the NP is [−q], Asp gets specified as [−q].

(31) $\left[ TP \left[ vP \left[ AspP \left[ Asp [VP ... NP ... ] \right] \right] \right] \right]$

Whether Asp gets specified as [+q] or [−q] makes a difference for the domain of aspectual interpretation in the following way:

(32) a. When [−q] Agree relation is established;
   \[
   \left[ TP \left[ vP \left[ AspP \left[ Asp [VP ... ] \right] \right] \right] \right]
   \]
   b. When [+q] Agree relation is established;
   \[
   \left[ TP \left[ vP \left[ AspP \left[ Asp [VP ... ] \right] \right] \right] \right]
   \]

When the [−q] Agree relation is established, the domain of aspectual interpretation is AspP alone, excluding VP, as illustrated in (32a). When the [+q] Agree relation is made, the relevant domain is the entire domain shaded in (32b). This means that only

the event features located within these shaded domains will contribute to the aspectual interpretation of the predicate involved. Crucially, only with the [+q] Agree relation, event features within VP are considered.

This much said, let us return to (30a, b). Notice that in (30a), Asp is valued by the [+q] NP a beer, and therefore, both <ie> and <fe> features are within the domain of aspectual interpretation, which results in a telic interpretation of the predicate involved. This is illustrated in (33):

\[
(33) \quad [vP \text{ Taroo } [\text{AspP} \quad \text{Asp} \quad [vP \text{ drank a beer}]]
\]

In contrast, in (30b), Asp is valued by the [–q] NP beer; accordingly, the domain of aspectual interpretation is Asp alone, as shown in (34):

\[
(34) \quad [vP \text{ Taroo } [\text{AspP} \quad \text{Asp} \quad [vP \text{ drank beer}]])
\]

In this case, only the <ie> feature is within the domain of aspectual interpretation, and thus, the predicate will be atelic. In the end, the <fe> feature modifies the whole event: the entire event of drinking beer has an ending. This explains the contrast between (30a) and (30b).

How about the example in (35)?

(35) John carried the bag into water in ten minutes/*for ten minutes.  
(MacDonald 2009: 8)

The fact that only the PP in ten minutes is acceptable in (35) shows that the predicate is telic. The question is why water, the complement of the P into, being [–q], cannot turn the predicate into atelic in this example. If it were possible, then (35) would be incorrectly predicted to be acceptable with the PP for ten minutes. Notice that this type of change/coercion is sometimes possible, as illustrated in (36), where sand successfully turns the predicate into an atelic predicate.

(36) John carried sand into the room *in ten minutes/for ten minutes.  
(MacDonald 2009: 8)
The reason why water, being a [−q] NP, cannot affect the [+−q] status of the Asp in (35) is that it is not the closest NP for Asp, as illustrated in (37).\footnote{In order to show my point clearly, I do not illustrate V-movement in (37) and (38).}

\[ (37) \]

In (37), the NP the bag is the NP closest to Asp; it is therefore this NP that values the [uq] feature of this functional head. Since the bag is a [+q] NP, the [+q] Agree relation with Asp is established, and the <fe> feature, which is assumed to be on the P into, will be in the domain of aspectual interpretation. As a result, the event will be interpreted as telic. This explains why it is only the PP in ten minutes, and not the PP for ten minutes, that is acceptable in (35).

In (36), on the other hand, it is the [−q] NP sand that is the NP closest to Asp, as shown in (38): it is therefore this [−q] NP that values the [uq] of Asp, which gets specified as [−q]. This means that the domain of aspectual interpretation is only AspP where only the <ie> feature is present. This results in the atelic interpretation of the predicate.
To summarize the discussion so far, MacDonald proposes that NPs play a role in setting the [+/-q] value of Aspect via Agree: the closest NP in the c-command domain of Asp enters into an Agree relation with Asp and values the [uq] feature of Asp. This valuation results in the [+/-q] interpretation of the predicate. Importantly, this Agree-based approach correctly predicts the presence of intervention effects: only the closest NP can value Asp. The contrast between (35) and (36) with respect to the (un)availability of the durative phrase (*for ten minutes) thus follows directly under MacDonald’s proposal.

Another important consequence of this proposal is that only elements within the c-command domain of Asp can participate in Agree with Asp. Consider (39):

(39) Livestock pushed the cart into the barn in ten minutes/*for ten minutes.  

(MacDonald 2009: 10)

The external argument *livestock* is generated in vP SPEC, which means it is outside of the c-command domain of Asp. This explains why subjects (generated higher than Asp, in vP SPEC) cannot Agree with Asp, while complements, within the c-command domain of Asp, can do so.
With MacDonald’s Agree-based approach with the two event features in mind, let us now turn to the aspectual properties of predicates and to the role that FQs play in Japanese in this regard. Consider the grammaticality of (40a) and (40b):

(40) a. \[TP \text{Tanaka-sensei-ga } [VP \text{kanja } -o \text{oozei shinsatsu-shita}](\text{-koto}) \]
Dr. Tanaka -NOM patient-ACC many examination-did(-fact)
‘Dr. Tanaka examined many patients.’

b. \[TP \text{Tanaka-sensei-ga } [VP \text{kangofu-o oozei yooshita}](\text{-koto}) \]
Dr. Tanaka -NOM nurse -ACC many needed (-fact)
‘Dr. Tanaka needed many nurses.’

Notice that if the FQ \text{oozei} is not present, either the time span adverbial \text{go-fun-de} “in five minutes” or the durative adverbial \text{go-fun-kan} “for five minutes” can be added to (40a).

(41) a. \[TP \text{Tanaka-sensei-ga go-fun-de } [VP \text{kanja-o} \text{shinsatsu-shita}](\text{-koto}) \]
Dr. Tanaka -NOM five-minute-in patient-ACC examination-did(-fact)
‘Dr. Tanaka examined the patient(s) in five minutes.’

b. \[TP \text{Tanaka-sensei-ga go-fun-kan } [VP \text{kanja-o} \text{shinsatsu-shita}](\text{-koto}) \]
Dr. Tanaka -NOM five-minute-period patient-ACC examination-did(-fact)
‘Dr. Tanaka examined patients/a (or the) patient for five minutes.’

This is not surprising since Japanese nouns are not accompanied by any determiner, and in (41a, b), the bare N \text{kanja} can be understood as either \ [+q] “the patient(s)” or \ [-q] “patient(s)”. In contrast with the examples in (41), which involve the predicate \text{shinsatsu-suru} “examination-do”, only the durative adverbial \text{nijuu-nen-kan} “for twenty years” is permitted in (40b):

(42) a. *\[TP \text{Tanaka-sensei-ga nijuu-nen-de } [VP \text{kangofu-o} \text{yooshita}](\text{-koto}) \]
Dr. Tanaka -NOM twenty-year-in nurse -ACC needed (-fact)
‘Dr. Tanaka needed (the) nurses/a (or the) nurse in twenty years.’

b. \[TP \text{Tanaka-sensei-ga nijuu-nen-kan } [VP \text{kangofu-o} \text{yooshita}](\text{-koto}) \]
Dr. Tanaka -NOM twenty-year-period nurse -ACC needed (-fact)
‘Dr. Tanaka needed (the) nurses for twenty years.’
This is again natural because the predicate in (40b) is the stative predicate yoosuru “need” (Kindaichi 1950).

Now, as a first step towards our analysis of the FQ oozei, let us discuss many in English. It has been argued (e.g. Partee 1989) that one of the readings many allows is a cardinal reading. On this reading, many is a “vague” cardinal quantifier, requiring the value of the number of the objects involved to be large in the context considered. It is then natural to assume that many + N is [+q]. If this is correct, the contrast in (43) also naturally follows:16

(43)  a. *Taroo made paper crafts in two hours.
    b. Taroo made many paper crafts in two hours.

The verb make is an accomplishment verb. This means that this verb has an <fe> feature. In order for the time span adverbial to be acceptable in (43a), this <fe> feature must be within the domain of aspectual interpretation. This requires the internal argument to be [+q] so that the VP will be part of the aspectual domain, as desired. However, in (43a), bare plurals like paper crafts are [-q]. Thus, (43a) is unacceptable. The grammaticality of (43b) then indicates that the internal argument, many paper crafts, is [+q]. The only difference between (43a) and (43b) is that in the latter example, many is present. Therefore, I conclude that the source of the internal argument being [+q] lies in the presence of many. This amounts to saying that many is [+q].

Extending MacDonald’s approach to inner aspect to Japanese, I propose that the FQ oozei, the Japanese counterpart of many, is also lexically specified as [+q]. I further propose that the modifiee of the FQ comes to inherit the [+q] feature of the FQ, because of the predication relation established between the two elements (the FQ and the modified NP). Under this view, the Agree relation that the FQ establishes with Asp is mediated by its modifiee.17, 18

16 The same contrast is observed with mass nouns. (i) and (ii) are cited from Borer (2005: 121)

(i) *Robin sifted sand in half an hour.
(ii) Robin sifted (too) much sand in half an hour.

The fact that the time span adverbial is acceptable only in (ii) shows that not only many but also much is [+q] and that the [+/-q] distinction is independent of the count/mass distinction of nouns (see, for instance, Borer (2005) and MacDonald (2008a,b)).


18 According to Mihara (2004), NPs, in principle, can be either [+q] or [-q]. For example, consider (i) and (ii):

(i) Masao-ga Barriers-o futsu-ka-de yonda.
    -NOM -ACC two-day-in read
    ‘Masao read Barriers in two days.’

    -NOM -ACC two-day-period read
    ‘Masao read Barriers for two days.’

The NP Barriers is [+q] in (i) since Masao finished reading the book, but it is [-q] since he read part of the book.
With this in mind, let us now examine (40a) and (40b). (44) shows the structure of (40a):\textsuperscript{19}

\[(44)\]

\[
\begin{array}{c}
\text{vP} \\
\text{Subject} \\
\text{AspP <ie> v} \\
\text{VP<fe> Asp <ie> [+q]} \\
\text{Object} \\
\text{Object-oriented FQ V<fe> [+q]} \\
\text{shinsatsu-shita}
\end{array}
\]

Here, the [+q] object NP, which has inherited the [+q] feature of the FQ, can agree with and value Asp. Accordingly, the domain of aspectual interpretation is AspP and VP in (44). Within this aspectual domain, we have <ie> and <fe> features, resulting in a telic interpretation of the predicate. Under this interpretation, (40a) means that Dr. Tanaka examined a patient one after another, and in the end, he examined many patients. The fact that only the Japanese counterpart of the time span adverbial in fifteen minutes is acceptable, as shown in the contrast between (45a) and (45b), is consistent with the present analysis, and provides support for the view that the FQ plays a crucial role in the aspectual interpretation of the predicate in (40a):

\[(45)\]  
a. \([\text{TP} \text{ Tanaka-sensei-ga juugo-fun-de [VP kanja-o oozei Dr. Tanaka -NOM fifteen-minute-in patient-ACC many shinsatsu-shita]](-koto)\]

‘Dr. Tanaka examined many patients in fifteen minutes.’

\textsuperscript{19} In this chapter, I consider shinsatsu-suru “examination-do” as a lexical verb in order to avoid unnecessary complication arising from illustrating incorporation operations to form the complex V in question from shinsatsu and suru. See Saito and Hoshi (2000) for relevant discussion on the incorporation operations in point. I also do not illustrate V-movement in (44), (46), and (64a, b), in order to avoid unnecessary complications and show my point clearly.
b. *\([TP \text{ Dr. Tanaka -NOM}}\) \text{juugo-fun-kan} \quad [VP \text{kanja-o oozei shinsatsu-shita}](\text{-koto})\]

‘Dr. Tanaka examined many patients for fifteen minutes.’

How about (40b)? The structure of this example is given in (46):

\[
\begin{aligned}
\nuP & \quad \nu' \\
\text{Subject} & \quad \nu \\
\text{VP} & \quad \nu' \\
\text{Object} & \quad \\
\text{Object-oriented FQ} & \quad \nu \\
\text{[+q]} & \quad \text{[+q]} \\
\end{aligned}
\]

\(\text{yooshita}\)

(40b) contains the stative V \text{yoosuru} “need”; following MacDonald’s proposal, no AspP is present in this structure. Therefore, no Agree relation with Asp can be established. The [+q] specification of the NP thus does not influence the aspectual interpretation of the predicate, and the predicate continues to be interpreted as atelic. The following contrast with durative and time span adverbials is again consonant with our analysis:

(47) a. *\([TP \text{ Dr. Tanaka -NOM}}\) \text{nijuu-nen-de} \quad [VP \text{kangofu-o oozei yooshita}](\text{-koto})\]

‘Dr. Tanaka needed many nurses in twenty years.’

b. \([TP \text{ Dr. Tanaka -NOM}}\) \text{nijuu-nen-kan} \quad [VP \text{kangofu-o oozei yooshita}](\text{-koto})\]

‘Dr. Tanaka needed many nurses for twenty years.’

In sum, under MacDonald’s framework, both (40a) and (40b) are correctly expected to be grammatical, and importantly, only the former involves Agree with Asp.
Consider now what happens when we have subject-oriented FQs; this is illustrated in (48). The subject NPs, as well as the subject-oriented FQs, are structurally too high (outside of the c-command domain of Asp). Following MacDonald’s analysis, we then expect the [+q] feature of the subject-oriented FQs in (48a, b) to be unable to enter into an Agree relation with the aspectual head. Consequently, the value of Asp is not affected by the [+q] of the FQ oozei in these examples, and both of the examples are grammatical.

(48) a. \([TP \text{ isha-ga} \text{ oozei} [VP \text{ Taroo-o} \text{ shinsatsu-shita}]](-koto)\]
\[\text{doctor-NOM many -ACC examination-did(-fact)}\]
‘Many doctors examined Taroo.’

b. \([TP \text{ isha-ga} \text{ oozei} [VP \text{ kangofu-o yooshita}]](-koto)\]
\[\text{doctor-NOM many nurse -ACC needed (-fact)}\]
‘Many doctors needed nurses.’

Since (48b) involves the stative predicate yoosuru “need”, and this verb is not subject to aspectual variation, the choice between the time span adverbial and the durative adverbial does not provide us with any clue to see the effect of the FQ modifying the subject NP. On the other hand, as I have shown above in the discussion of (41) and (45), in the case of the verb shinsatsu-suru “examine,” it is possible to check whether the subject-oriented FQ oozei affects the aspectual interpretation of the predicate in (48a). If the FQ in point plays a role in deciding the aspectual interpretation, we would expect to find a grammatical contrast between examples involving the time span adverbial and ones containing the durative adverbial, in a way parallel to the contrast observed in (45). Consider (49a) and (49b):

(49) a. \([TP \text{ isha-ga} \text{ oozei} \text{ go-fun-de} [VP \text{ Taroo-o} \text{ shinsatsu-shita}]](-koto)\]
\[\text{doctor-NOM many five-minute-in -ACC examination-did(-fact)}\]
‘Many doctors examined Taroo in five minutes.’

b. \([TP \text{ isha-ga} \text{ oozei} \text{ go-fun-kan} [VP \text{ Taroo-o} \text{ shinsatsu-shita}]](-koto)\]
\[\text{doctor-NOM many five-minute-period -ACC examination-did(-fact)}\]
‘Many doctors examined Taroo for five minutes.’

But see Tanaka (2008) for a different view on subject-oriented FQs, and for the proposal that a different functional category located above vP holds a feature-checking relationship with subject-oriented FQs. However, considering MacDonald’s (2006, 2008a,b, 2009) arguments on inner aspect, I assume that aspectual feature-checking is limited to object-oriented FQs.
Unlike the cases in (45a, b), no contrast emerges between the two types of adverbials in (49). This means that the subject-oriented FQ oozei does not contribute to the aspectual interpretation of the predicate, which in turn indicates that no Agree relation between the subject-oriented FQ and Asp is established. Consequently, given the assumption that AspP is absent with statives, I conclude that there should be an underlying difference between (40a) on the one hand, and (40b), (48a), and (48b) on the other, with respect to the availability of their FQ oozei to Agree with Asp.

Based on this hypothesis, I next present my proposal to account for the extraction asymmetries that we observed in Japanese comparative deletion in section 14.3. I will first examine those cases involving comparative deletion in the context where the complex FQ is object-oriented, and I will then turn to the cases of comparative deletion where the complex FQ is subject-oriented.

In (50) we have two examples involving comparative deletion where the complex FQ modifies the matrix object. (50a) involves the predicate shinsatsu-suru “examine” and (50b) the predicate yoosuru “need.” The examples in (50) are the comparative deletion counterparts of those in (40). Note first of all that, parallel to the examples in (40a, b), there is no grammatical contrast between the example with shinsatsu-suru “examine” and the one with yoosuru “need” when they show up with the object-oriented oozei.

(50) a. [TP[vP Tanaka-sensei-ga [vP kanja-o [QP[CP Yamada-sensei-ga shinsatsu-shita-yori(-mo) oozei] shinsatsu-shita]](-koto)]
ed many examination-did(-fact)
‘Dr. Tanaka examined more patients than Dr. Yamada examined e.’

b. ?[TP[vP Tanaka-sensei-ga [vP kangofu-o [QP[CP Yamada-sensei-ga yooshita-yori(-mo) oozei] yooshita]](-koto)]
needed-many needed (-fact)
‘Dr. Tanaka needed more nurses than Dr. Yamada needed e.’

The parallelism between (50a) and (50b) is not surprising: we have only added the than-clause to the FQ oozei to the examples in (40a) and (40b).

Consider now the cases involving comparative deletion when the complex FQ modifies the subject. There is also no difference in the grammaticality status of the examples in (51), both of which involve the complex FQ formed by the yori-clause and the subject-oriented FQ oozei. The examples are both well formed: the verb shinsatsu-suru “examine” and the verb yoosuru “need” can both co-occur with this FQ.

(51) a. ?[TP kinoo [vP isha-ga [QP[CP e Taroo-o shinsatsu-shita-yori(-mo) oozei] Hanako-o examination-did-than(-also) many shinsatsu-shita]](-koto) examination-did(-fact)
‘More doctors examined Hanako than e examined Taroo yesterday.’
Again, this is exactly what I predict. In these examples, the subject and the complex QP formed by the yori-clause and the FQ oozei are located too high to enter into an Agree relation with Asp, so we expect these elements not to affect the aspectual interpretation of the predicate.

However, crucially, as we observed in section 14.3, if the yori-clause is scrambled, an interesting contrast emerges. Consider (50a, b) first, which involve an object-oriented FQ; as shown in (52), a surprising grammatical asymmetry emerges between these two examples when the yori-clause is scrambled.

(52) a. (?)[TP[Cp Yamada-sensei-ga e shinsatsu-shita-yori(-mo)]
   Dr. Yamada -NOM examination-did-than(-also)
   [TP[vP Tanaka-sensei-ga [VP kanja-o [QP t oozei]
   Dr. Tanaka -NOM patient-ACC many
   shinsatsu-shita]])(-koto)
   examination-did (-fact)
   ‘Dr. Tanaka examined more patients than Dr. Yamada examined e.’

b. ?*[TP[Cp Yamada-sensei-ga e yooshita-yori(-mo)]
   Dr. Yamada -NOM needed-than (-also)
   [TP[vP Tanaka-sensei-ga [VP kangofu-o [QP t oozei]
   Dr. Tanaka -NOM nurse -ACC many
   yooshita]])(-koto)
   needed (-fact)
   ‘Dr. Tanaka needed more nurses than Dr. Yamada needed e.’

In contrast, if we scramble the yori-clause out of the subject-oriented FQ (formed by the yori-clause and oozei) in (51a, b), both examples become severely degraded, as shown in (53):

(53) a. ?*[TP[Cp e Taroo-o shinsatsu-shita]-yori(-mo), [TP kinoo
   -ACC examination-did-than(-also) yesterday
   [vP isha-ga [QP t oozei] [VP Hanako-o shinsatsu-shita]])(-koto)
   doctor-NOM many -ACC examination-did (-fact)
   ‘More doctors examined Hanako than e examined Taroo yesterday.’
b. ?*[TP[CP e kaminshitsu-o yooshita-yori(-mo)]i [TP odoroita-koto-ni nap room -ACC needed-than(-also) surprisingly
[ VP isha-ga [QP t oozei] [VP kitsuenshitsu-o yooshita]](-koto)
doctor-NOM many smoking room-ACC needed (-fact)
‘Surprisingly, more doctors needed a smoking room than e needed a nap room.’

What does the asymmetry between (52a) on the one hand, and (52b), (53a), and (53b) on the other, follow from?

In section 14.3.1 and in fn. 11, I have shown that the complex FQ formed by oozei and the yori-clause is an adjunct, whether it is a subject-oriented or an object-oriented secondary predicate. It thus follows that the asymmetries in the possibility of extracting the yori-clause cannot be explained as resulting from the +/-complement status of the domain of extraction. Therefore, we need to look for an alternative solution which does not rely on the +/-complement asymmetry in order to give a principled explanation of all these contrasts.

I propose that it is the Agree relation between the [+q] object NP (which inherits this [+q] feature from the complex QP) and Asp that exempts the complex QP from its status as a barrier in the relevant examples. Consider the structure in (54), which illustrates the structure I propose for subject-oriented and object-oriented complex FQs in the contexts of comparative deletion under analysis. If this structure is correct, the complex FQ formed by the yori-clause and oozei can agree with Asp via the NP it modifies when this element is an object, but not when it is a subject, as the latter is located too high in the structure.

(54)

```
Subject [+q]
  vP
    v'
      v'
        v
          AspP
          v
            QP
              Asp
                v
                  QP
                    oozei
                      [+q]
                        Object [+q]
                          V'
                            QP
                              v
                                CP
                                  QP
                                    yori-clause
                                      CP
                                        QP
                                          yori-clause
                                            CP
                                              QP
                                                yori-clause
                                                  CP
                                                    QP
                                                      yori-clause
                                                        CP
                                                          QP
                                                            yori-clause
                                                              CP
                                                                QP
                                                                  yori-clause
                                                                    CP
                                                                      QP
                                                                        yori-clause
                                                                          CP
                                                                            QP
                                                                              yori-clause
                                                                                CP
                                                                                  QP
                                                                                    yori-clause
                                                                                      CP
                                                                                       QP
                                                                                        yori-clause
                                                                                             CP
                                                                                               QP
                                                                                                yori-clause
                                                                                                    CP
                                                                                                      QP
                                                                                                          yori-clause
                                                                                                               CP
                                                                                                                QP
                                                                                                                    yori-clause
                                                                                                                        CP
                                                                                                                            QP
                                                                                                                                yori-clause
                                                                                                                                 CP
                                                                                                                                 QP
                                                                                                                                 yori-clause
                                                                                                                                 QP
                                                                                                                                 yori-clause
                                                                                                                                 CP
                                                                                                                                 QP
                                                                                                                                 yori-clause
                                                                                                                                 QP
                                                                                                                                 yori-clause
                                                                                                                                 CP
                                                                                                                                 QP
                                                                                                                                 yori-clause
                                                                                                                                 QP
                                                                                                                                 yori-clause
                                                                                                                                 CP
                                                                                                                                 QP
                                                                                                                                 yori-clause
                                                                                                                                 QP
                                                                                                                                 yori-clause
```
Since, as we have seen in the discussion of the paradigms above, it is only when an element can enter into an Agree relation with Asp that extraction from within this element is allowed, I propose that what explains the contrast between (52a) on the one hand, and (52b), (53a), and (53b) on the other, is that Agree exempts the complex FQ formed by the yori-clause and oozei from being a barrier. Consequently, this object-oriented complex FQ is a transparent domain and does not constitute a barrier for extraction of the yori-clause from within: this explains the grammaticality of (52a). In (52b) however, no AspP is present due to the fact that yoosuru “need” is a stative predicate, and thus, no Agree relation between the object-oriented FQ and Asp can be established. Consequently, no extraction can take place from within the complex QP formed by the yori-clause and oozei.

In contrast with object-oriented FQs, the subject-oriented FQ in (53) cannot enter into an Agree relation with Asp via the subject NP, and thus, it counts as a barrier for extraction of the yori-clause from within. Under this analysis, the contrast between (52a) on the one hand, and (52a), (53a), and (53b) on the other, derives from whether or not the FQ can Agree with Asp via the NP that it modifies.

Support for the proposed analysis comes from Borgonovo and Neeleman’s (2000:200) observation that there is a grammatical contrast between (55b) and (55d):

(55) a. The athlete ran [singing Carmen].
   b. *What1 did the athlete run [singing t1]?
   c. The athlete ran home [singing Carmen].
   d. What1 did the athlete run home [singing t1]?

Hoekstra (1984) and Levin and Rapoport (1988), among others, show that verbs of motion like run behave as unergatives in isolation, but behave as ergatives with a directional adverb. Accordingly, the contrast between (55b) and (55d) can be understood as another instance of the contrast between unergatives and unaccusatives discussed in section 14.2.2.

With the contrast between (55b) and (55d) in mind, consider (56a, b). (56a) contains two directional PPs whereas (56b) involves two non-directional comitative PPs, which indicate with whom the students swam.

(56) a. [TP kinoo [vP gakusei-ga [Qp[Cp e kishi-made oyoida-yori(-mo)] yesterday -NOM shore-to swam -than(-also) oozei] [vP booto-made oyoida]](-koto)
   many boat -to swam (-fact)
   ‘More students swam to the boat than e swam to the shore yesterday.’

b. [TP kinoo [vP gakusei-ga [Qp[Cp e kodomo-to oyoida-yori(-mo)] yesterday -NOM child -with swam -than(-also) oozei] [vP tomodachi-to oyoida]](-koto)
   many friend -with swam (-fact)
   ‘More students swam with friends than e swam with children yesterday’
What is interesting is that if we scramble the yori-clause in these examples, a clear grammatical contrast emerges, as shown in (57a, b). This contrast is parallel to the contrast between (55b) and (55d):

(57)  
a. \( ?[\text{TP}\left[\text{CP} e \text{kishi-made oyoida-yori(-mo)}\right]\left[\text{TP} \text{kinoo}\right] \left[\text{vP gakusei-ga} \left[\text{QP t1 oozei}\right]\left[\text{vP booto-made oyoida}]\right](-koto)\right] \left[\text{shore-to swam-than(-also) yesterday}\right] \) 

\[\text{student-NOM many boat -to swam(-fact)}\]

b. \( ?^*\left[\text{TP}\left[\text{CP} e \text{kodomo-to oyoida-yori(-mo)}\right]\left[\text{TP} \text{kinoo}\right] \left[\text{vP gakusei-ga} \left[\text{QP t1 oozei}\right]\left[\text{vP tomodachi-to oyoida}]\right](-koto)\right] \left[\text{child -with swam-than(-also) yesterday}\right] \) 

\[\text{student-NOM many friend -with swam(-fact)}\]

Under the Agree-based approach, we can account for the grammaticality of (57a) and (27b), repeated here as (58b), in a uniform manner. Recall that (58b) is derived from (58a) via the scrambling of the yori-clause.

(58)  
a. \( [\text{TP} \text{kesa} \left[\text{TP} \text{gakusei-ga} \left[\text{QP}\left[\text{CP} e \text{kinoo}\right]\left[\text{vP toochaku-shita-yori(-mo)}\right] \left[\text{oozei}\right]\right] \left[\text{oozei}\right]\right] \left[\text{toochaku-shita}]\right)(-koto) \left[\text{this morning student-NOM yesterday arrival -did -than(-also) many arrival -did (-fact)}\right] \) 

“This morning more students arrived than e arrived yesterday.’’

b. \( ?[\text{TP}\left[\text{CP} e \text{kinoo toochaku-shita-yori(-mo)}\right]\left[\text{TP} \text{kesa} \left[\text{TP} \text{gakusei-ga} \left[\text{QP t1 oozei}\right]\left[\text{toochaku-shita}]\right](-koto) \right] \) 

\[\text{[QP t1 oozei] toochaku-shita}]\right)(-koto) \)
In (56a), where the verb oyoida "swam" is accompanied by the directional PP kishi-made "to the shore" or booto-made "to the boat," the unergative verb behaves as an unaccusative predicate. Following standard assumptions, I propose that in this case the subject is generated as a complement of the verb. From this point of view, (56a) is parallel to (58a). This means that the complex QP/FQ formed by the yori-clause and oozei is also generated within VP in (56a) and (58a). The consequence of this is that this complex FQ, via the NP it modifies, can agree with Asp. This explains why the extraction of the yori-clause is permitted in (57a) and (58b).

Before closing this chapter, I would like to provide further support to the hypothesis that it is Agree with Asp that exempts the adjuncts under analysis from being a barrier. In order to do so, I will now examine another form that Japanese has to express "many:" takusan. Unlike oozei, takusan can be used both for humans and non-humans; this is illustrated in (59a, b):

(59) a. [TP[vP gakusei-ga takusan [VP kesseki-shita]](-koto)
   student-NOM many absence-did (-fact)
   ‘Many students were absent.’

b. [TP[vP Akira-ga [VP biiru-o takusan hiyashita]](-koto)
   -NOM beer-ACC much cooled (-fact)
   ‘Akira cooled lots of beer.’

Given this much background, let us consider the example in (60):

(60) [TP Hanako-ga ni-ji-kan/ni-ji-kan-de [VP biiru-o
   -NOM two-hour-period/two-hour-period-in beer-ACC
   [QP[CP Akira-ga e hiyashita -yori(-mo) takusan] hiyashita]](-koto)
   -NOM cooled -than(-also) much cooled (-fact)
   ‘Hanako cooled more beer than Akira cooled e for two hours/in two hours.’

It is important to notice that the time span adverbial and the durative adverbial are both acceptable in (60). This fact in turn indicates that both telic and atelic readings are available in this example. This is reminiscent of the fact that English V cool is also ambiguous in the same way:

(61) John cooled the soup for ten minutes/in ten minutes.

(MacDonald 2009: 85)

It is then natural to assume that hiyasu “cool” is also ambiguous between the activity and the accomplishment use in (59b) and (60).

---

In this chapter, I will not address the question of what the precise position of the subject is in (56a). See, for example, Ramchand (2008) who argues that a phrase referring to the entity that holds the result state occupies R(esult)P SPEC. This RP SPEC position is claimed to be within the c-command domain of Asp. Thus, under Ramchand’s framework, the subject NP can still establish an Agree-relationship with Asp in (56a).
As may be expected, in (60), the *yori*-clause can be extracted out of the complex QP formed by the *yori*-clause and *takusan*, and move to sentence-initial position, as shown in (62):

\[
\text{(62) (?) TP} \left[ \text{CP Akira-ga e hiyashita-yori(-mo)}, \text{ TP Hanako-ga} \right. \\
\quad \text{NOM cooled -than(-also) -NOM} \\
\quad \text{ni-ji-kan/ni-ji-kan-de [VP biiru-o [QP t takusan] beer-ACC much]} \\
\quad \text{hiyashita]}(-koto) \\
\text{cooled (-fact)} \\
\text{‘Hanako cooled more beer than Akira cooled e for two hours/in two hours.’}
\]

The grammaticality of (62) with either the durative adverbial *ni-ji-kan* ‘for two hours’ or the time span adverbial *ni-ji-kan-de* ‘in two hours’ shows that the scrambling of the *yori*-clause out of the complex QP formed by the *yori*-clause and *takusan* is available in either a telic or an atelic context. This fact in turn indicates that the availability of scrambling of the *yori*-clause out of the complex QP is independent of the telic/atelic distinction of the predicate involved. Notice that this is exactly what we predict under the present proposal incorporating MacDonald’s analysis of inner aspect, which I now turn to.

Let us go over the derivations of (59b), which is crucial in understanding why the *yori*-clause can scramble to sentence-initial position, as shown in (62) above. Here I assume that not only the FQ *oazei* but also the FQ *takusan* is lexically specified as [+q], and it establishes an Agree relation with Asp via the NP it modifies. Evidence for the assumption that the FQ *takusan* is [+q] comes from, again, the (un)availability of durative and time span adverbials in a sentence with the FQ in point. Consider the contrast between (63a) and (63b):

\[
\text{(63) a. TP} \left[ \text{vP Taroo-ga juugo-fun-de/juugo-fun-kan} \right. \\
\quad \text{NOM fifteen-minute-in/fifteen-minute-period} \\
\quad \text{[VP tegami-o kaita]}(-koto) \\
\quad \text{letter -ACC wrote (-fact)} \\
\quad \text{‘Taroo wrote letters in fifteen minutes/for fifteen minutes.’}
\]

\[
\text{b. TP} \left[ \text{vP Taroo-ga juugo-fun-de/*juugo-fun-kan} \right. \\
\quad \text{NOM fifteen-minute-in/fifteen-minute-period} \\
\quad \text{[VP tegami-o takusan kaita]}(-koto) \\
\quad \text{letter -ACC many wrote (-fact)} \\
\quad \text{‘Taroo wrote many letters in fifteen minutes/*for fifteen minutes.’}
\]

In (63b), the durative adverbial cannot appear. Again, the only difference between (63a) and (63b) is the presence of *takusan* in the latter example. Considering the fact that only the time span adverbial *juugo-fun-de* “in fifteen minutes” is acceptable in (63b), I therefore conclude that *takusan* is also [+q].
I also assume that the Agree relation with Asp works in the same way as the cases involving the FQ *oozei* discussed above. (64a, b) show the derivations with the accomplishment and the activity use of the V *hiyasu* 'cool' respectively.

(64) a. Accomplishment Use

```
<table>
<thead>
<tr>
<th>vP</th>
</tr>
</thead>
<tbody>
<tr>
<td>v'</td>
</tr>
<tr>
<td>( Asp &lt;i_e&gt; )</td>
</tr>
<tr>
<td>v</td>
</tr>
<tr>
<td>( VP &lt;f_e&gt; Asp &lt;i_e&gt; )</td>
</tr>
<tr>
<td>( Object )</td>
</tr>
<tr>
<td>( V' )</td>
</tr>
<tr>
<td>(+q)</td>
</tr>
<tr>
<td>( Object-oriented FQ )</td>
</tr>
<tr>
<td>( V &lt;f_e&gt; )</td>
</tr>
<tr>
<td>(+q)</td>
</tr>
<tr>
<td>hiyashita</td>
</tr>
</tbody>
</table>
```

b. Activity Use

```
<table>
<thead>
<tr>
<th>vP</th>
</tr>
</thead>
<tbody>
<tr>
<td>v'</td>
</tr>
<tr>
<td>( Asp &lt;i_e&gt; )</td>
</tr>
<tr>
<td>v</td>
</tr>
<tr>
<td>( VP Asp &lt;i_e&gt; )</td>
</tr>
<tr>
<td>( Object )</td>
</tr>
<tr>
<td>( V' )</td>
</tr>
<tr>
<td>(+q)</td>
</tr>
<tr>
<td>( Object-oriented FQ )</td>
</tr>
<tr>
<td>( V &lt;f_e&gt; )</td>
</tr>
<tr>
<td>(+q)</td>
</tr>
<tr>
<td>hiyashita</td>
</tr>
</tbody>
</table>
```
The only difference between (64a) and (64b) is that the $<fe>$ feature is present in the former but not in the latter example. This very difference results in the telic and atelic interpretation in (59b), thus in (60) as well. What is important here is the fact that in both (64a) and (64b) an Agree relation of the NP with *takusan* with Asp has been established. Under the present proposal, the grammaticality of (62) follows from the fact that the [+q] object NP agreeing with Asp has exempted the adjunct QP from being a barrier, independent of whether the predicate is interpreted as telic or atelic.

To summarize, I have shown that the complex FQ formed by the *yori*-clause and *takusan* can agree with Asp via the NP it modifies. We have also seen that when this feature relation takes place, the adjunct QP is exempted from its barrierhood and allows extraction of the *yori*-clause from within. If this is correct, Agree with Asp, not the telicity of the predicate involved per se, plays an important role in defining whether an “adjunct” will be a barrier and when it will allow extraction of an element from within.

Crucially, this is independent of the complement/non-complement status of adjuncts. In all the cases I have analyzed in this chapter involving comparative deletion in Japanese, the complex QP formed by the *yori*-clause and *oozei/takusan* had the status of an adjunct, whether it was subject-oriented or object-oriented and whether it entered into an Agree relation with Asp or not. However, some of these adjuncts (the object-oriented FQs) can enter into an Agree relation with Asp via the NP they modify. It is in this case that these adjuncts allow extraction from within. Furthermore, I have also shown that extraction is not always granted from object-oriented complex FQs: when the FQ (via its associated NP) cannot establish an Agree relation with Asp (for instance, when Asp is not present in the structure, as was the case in (52b)), extraction becomes impossible. The asymmetries in the possibility of extracting out of a subject in examples with predicates subject to the unergative/unaccusative alternation, provide further support to the analysis I have advocated for, as well as to the syntax-based approach to inner Asp I have defended in this chapter.

The relevant difference between the cases analyzed in this chapter and the vast majority of cases that have been analyzed in the literature to support the inherent islandhood of adjuncts is that, in the latter case, the constituent/adjunct that behaved as a barrier and blocked extraction out of its domain did not undergo any Agree relation with Asp (or with another relevant head in the main clause). This is why these adjuncts always blocked extraction and why they have always been taken to be inherent barriers for movement. I have concluded that the classical analysis of adjuncts as inherent barriers is not correct, based on a detailed analysis of Japanese comparative deletion.
14.5 Concluding remarks

In this chapter I have shown that there is an asymmetry, not discussed in the literature until now, in the possibility of scrambling a *yori*-clause out of a complex QP in Japanese comparative deletion, in configurations where the complex QP (formed by the *yori*-clause and the QP *oozei/takusan*) is a secondary predicate and modifies an NP in the matrix clause. Although, as I have shown, this complex QP has the status of an adjunct, the *yori*-clause can be raised out of this complex QP/FQ when it is object-oriented, but not when it is subject-oriented. This is in line with the results previously reported by Demonte (1988) and Borgonovo and Neeleman (2000).

I have proposed that what makes some object-oriented secondary predicates transparent for extraction is the fact that object-oriented FQs can enter into an Agree relationship with Asp via the NP they modify during the course of the derivation; in contrast, subjects in *vP SPEc* and the complex QPs that modify these elements are too high in the structure and cannot establish any Agree relation with Asp. I have provided further evidence for this approach on the basis of the unergative/unaccusative alternation exhibited by some predicates and the way in which the aspectual properties of the predicate and the complex QP interact in allowing or disallowing extraction.

If this analysis is correct, I have to conclude that adjuncts are not inherent barriers, or at least, that they can be exempted from this status in certain well-defined contexts. The conclusion is that structure building operations play a crucial role in accounting for the +/-islandhood of adjuncts: we need to examine where adjuncts are merged, and determine whether any (relevant) Agree relationship with the matrix clause is established along the course of structure building (Boeckx 2003, 2008, MacDonald 2006, 2008a,b, 2009).
Feature-Splitting Internal Merge: The Case of Tough-constructions*

MIKI OBATA AND SAMUEL DAVID EPSTEIN

15.1 Introduction

The goal of this chapter is to shed new light on some properties of Internal Merge application, Merge being a, or the, fundamental operation of structure building. Chomsky’s (2007, 2008) feature-inheritance analysis stipulates that two heads such as T and C simultaneously attract a single element.1 (For a possible explanation, see Epstein, Kitahara, and Seely, this volume.) Obata and Epstein (2008, 2011) extend Chomsky’s idea of simultaneous attraction and propose a new kind of structure building operation called Feature-Splitting Internal Merge, which “splits” (by featural decomposition; see Chomsky 1964, 1995b) a single element into two independently “moveable” syntactic objects. This new kind of Internal Merge yields structure building enabling us to give a local, featural, non-universal account of improper movement phenomena, which has thus far been considered a universally prohibited phenomenon requiring long-distance postulates (cf. May’s (1979) Condition C approach and Fukui’s (1993) chain uniformity condition account). Following Obata and Epstein (2008, 2011) we assume here that, under our analysis of improper movement, it is correctly predicted that certain kinds of improper movement are in fact allowed. This will play a crucial role in our analysis of the “perennially vexing” English tough-construction.

In this chapter, we further apply our analysis of Feature-Splitting Internal Merge to other types of structure building by mainly considering the following three questions:

---

* We would like to thank the audience at the Ways of Structure Building conference (November, 2008 at University of the Basque Country) for their helpful discussion. Also, we are very grateful to Betty McNulty Epstein, Kleanthes Grohmann, Terje Londahl, Jairo Nunes, Yukio Otsu, Milan Rezac, Shigeo Tonoike, and especially to Noam Chomsky, Hisatsugu Kitahara, and T. Daniel Seely for their extremely important comments and suggestions. We are also very grateful to the editors of this volume, Myriam Uribe-Etxebarria and Vidal Valmala, for their hospitality at the conference and for their interest in our research.

1 Aspects of this chapter are also discussed in Obata (2010).
(1) What kind of mechanism allows us to derive tough-constructions but rule out (other) improper movement? (2) What feature(s) distinguish a finite T which assigns nominative Case, from a control T, which assigns null Case to PRO, given that both types of T have identical [uPhi] under Chomsky (2001, 2007, 2008)? (3) Why is there a null Case assigning T but asymmetrically and implicitly there is assumed to be no null Case assigning V? With respect to these issues, we assume, as is standard, that there are at least three kinds of T/C (“finite,” infinitive-Control and “Raising”) but we also propose that, correspondingly, there exist three kinds of v/V analogously distinguished in terms of their feature content. That is, we eliminate the (standardly assumed) stipulated implicit asymmetries between T/C (3 kinds) on the one hand and v/V (2 kinds) on the other—and explore the consequences of the proposed symmetry. This three-way distinction predicts that there are also three ways of feature-splitting under Internal Merge depending on what features appear on probes, goals, and how they are split when Internal Merge applies. The result is a more natural system, (eliminating stipulated asymmetries) and one which provides us with new derivational possibilities of structure building. (“Simpler theory, increased generative capacity”). Given the proposed system, tough-constructions are, in fact, derived by “proper improper movement” moving a lexical DP, not a null operator (contra. Chomsky 1982). (See Epstein 1989, Brody 1993, Hornstein 2001 for similar analyses, and Hicks 2009 for a proposal that both a lexical DP and a null operator undergo movement in the derivation of tough-constructions.)

In addition, our system implies that Case on a tough-subject is “revised” in the course of the derivation, a phenomenon which has been reported in e.g. Bejar and Massam (1999) for “untough” phenomena in other languages e.g. Hungarian, but which they hypothesized to be non-existent in English.

The chapter is organized as follows: section 15.2 provides a brief review of feature-splitting and makes the issues discussed in the chapter explicit. Section 15.3 advances a hypothesis regarding feature distribution on heads and demonstrates how it works. Section 15.4 addresses new ways of structure building under feature-splitting and gives a derivational account of tough-constructions. Section 15.5 concludes our study.

15.2 Framing the issues

15.2.1 A new way of structure building: Feature-Splitting Internal Merge

In this section, we consider Chomsky’s (2007, 2008) feature inheritance system and its predictive content. Since On Phases, the treatment of T (and also V) has radically

2 Although our approach and Hornstein’s (2001) approach both suggest that a lexical DP, not a null operator, moves in tough-constructions, we still assume PRO in contrast to the movement approach to control originating in O’Neil (1994) and developed in Hornstein (1999, 2001).
changed: T does not bear phi-features inherently, but rather inherits phi-features from C. Consequently, T cannot operate as a probe until C is introduced into the derivation. This feature inheritance analysis from C to T makes it possible to non-lexically distinguish the phi-feature constitution of a finite/control T from the one selected by raising predicates. T occurring with raising predicates is not selected by C, so T does not inherit phi-features. Lacking phi-features, T appearing with raising predicates does not have the ability to value Case. That is, the availability of C determines the potential of T. Therefore, we no longer need to stipulate that the lexicon contains two different Ts. Rather, the features of the sole lexical T are determined “functionally” by the (selectional) syntactic context in which the T appears. Crucially, under this system, movement to Spec-CP and to Spec-TP takes place simultaneously and independently. (See Chomsky 2007 and 2008 for more details.) This analysis, whereby certain movements to Spec-CP do not proceed through Spec-TP (More generally, A-movement does not feed A’-movement), is empirically motivated to account for asymmetries in the suppression of subject condition effects as presented in Chomsky (2008). (See Obata and Epstein 2008, 2011 for more details regarding simultaneous application of syntactic operations.)

What happens when C and T both attract a single element simultaneously? In such cases, a single element is “separated” into two syntactic landing sites because of Internal Merge triggered by the edge features of two distinct heads. Exactly how is this executed? In Obata and Epstein (2008, 2011), we proposed that features on a single element are “split” into two landing sites, via Feature-Splitting Internal Merge. For instance, consider the following case:

(1)  Who left?
    Step1: [vP who[Φ[iCase][iQ]] V [vP left]]
    Step2: [TP T [vP who[Φ[iCase][iQ]] ...]]
    Step3: [CP C [TP T [vP who[Φ[iCase][iQ]] ...]]]
    Step4: [CP C [TP T[Φ] [vP who[Φ[iCase][iQ]] ...]]]
    C-to-T Feature Inheritance
    Step5: [CP C [TP T[Φ] [vP who[Φ[iCase][iQ]] ...]]]

Simultaneous attraction of “who” by both T and C

In (1), the subject “who” is merged into the edge of vP as in Step 1. After VP is transferred, T is merged as in Step 2 and then C is merged as in Step 3. [uΦ] on C is inherited by T in Step 4. In Step 5, after T phi-agrees with [uΦ] on “who”, T and C
simultaneously attract the single element “who”. (2) illustrates feature distribution (obeying the No Tampering Condition) after simultaneous attraction in Step 5.3

\[
(2) \quad \text{[CP who[iQ] C [TP <who[iPhi][uCase]> T [vP <who[iPhi][Cause][iQ]>...}
\]

In Obata and Epstein (2008, 2011), we suggest that T attracts all and only the features affected by T-DP agreement, namely [phi] and [Case] while C attracts the rest of the features. Arguably, this can be induced from Chomsky’s (2007) A/A’-distinction: “A-movement is IM (Internal Merge) contingent on probe by uninterpretable inflectional features, while A’-movement is IM driven by EF” (Chomsky 2007: 24).4 Since Chomsky (2000), it has been assumed that [uCase] on DP is valued as a reflex of phi-agreement. That is, in (2), both [uCase] and [iPhi] on “who” occurring in the edge of vP are “involved” in or affected by, phi-agreement with T, while [iQ] on “who” is not. Given this analysis, we propose that “who” at the edge of CP bears only [iQ] but loses [Phi] and [Case] as illustrated in (2).5

If the sentence in (1)/(2) is embedded, the prediction is that “who” at the edge of CP, which lacks phi/case features can never undergo phi-agreement, i.e. it can never be the goal of a higher phi-probe later in the derivation. This prediction is upheld as exemplified by the ungrammaticality of (3):

(3) *Who seems (that) will eat the cake?

(4) Who seems to eat the cake?

3 If the No Tampering Condition says Merge X, Y leaves X, Y unchanged. Then, our feature-splitting seems to satisfy No Tampering Condition in the following way:

(i) [iPhi] movement leaves an identical copy of [iPhi] in the departure site.
(ii) [iQ] movement leaves an identical copy of [iQ] in the departure site.
(iii) The departure site has all and only the features it had prior to application of Internal Merge.

A reviewer points out that Chomsky (2008: 144) argues that C-to-T feature inheritance violates the No Tampering Condition (because phi-features are not left in C) yet nonetheless satisfies the SMT.

4 Note that in fact both movements are equivalently driven by EF, not just A’-movement. Hence, “EF-driven” does not distinguish A-movement from A’-movement. Rather, A-movement is preceded by phi-agreement by the probe-goal system while A’-movement is not. But it is not clear how these pre-movement relations are encoded in output relations that must be present in interface representations where A/A’ interpretation differs. For a proposed solution, see Obata and Epstein (2008, 2011).

5 For us, the same mechanism applies also to v and V. (See Epstein, Kitahara, and Seely 2011.) That is, features on an object wh-phrase are split into two positions, namely the edges of vP and VP. Since [iPhi] and [uCase] are both involved in phi-agreement by V, those features are attracted by EF on V. On the other hand, [iQ] is attracted by v. However, Grohmann (2003) proposes anti-locality conditions, which ban movement from a complement position to a specifier position within a single XP, so that movement of the object from Comp-VP to Spec-VP should be blocked. This issue deserves further consideration, but here we provisionally adopt Chomsky’s feature-inheritance system and assume it applies to v-V as it does to C-T. (See also Richards 2007.)
(3) has been regarded as a case which can be derived by improper movement. (cf. Chomsky 1973) In (3), “who” moves from the embedded subject position to embedded Spec-CP then back to an A-position (i.e. the matrix Spec-T). Under our analysis, [iPhi] on “who” has been split off, appearing on the occurrence of “who” at the embedded Spec-TP, but absent on the occurrence of “who” in the embedded Spec-CP, which bears only [iQ]. As a result [uPhi] on matrix finite T remains unvalued because of lack of [iPhi] on “who” in Spec-CP, “who” being the only visible DP goal, given Transfer of the embedded TP. That is, lower occurrences of “who,” e.g. the occurrence in the embedded Spec-TP, which does have phi-features, are unavailable as a goal for the probing matrix T, given the Phase Impenetrability Condition. Thus, as a result of feature-split, the improper movement derivation crashes as desired—and does so on the basis of unvalued features on a single category—without appeal to unbounded relations crucial to c-command based Binding or Chain-based (uniformity) accounts of improper movement. On the other hand, (4) also involves movement of “who” out of the embedded clause but there is no finite T in the embedded TP which phi-agrees with “who.” Therefore, feature-split does not happen in the embedded clause. This is why [uCase] on “who” can be valued by the matrix T, which in turn gets its phi-features valued by “who,” and the derivation of (4) converges.

In a nutshell, Obata and Epstein (2008, 2011) proposed Feature-Splitting Internal Merge, where features on a single element are split into two syntactic positions in simultaneous attraction of a single element by two distinct heads. How features split depends on whether features are involved in Agree with T (or V). Features such as [iPhi] and [uCase] on DP are attracted by T because they are involved in Agree while [iQ] on DP, a feature irrelevant to Agree with T, is attracted by C. As a consequence, improper movement phenomena are ruled out by causing local featural crash as explained in this section.

15.2.2 Three research questions

This chapter extends the feature-split system by considering the following three questions. The first issue concerns tough-constructions:

(5) John is easy to please.

---

6 Obata and Epstein (2008, 2011) following Carstens (2008 and p.c.) discuss data from Kilega and Lusaamia, two Bantu languages, in which Case-valuation takes place independently of phi-agreement. That is, [iPhi] on DP does not take part in the Agree relation with T in these languages. Given our analysis, the prediction is that [iPhi] can be attracted by C in the configuration of (2) and improper movement should be allowed. In fact, many of the Bantu languages including Kilega and Lusaamia allow improper movement as predicted. Notably, other researchers have been compelled to propose convergent derivations (for recalcitrant data) which appear to us to involve improper movement. (See Polinsky and Potsdam 2006, Drummond, Kush, and Hornstein 2008.)

7 Regarding phonological features and pronunciation of the top copy, see Obata (2009).
One of the reasons why tough-constructions have always attracted attention is that one path “John” might in principle take is to move from the object position in the embedded clause to the matrix Spec-TP via the embedded Spec-CP. But of course, this is improper movement, as in “*John seems that Bill likes.” This is why a lot of attempts have been made to explain this construction without appeal to improper movement which has been widely assumed to be universally excluded and has been thus far regarded as a phenomenon completely unrelated/insensitive to variant feature constitutions of different types of T (cf. Lasnik and Fiengo 1974, Chomsky 1982, Browning 1987 among others). As summarized in the last section, given our system, improper movement is ruled out causing featural crash by unvalued [uPhi] on the matrix T. That is, another way to possibly account for tough-constructions is to hypothesize that in those constructions, [uPhi] on the matrix T can somehow successfully phi-probe the tough-subject (“John” in (5)) when it has moved improperly into the embedded Spec-CP and still bears its inherent [iPhi]. This kind of convergent improper movement derivation must somehow be blocked in cases such as (2)/(3). We will discuss this issue in section 15.4.

The second issue concerns not the matrix T but the embedded T, specifically the differences between finite T and control T. Since Chomsky (2001), it has been assumed that both control T and finite T consist of [uPhi], which triggers phi-agreement with a subject and values Case on the subject as a consequence. That is, finite T and control T are not distinguishable in terms of their phi-feature content, i.e. both have a full set of [uPhi]. However, a well-known question confronting this analysis is why finite T cannot license PRO:

(6) *(John thinks that) PRO eats apples.

This data implies that control T and finite T behave differently with respect to subject-licensing. But how do they differ? One possibility to be pursued in section 15.3 is that finite T bears [uPhi] while control T bears [uCase]. (See e.g. Pesetsky and Torrego 2001.) Assuming different Case-valuation mechanisms between these two Ts provides us with a new way of feature-splitting, which is dependent on features on probes.

The third issue is why there is a null Case assigning T, but there is implicitly assumed to be no null Case assigning V. As addressed in the last paragraph, control T and finite T should be distinguished in terms of their feature content. The same logic arguably applies to V. After V inherits features from v, it phi-agrees with an object DP. Then, [uCase] on DP is valued as a consequence of phi-agreement and is realized as accusative Case, while finite T values [uCase] on a subject DP and nominative Case is realized. But what about control T? Control T can license PRO valuing null Case but not a lexical DP, unlike finite T. But why doesn’t a corresponding V, licensing a PRO object, exist? Is there any logical reason which makes V licensing PRO impossible—or is
this just an implicit stipulation? Is there empirical evidence in favor of null (PRO) objects? Arguably, “Yes,” even in English as in e.g. (7) which appears to take object PRO:

(7)  
a. John washes PRO.
   (cf. John washes himself.)

b. John is eager to please PRO
   (cf. John is easy to please, Chomsky 1965)

(7a) is interpreted as “John washes himself” without having any visible object. That is, it is possible that the object position is filled by PRO controlled by the subject “John” (cf. Rizzi 1986 for Italian, Epstein 1984 for English, and Chomsky 1981 for his analysis of tough-constructions assuming an object PRO). In (7b), there “must be” an argument in object position to saturate the internal theta-role assigned by “please,” so that there exists by hypothesis a discourse-determined antecedent for the null object, again, possibly PRO. If this speculation regarding null objects is on track, then there should be two Vs, just like there are two Ts: one licensing a lexical object and the other licensing PRO (leaving aside Raising T, a third kind irrelevant to this particular discussion). This assumption—that there exists finite V and also a control V—renders T and V symmetrical in terms of their feature content and syntactic behavior, thereby eliminating an unexplained, if even recognized, lexical asymmetry in the feature content of Phase-head complements.

In the next section, the second and third issues regarding T and V are discussed in detail by advancing a more explicit hypothesis regarding their (parallel) feature content.

15.3 Phase-head complement feature symmetry

15.3.1 Three types of T/C and (correspondingly) three types of V/v

In this section, we further consider two of the three issues raised in the last section regarding properties of T and V. As summarized earlier, it has not been clearly stated how to distinguish control T and finite T and also why there is no V licensing PRO. In order to overcome these unanswered problems, we advance the following hypothesis:

(8) Hypothesis:
    There are uniformly 3 kinds of C/T and 3 kinds of v/V distinguished in terms of their feature content.

8 A comprehensive examination of the voluminous and important literature concerning cross-linguistic variation concerning null objects lies beyond the scope of this chapter.
(9)\[
\begin{array}{|c|c|}
\hline
C/T & v/V \\
\hline
a. C \rightarrow T_{[u\Phi]} & d. v \rightarrow V_{[u\Phi]} \\
\text{Nominative} & \text{Accusative} \\
b. C \rightarrow T_{[u\text{Case}]} & e. v \rightarrow V_{[u\text{Case}]} \\
\text{Null Case} & \text{Null Case} \\
c. T[---] & f. V[---] \\
\text{No Case} & \text{No Case} \\
\hline
\end{array}
\]

In (9), finite $T$ and control $T$ can be distinguished by their feature make-up. That is, nominative Case is assigned as a consequence of phi-agreement triggered by $[u\Phi]$ on $T$. Meanwhile, null Case is assigned by Case-agreement triggered by $[u\text{Case}]$ on $T$, so that there is no phi-agreement with a control $T$. Also, under this system, null Case assignment is not an ad hoc property of $T_{[u\text{Case}]}$ but there similarly exists a null Case assigning $V$, namely $V_{[u\text{Case}]}$. Furthermore, suppose these three kinds of $T$ and $V$ are freely combined, i.e. there is no selectional constraint on the structure building combination of these.

(10) T$[u\Phi]$ a.c. V$[u\Phi]$
   a. V$[u\text{Case}]$
   b. V[---]

(11) T$[u\text{Case}]$ a.c. V$[u\Phi]$
   d. V$[u\text{Case}]$
   e. V[---]

(12) T[---] a.f. V$[u\Phi]$
   g. V$[u\text{Case}]$
   h. V[---]

In the following sections, we clarify what types of sentences are derived by each of these nine combinations.

15.3.2 $[u\text{Case}]$: on null Case assignment

Let us begin with considering how $[u\text{Case}]$ valuation takes place. In our analysis, $[u\text{Case}]$ plays an important role in distinguishing between $T/V$ licensing a lexical DP and $T/V$ licensing PRO. The first issue in need of clarification is how an unvalued feature ($[u\text{Case}]$ on $T$) values another unvalued feature ($[u\text{Case}]$ on DP). The basic properties of the feature valuation system are: [1] an unvalued feature on a probe triggers Agree with a goal bearing its valued counterpart, and then [2] the unvalued probe feature receives a value from the goal. In the case of $[u\text{Case}]$, however, this scenario is
not straightforwardly applicable in that a goal DP also has [uCase], not the valued counterpart. That is, neither the goal nor the probe has a value. (See also Epstein and Seely 2006 for discussion of Agree between unvalued features.) With respect to this problem, we propose that it is in this context that “null/zero” Case is valued.

(13) The mechanism of null Case assignment:

Null Case is, we propose, to be understood as phonetically “unspecified” Case in this context. In other words, the computational system recognizes that [uCase] is valued, here as [null Case]. That is, [uCase] valued by T/V[Case] as in (13) is regarded as null Case. Since the assigned value is “null”, however, phonological features on PRO/DP in (13) are unspecified. More concretely, a feature bundle comprising, e.g. [pronoun], [3rd person], [singular], [masculine], and [Nom], is phonetically specified as “he” when valued by T[uPhi]. On the other hand, the same feature bundle is NOT phonetically specified if valued by T/V[uCase], though it is valued. That is, valuation (as null Case) has occurred, but phonetic encoding of DP/PRO (at this point) is lacking. This is a desired result for PRO because PRO is phonetically empty to begin with, hence does not require/receive phonetic encoding. However, what happens if a lexical, i.e. phonetically overt, DP receives null Case? It has phonological features which must undergo interpretation by the Sensory-Motor system (SMS). Consider the following standard paradigm:

(14) a. Mary hopes to go.
    b. *Mary hopes Bill to go.

T[uCase] assigns null Case to PRO in (14a) and to “Bill” in (14b). Why is (14b) ungrammatical? Following Epstein (1990), we make the empirical claim that what is wrong with (14b) is phonetic, not theta-theoretic or semantic, as arguably evidenced by the coherent semantic interpretation we (our CI-systems) can impose on (14b). Let us see the PF and LF representations for (14b):

(15) PF/LF representation for (14b)
    PF: Mary hopes to go.
    LF: Mary hopes Bill to go.

Crucially, there is no Case crash in the PF representation, i.e. the derivation converges at PF. That is, Case is in fact valued, specifically it is valued as [null Case], which is a legitimate feature at the PF interface. Since “Bill” has [iPhi], the representation at the LF
side contains “Bill”. On the other hand, phonological features on “Bill” are unspecified by null Case, so that it is unpronounceable with null Case by SMS reading the PF representation, i.e. only specified phonological features are “pronounceable.” Consequently, the inherent lexical phonological features on “Bill” never undergo their required PF interpretation. This violates the Recoverability Condition, which requires that no lexical, interface-interpretable information be “deleted” or rendered uninterpretable by narrow syntax operations (Chomsky and Lasnik 1993). In this case, phonological information on “Bill” is lost. This is why (14b) is ruled out. Again, it will be crucial to our analysis of tough-constructions that they do NOT crash, i.e. the Case feature of “Bill” *is* valued (as null Case).

With tough-constructions in mind, notice that there is another possible outcome for such phonetically unlicensed but Case-valued lexical items to, in principle, become specified by other (higher) probes later on in the derivation. In other words, the valued null Case (which does NOT cause crash) might be later “revised” by another higher Case-valuer, so that it is represented at the PF side observing the Recoverability Condition, informally it *is* pronounceable. We claim that this is exactly what happens in the derivation of tough-constructions. Putting aside the feature-splitting system for the time being, consider:

(16) He is easy to please.

We propose that the DP is first merged into the direct object position of “please.” In this example, suppose we have selected the null Case valuing form of “please” (not the accusative valuer which inherits phi-features from v). Then, in the embedded clause, Case on the object is valued as null Case. As a result, Transfer of the embedded VP does not crash—precisely because Case *is* valued, in this case, as null.

(17) In the embedded clause:

\[
\begin{array}{c}
\text{please}_{[\text{uCase}]} \quad \text{he/him}_{[\text{iPhi}]} \\
\text{Case-Agree}
\end{array}
\]

But, as noted above, null Case is insufficient to license SMS pronunciation of a lexical DP, i.e. a DP bearing inherent phonological features. Suppose (somehow) the matrix T (“is”) also agrees with the DP, changing its Case from null to Nom.

(18) In the matrix clause:

\[
\begin{array}{c}
\text{is}_{[\text{uPhi}]} \quad \text{he/him}_{[\text{iPhi}]} \\
\text{Phi-Agree}
\end{array}
\]

he/him becomes HE.
In (16), the *tough*-subject “he” is interpreted as the object of “please” but pronounced with nominative morphology at the matrix subject position. That is, “he/him” is first merged into the object position of “please.” If V (“please”) has [uPhi], it assigns accusative Case to the object, which is realized as “him” as in “It is easy to please him.” If V has [uCase] as illustrated in (17), Case-valuation takes place between V and the [uCase] object and null Case is assigned. Again, [uCase] on the object is valued (there is no crash after Transfer) but the phonological features are still unspecified. Then, the object undergoes movement to the matrix clause. Since this is a lexical DP, it has [iPhi]. The matrix T[uPhi] can phi-agree with the moving element as shown in (18). As a consequence of this phi-agreement, phonological features are finally specified as “he,” which is represented at PF. That is, no recoverability violation occurs. This is why (16) is grammatical. (We will further consider the derivation for *tough*-constructions in section 15.4 highlighting how they differ from other improper movement phenomena under the feature-split analysis.)

Thus, under the analysis proposed here, Case on a *tough*-subject is “revised” (or perhaps stacked) in the course of the derivation, which makes it possible to specify phonological features. But is such Case revision an ad hoc mechanism lacking any independent motivation, beyond our analysis of *tough*-constructions? Interestingly, Case-revision is also observed in other languages. For example, Bejar and Massam (1999) present data from languages such as Hungarian, where Case on DP can be “revised” after the DP undergoes A’-movement. They call this phenomenon “multiple Case-checking.”

(19) kiket montdad hogy szeretnél ha eljönnének
   who-ACC you-said that you-would-like if came(3PL)
   ‘Who did you say that you would like it if they came?’
   (Bejar and Massam 1999: 66, cited from Kiss 1985)

“Kiket” is extracted out of the subject position of the finite if-clause, where it receives nominative Case. On the way to the final landing site, it subsequently gets accusative Case from the intermediate verb (“said”) and the second Case—accusative Case—is always the one that is phonetically realized. That is, Case is multiply assigned to “kiket,” although, as Bejar and Massam point out, this is (interestingly in our view) a construction-specific phenomenon (like *tough*-constructions in English).9

9 A reviewer pointed out that there is something different between our Case-revision and Bejar and Massam’s multiple Case-checking: In the former case, the first Case is always null Case while in the latter, the first Case is not null but phonetically realizable.
Summarizing the discussion so far, [uCase] on a V-probe triggers Case-agreement with a goal and null Case is assigned. Since null Case is a valued Case, its appearance when transferred to PF does NOT induce crash. But null Case is phonetically unspecified Case; phonological features are not represented at the PF representation. Therefore, if the assignee is a lexical DP, not PRO, recoverability requires it to be phonetically specified by agreeing with (hence being valued by) another probe at some subsequent point of the derivation.

15.3 Nine possibilities
The previous section explained how [uCase] behaves in the derivation. Now, let us examine the nine possible combinations one by one.

(20) T[\uPhi] a. V[\uPhi] → e.g. He likes her.\(^\text{10}\)
b. V[uCase] → e.g. He washes.
c. V[---] → e.g. He was arrested.

The first set includes three combinations each with T[\uPhi]. T[\uPhi] is a finite T, so that it can assign nominative Case whose phonological features are specified as “he” in (20). V[\uPhi] values accusative Case on the object, so that it is realized as “her” as in (20a). V[uCase] values null Case as summarized in the last section. If the object is PRO, which is controlled by the subject “he” as in (20b), the sentence is fine. But if a sentence like “He washes his car” is derived under this combination, the derivation is ruled out because the object “his car,” which gets null Case, never undergoes Case-revision thereby causing a recoverability violation. In (20c), V which transmits phi/Case-features to V is missing. This is why V does not have any unvalued feature triggering Agree. This corresponds to e.g. the passive case, so that the object needs to get nominative Case from T[uPhi].

Next, consider the second set of T/V combinations which contains three different types of V with T[uCase].

(21) T[uCase] a. V[uPhi] → e.g. It is easy to please him.
        Mary wants to please him.
b. V[uCase] → e.g. He is easy to please.
c. V[---] → e.g. He wants to be arrested.

\(^{10}\) In Obata and Epstein (2008, 2011) summarized in section 15.2.1, we assumed that ungrammatical improper movement cases fall into this combination, e.g. the embedded T and V in “*John seems that Mary likes.” The verb “seem” selects either a finite CP or an infinitival TP but not an infinitival CP as in “*It seems [CP PRO to go].”
The underlined parts/embedded clauses exemplify the T-V combination at issue. T[uCase] values null Case of the subject. This is why the subjects in (21) are all PRO, which does not require Case-revision. In (21a), V[\text{uPhi}] assigns accusative Case to the object, so that the objects are realized as “him.” In (21b), the tough-subject “he” first gets null Case from V[uCase] “please” when occupying the direct object position. Since null Case is valued case, there is no crash when the embedded VP containing the lexical DP with null Case is transferred. A second Case (i.e. nominative) is later assigned by the matrix finite T. Case is successfully revised and the pronominal DP can be pronounced as “he” and so there is no recoverability violation. In (21c), the embedded V does not assign any Case to PRO. Rather, PRO gets null Case from the embedded T[uCase]. The derivation converges.

Finally, the third set includes three V types co-occurring with T[\text{---}] as follows:

\begin{equation}
(22) \quad T[\text{---}] \quad \begin{array}{ll}
a. & V[\text{uPhi}] \rightarrow \text{e.g. He seems to } \underline{\text{please Mary.}} \\
b. & V[\text{uCase}] \rightarrow \text{e.g. He seems to } \underline{\text{wash.}} \\
c. & V[\text{---}] \rightarrow \text{e.g. He seems to } \underline{\text{be arrested.}} \\
\end{array}
\end{equation}

In (22), T does not have any phi-features triggering Agree because C is missing. That is, this T is the kind selected by a raising predicate. In (22a), V[\text{uPhi}] assigns accusative Case to the object “Mary.” Then, the subject of the embedded clause “he” cannot get Case from T[\text{---}], so that the matrix finite T assigns nominative Case to it. In (22b), [uCase] on “wash” triggers Case-agreement with an object PRO. Again, “he” moves and gets Case from the matrix finite T. In (22c), V[\text{---}] does not trigger any agreement just as T[\text{---}] does not, so “he” finally agrees with the matrix finite T.

As illustrated above, each of the nine combinations works to derive each type of sentence. Given this, there is no need to be troubled with two of the three questions we discussed in section 15.2.2: (1) What distinguishes a finite T which assigns nominative Case, and a control T, which assigns null Case to PRO, given that both types of T have identical [\text{uPhi}] under Chomsky (2001, 2007, 2008)? and (2) why is there a null Case assigning T but there is implicitly assumed to be no null Case assigning V? For the first question, our answer is that those two different Ts are distinguishable in terms of their feature content. In response to the second question, now we can say that there is a null Case valuing V like there is a null Case valuing T, thereby eliminating the cross categorial asymmetry. In the next section, we will tackle the remaining question about tough-constructions and improper movement based on Feature-Splitting Internal Merge.
15.4 Feature-splitting: separation of [Case] and [Phi]

15.4.1 How to distinguish ("ungrammatical") improper movement from ("grammatical") tough-constructions

In this section, we especially consider what kind of mechanism allows us to derive tough-constructions (see (16)–(18)) but rule out classic improper movement (see example (3)). Under the view that a tough-subject itself moves from the embedded object position to the matrix subject position, the derivation for tough-constructions goes through the same derivational steps as improper movement does, namely movement from A-to-A’-to-A positions as follows:

(23) *John seems (that) leaves.
    [[TP John seems [CP <John> (that) [TP <John> leaves]]]]

(24) John is easy to please.\(^{11}\)
    [[TP John is easy [CP <John> [TP to please <John>]]]]

However, the outcome is different: (23) is ungrammatical while (24) is grammatical. The main goal of this section is to explain why.

15.4.2 Another way of feature-splitting

In section 15.2.1, we briefly reviewed Feature-Splitting Internal Merge suggested in Obata and Epstein (2008, 2011). Recall the way in which features get split depends on whether those features participate in Agree with T/V. Consider the following improper movement case:

(25) *Who seems left?

After “who” is merged into Spec-\(v\)P in the embedded clause, the embedded finite T whose features are inherited from C phi-agrees with it. Then, EF on T and C attracts “who” simultaneously as follows:

---

\(^{11}\) Tough-constructions are prohibited when a subject in the embedded clause is moved as in “*John is easy to please Mary.” This is an interesting case of overgeneration made possible by our more permissive system which allows improper movement derivations of tough-constructions, and allows a lexical DP to receive null Case, provided it is revised—a condition satisfied in this ungrammatical example. We speculate that this subject-object asymmetry is properly ascribed to ECP which, in GB, prohibited movement to be launched from an ungoverned position licensing PRO (cf. Epstein 1984). Also, the same discussion applies to another case such as “*John is easy to be arrested.” (See Obata and Epstein, in prep, for further discussion of comp-trace effects and “for” as a Case-valuer.)
In this context, features on “who” are split into the two syntactic landing sites as a consequence of simultaneous attraction by C and T:

\[
\begin{array}{c}
\text{(26) } \text{[CP C [TP T[U\Phi] [VP who[U\Phi][I\Phi][U\Case][I\QP] \ldots]]]} \\
\end{array}
\]

We suggested that T attracts features it agreed with and C attracts the rest of the features, which can be induced from Chomsky’s (2007) A/A’-distinction: “A-movement is IM (Internal Merge) contingent on probe by uninterpretable inflectional features…” (Chomsky 2007: 24). Therefore, since both [uCase] and [iPhi] on “who” were involved in phi-agreement with T both go to Spec-T, while [iQ] on “who” moves to the edge of CP, so that “who” at the edge of CP bears only [iQ] but has no [Phi] and [Case]. Since the embedded TP is transferred, and “who” in the embedded SpecCP lacks phi-features, this leaves [uPhi] on the matrix T unvalued, since it can find no goal DP bearing [iPhi]. Therefore, (25) crashes.

If some mechanism were to make it possible that “who” at the edge of CP were to keep [iPhi], then this derivational variant of (27) would successfully converge, since matrix T could phi-agree with “who” in Spec-CP. What is “some mechanism?” Recall that we suggested in section 15.3 that there are three kinds of T/C and three corresponding kinds of V/v and demonstrated what kinds of sentences are derived in each of the combinations. Given that, let us examine exactly how tough-constructions are derived:

\[
\begin{array}{c}
\text{(27) } \text{[CP who[I\QP][U\Phi][I\Phi][U\Case][I\QP] \ldots]} \\
\end{array}
\]

(28) John is easy to please.

Tough-predicates do not take a finite CP, so that sentences like “John is easy that Bill pleases” and “It is easy that Bill pleases John” are ungrammatical. That is, tough-predicates always select T[U\Case], which is a null Case assigner. With respect to the embedded V, since “please” in (28) is optionally an accusative or null Case valuing verb, it is either V[U\Phi] or V[U\Case]. If V[U\Phi] is chosen, then the object receives accusative Case. That is, a sentence like “It is easy to please John” is derived. As discussed earlier, tough-subjects get Case in the course of the derivation first from “please” (null Case) and later from the matrix T (nominative Case). In other words, Case is revised, stacked, or multiply assigned. V[U\Case] values null Case, so there is no crash of previously transferred phase head complements as explained in the last section. Now, let us see how a tough-construction is derived,
precisely. (NB: The capital letters “HE” below stand for a phonetically unspecified pronoun.)

(29) He is easy to please.

\[\text{Step 1: } [\text{VP PRO } \text{v} [\text{VP } \text{V} [\text{uCase}] \text{HE} [\text{iPhi}] [\text{uCase}] [\text{Op}]]] \]

\[V \text{ Case-agrees with HE and null Case is valued.}\]

\[\text{Step 2: } [\text{VP } \text{HE} [\text{iPhi}] [\text{Op}] \text{ PRO } \text{v} [\text{VP } \text{HE} [\text{null}] \text{ V} [\text{null}] \text{HE} [\text{iPhi}] [\text{uCase}] [\text{Op}]]] \]

Feature-Splitting: V attracts only the MATCHING [uCase] and v attracts the rest, crucially, [Op] and [iPhi].

\[\text{Step 3: } [\text{CP } \text{HE} [\text{iPhi}] [\text{Op}] \text{ PRO } \text{T} [\text{uCase}] [\text{VP } \text{HE} [\text{iPhi}] [\text{Op}] \text{ PRO } \text{v} [\ldots]]] \]

\[EF \text{ on C attracts HE bearing [iPhi] and [Op] to Spec-CP.}\]

\[\text{Step 4: } [\text{TP PRO } \text{HE} [\text{iPhi}] [\text{Op}] \text{ V} [\text{null}] \text{HE} [\text{iPhi}] [\text{uCase}] [\text{Op}]]] \]

\[\text{Matrix } T [\text{uPhi}] \text{ phi-agrees with HE in Spec-CP.}\]

(29) is the derivation for tough-constructions. Notice in Step 2 that feature-splitting takes place in a different way from the ungrammatical improper movement case (25). Here, the only feature which participated in Agree with “please” \(V_{\text{uCase}}\) is [uCase] on “HE”. That is, [iPhi] and [Op] are not involved in this relation. This is why V attracts only [uCase] to Spec-VP and the rest of the features are attracted by v. This makes a crucial featural distinction between the ungrammatical improper movement case and the grammatical (also improper movement) tough-construction. Only in the latter does, “HE” in Spec-CP retain [iPhi], thereby allowing the matrix finite T to avoid crash by agreeing with it. A recoverability violation is also averted since the under-specified phonetic features on “HE” are specified nominative under phi-agreement with the matrix finite T. Another point which is worth mentioning is that in Step 3, \(T_{\text{uCase}}\) Case-agrees with subject PRO across the intervening “HE.” How is this possible? Assuming with Chomsky (2001) that each feature functions as an independent prober, \(T_{\text{uCase}}\) can indeed agree with PRO\(_{\text{uCase}}\) across the intervening “HE”, and this is possible precisely because the [uCase] feature of “HE” has already been split off and was moved to Spec-VP. Thus, there is no Case-feature intervention effect in such derivations. (See Obata and Epstein 2008, 2011 for discussion of other predictions.)
(30) *John seems that Mary likes.

If the $V_{[\text{uCase}]}$ form is selected for the embedded verb “likes,” then “John” can move keeping $[\text{iPhi}]$, and it would seem as if $[\text{iPhi}]$ on “John” in embedded Spec-CP can then value $[\text{uPhi}]$ on the matrix T overgenerating (30). However, consider the configuration right after features on “John” are split and T and C are introduced:

\[
[\text{CP} \ C \ [\text{TP} \ [u\Phi] \ \text{John} [\text{iPhi}][\text{Op}] \ \text{Mary} [\text{iPhi}][\text{uCase}] \ \text{v} […]]]
\]

Recall, $[\text{uCase}]$ on “John” has already been split off having moved to Spec-VP. $[\text{iPhi}]$ and $[\text{Op}]$ on “John” are attracted by $\text{v}$ and land at the outer SpecvP position. Then, T and then C are introduced. After $[\text{uPhi}]$ on C is inherited by T, T attempts to phi-agree with “Mary” across the intervening “John,” which recall also has $[\text{iPhi}]$. However, the intervening $[\text{iPhi}]$ on “John” causes an intervention effect, so that phi-agreement between T and “Mary” is NOT allowed. As a result, $[\text{uCase}]$ on “Mary” remains unvalued causing crash of the derivation. This configuration contrasts with Step 3 in (29), where T Case-agrees with “PRO” across “HE” lacking [Case]. Our analysis can also exclude object improper movement.

Notice our analysis predicts that if the downstairs verb “likes” has a full phi-feature set, then $[\text{iPhi}]$ on “John” will value these features and will split off and move to Spec-VP. As a result, John in outer SpecvP lacks $[\text{Phi}]$ (and [Case]) thereby allowing the embedded T to agree with “Mary” across “John.” However, now $[\text{uPhi}]$ on the matrix T cannot be valued by “John,” since “John” lacks phi-features. But now two different scenarios require consideration: [1] Suppose an expletive, not John, is used to value $[\text{uPhi}]$ on the matrix T, or [2] Suppose the matrix T is the phi-less “raising-type T.” Case [1] is simply the grammatical: “It seems that Mary likes John.” Case [2] is an ungrammatical example such as “*John to seem that Mary likes _ is unfortunate,” (it means “for it to seem (to Sam) that Mary likes John is unfortunate”). This important possible overgeneration needs to be further investigated but we leave it for future discussion. Note under Epstein and Seely’s (2006) analysis (different from the principles assumed here), the structure is not generable precisely because “pure-EPP movement,” i.e. movement to Spec of phi-less T, never occurs (since there is no EPP). (This and other possible landing sites for “John” in such a derivation require further investigation.)

As demonstrated in this section, our feature-splitting system can derive tough-constructions by “proper improper movement” but concomitantly excludes “classic”
improper movement. Now, we can answer the last question “what kind of mechanism allows us to derive tough-constructions improperly but rule out classic improper movement?” These two constructions select different types of T and V (in accordance with their selectional restrictions), so that features are split in different ways depending on the feature content of the probes. Only in tough-constructions, (with a null-Case valued object) does the agreeing downstairs V (“please”) bear only [uCase]. This in turn makes it possible for the phonologically unlicensed moving direct object to keep its phi-features, carrying them to the edge of the embedded CP. This in turn permits a higher [uPhi] T to undergo agree with this Spec-CP, thereby permitting convergence and averting a recoverability violation, even though “improper movement”—reanalyzed here as a feature-variant local agreement phenomenon—has occurred. The corresponding derivation in which the tough predicate is replaced by a raising predicate crashes.

A final note concerns the noted absence of PP null operators. Chomsky (1982) suggests that null operator movement takes place in tough-constructions. Browning (1987) presented the following data to show that null operators can not replace PP:

(32)  a. *With a hammer is dangerous for John to fix the television.
     b. *On Tuesday is difficult for John to come to a meeting.
     c. *To Bill is difficult for John to give a book

(Browning 1987)

That is, under the null operator analysis, it is necessary to stipulate that tough-constructions are allowed only for DP but not PP. On the other hand, our analysis (which does not appeal to null operators) excludes the sentences in (32) since PP does not have [iPhi], so that [uPhi] on the matrix T remains unvalued, just like in ungrammatical improper movement cases.12

This lends further empirical support to our approach to tough-constructions based on feature-splitting.

15.5 Discussion and outstanding issues

We have proposed the following:

- There are three kinds of C/T and also three corresponding kinds of v/V.
- V (like T) can assign null Case to an object.

12 We must assume that Agree does not go into PP, so that it cannot “see” DP bearing [iPhi] inside of the PP, but this too requires further research.
Case-revision in tough is explained by saying that a matrix finite T phi-agrees with a moving element whose phonetic features are underspecified because of null Case assigned by V_{[uCase]}.

- A-A’-A (so-called “improper”) movement is NOT universally excluded.

- Both ungrammatical improper movement and grammatical tough-constructions are explicable by Feature-Splitting Internal Merge and the (non-)intervention effects which splitting induces, given individual feature probing.

If the discussion here is on the right track, several hidden problems regarding properties of T and V are clarified while Feature-Splitting Internal Merge gains further theoretical and empirical support. Classic improper movement is re-analyzed as a local, featural agreement failure (i.e. crash) (unrelated to unbounded Condition C application or Chain Conditions). The analysis, if successful, is restrictive enough to prohibit classic improper movement while concomitantly allowing “improper movement” derivations of tough-constructions. Other aspects of tough-constructions not explored here of course require further research, as does the ungrammaticality of subject-tough movement and PP-tough movement, as (all too) briefly discussed here. One very promising general aspect of this type of Minimalist analysis, if on track, is that “construction-specificity” might be accurately reducible to an independently motivated, partially parameterized theory of arguably irreducible syntactic features and morpho-syntactic variation.
References


Abels, Klaus and Kristine Bentzen (2008). ‘Are movement paths punctuated or uniform?’, manuscript, UCL and University of Tromsø.


References


References


Cardoso, Adriana and Mark de Vries (2011). ‘Internal and External Heads in Appositive Constructions’, manuscript, University of Lisbon and University of Groningen.


References
References


Frampton, John (2004). ‘Copies, traces, occurrences, and all that: evidence from Bulgarian multiple wh-movement’, manuscript, Northeastern University, Boston, Mass.


References


References


Iatridou, Sabine and Ivy Sichel (2009). ‘Negative DPs and Scope Diminishment’, manuscript, MIT and Hebrew University.


Janke, Vikki and Ad Neeleman (2009). ‘Ascending and descending VPs in English’, manuscript, University College London (lingBuzz/000861).


Lin, Vivian (2002). *Coordination and Sharing at the Interfaces*. Doctoral dissertation, MIT.


Maastricht, Lieke van, Marlies Kluck, and Mark de Vries (2010). ‘How invisible are Transparent Free Relatives?’ Talk presented at the Tabu-day, 3 June 2010, University of Groningen.


Obata, Miki (2009). 'Feature-Splitting of Phonological Features', manuscript, University of Michigan, Ann Arbor.

Obata, Miki and Samuel David Epstein (2008). 'Deducing Improper Movement from Phase Based C-To-T Phi Transfer: Feature-Splitting Internal Merge', manuscript, University of Michigan. [Presented at Carson-Newman College (Exploring Crash-Proof Grammars, 29 February to 1 March 2008), at Newcastle University, (The 31st GLOW, 26–28 March 2008), and at University of California, Los Angeles (The 27th WCCFL, 16–18 May 2008); written version to appear in *Proceedings of WCCFL 27.*]


Obata, Miki and Samuel David Epstein (in preparation) “‘For-To’ Revisited’, manuscript, University of Michigan, Ann Arbor.


Richards, Norvin (2001a). ‘A Distinctness Condition on Linearization’, manuscript, MIT.


Rizzi, Luigi (2002). Locality and Left Periphery, manuscript, Università di Siena.


References

Sauerland, Uli (2000). ‘Syntactic Economy and Quantifier Raising’, manuscript, University of Tübingen.


References


References


Language Index

Afrikaans 209, 218–19, 242
Assamese 133–4

Bangla 218, 229
Basque 207–8, 284, 286
Bengali 224–5, 238
Blackfoot 284
Breton 283, 285
Bulgarian 56, 289, 291

Chukchee 284
Croatian 93–4, 96–8, 101, 103, 112
Czech 131–2

Dutch 84, 154, 162–6, 219, 220–1, 229, 238, 331


Finnish 208–9, 215
French 209, 291, 294

German 21, 24, 30, 32, 43, 94, 97, 105, 107, 112, 176, 208, 211, 215, 217–18, 219–223, 225–7, 229, 235, 238

Hindi-Urdu 210, 219, 221, 229, 238, 284
Hungarian 367, 376

Icelandic 245, 248–50, 252, 255, 269–70, 277, 290
Italian 71, 84, 213, 291, 372
Itelmen 284


Khwarshi 284
Korean 324
Kutchi Gujarati 284

Malayalam 207
Mandarin Chinese 212
Mande languages 207
Marathi 210, 224–5
Mumuye 212

Persian 218–9, 221, 223, 229
Portuguese: Brazilian 136, 138
European 118, 136–8, 218
Quichua 236

Romanian 129–30

Spanish 66–9, 71–3, 75–6, 79, 81, 84–5, 87, 91–2, 293, 331, 333, 343, 345

Telugu 133–4
Tsez 284–6
Turkish 207, 222–3, 231, 238

Vata 132–3
Vietnamese 213

West Flemish 213, 265
West Germanic (general) 265
Subject Index

*-marking 200–2
3rd factor considerations 254, 261–3, 268

A*-movement, see movement
Across-the-Board (ATB) movement, see movement
activity condition, see also
agreement, adjunction
A-over-A Condition 285
3

adject:
adjunct control 118–21, 133–8
adjunct islands 123, 337
Adjudgent Condition (effects) 330–333
adjudication 42, 56, 117–18, 120, 158, 256, 258
adverb(ial) 95, 220–1, 242, 291–3, 335, 341, 351–6, 359, 361–2
agent, see argument structure
Agree-related movement 214, 219
Agree-related ^ 235, 239
agreement, see also long-distance agreement 25, 140, 240, 248, 255, 258, 265, 269–73, 275–6, 278, 283–7, 294–6, 369–71, 373, 376–8, 380–4
number agreement 140, 278
phi-agreement 265, 284–5, 369–71, 373, 376, 380–1
rich agreement 283
amalgam 46, 144, 151, 159–166
A-movement, see movement
anaphor, see also binding 157, 164, 176, 251, 295, 328
Antecedent Contained Deletion 176, 306
antilocality 86
anti-reconstruction, see reconstruction
anti-superiority, see also superiority 272, 291
A-over-A Condition 285, 287–8
argument structure 67–8, 71–2, 74, 76, 82, 84, 89, 92, 236, 273
agent 66–70, 79, 81–5, 87–92, 140
initiator 69–70, 74–5, 77, 80, 85, 88, 90
instrument 69–70, 79, 90
path 72–3, 82
patient 84
aspect:
Aspect Phrase (AspP) 90, 347
inner aspect 331, 346, 355, 362
ATB extraction 67, 131–2
atomization 228, 234, 236, 239–41
Attract (operation), see also Specificity Principle 293, 297, 299–30, 318, 321–2, 327, 329
Bare Phrase Structure (BPS) 42, 76, 78, 80–1, 207, 231
barrier 152, 199–200, 330–3, 340, 352, 358–9, 361, 364–5
binding:
Condition A 251, 327
Condition C 155, 158, 164, 166, 176, 218, 366, 384
bottom-up computations/derivations 123, 131, 157, 164, 209
Case, see also valuation (of Case):
accusative Case 82, 87, 223, 250, 371, 376–8, 380
Case features, see (Case) features
Case Resistance Principle (CRP) 221
Case revision 376, 378, 384
dative Case 73, 101, 277–8
ergative Case 287
inherent Case 126
nominative Case 222, 269, 277, 284, 367, 371, 373, 376–8, 380
null Case 367, 371, 373–81, 383–4
quirky Case 147, 278
c-command (definition of) 100, 156, 232
chains, see also copies; linearization; Linear Correspondence Axiom (LCA) 30, 34, 121, 126, 127, 129–30, 144, 148, 169, 183–5, 190, 197, 202, 266, 298, 305
Chain Reduction (operation) 127–32
clitic 93–105, 110, 112, 131–2, 207, 293–5
cliquic doubling 293–4
complementizer 81, 124, 136, 140, 210–11, 224–5, 239
Complex NP Constraint 228
compositionality 35, 322
computational system 114–17, 119–21, 123–5, 127, 137, 227, 241, 261, 374
Conceptual–Intentional (CI) features, see features
Conceptual–Intentional Interface (CI) 193, 254, 257–8, 260–1, 264, 374
Condition on Extraction Domains (CED) 219, 231, 330
constituent structure set (CSS) 19–20, 24–5, 27–9
Constraint on Sharing (COSH) 94, 107–11, 113
Control (configurations) 118–21, 133–38, 222, 224, 239–40, 267, 367–8, 371–3, 377, 378
coordination 48, 50, 53–5, 58, 106–7, 112, 131, 139, 140–1, 151–2, 155, 159, 165–6, 172, 185, 237
copies, see also chains; linearization; Linear Correspondence Axiom (LCA) phonetic realization of copies 128
pronunciation of lower copies 129, 132
pronunciation of multiple copies 133–4
copy operation 122, 126
copy theory (of movement) 20, 62, 116, 118, 126, 133, 142, 150, 231–2
counter-cyclic Internal Merge, see Merge
counter-cyclic merger, see Merge
covert movement, see movement
c-selection 216
c-selection-related 214, 226–7, 235, 238
cyclic Transfer, see Transfer
deletion, see also ellipsis; Antecedent Contained Deletion:
non-constituent deletion 204
scattered deletion 204, 220, 232–3, 235
VP deletion 189
defective intervention (in Icelandic), see intervention effects
Dependency Grammar 22, 39, 42–3
Dislocation 293, 297–9, 302, 308, 310, 319, 321–2, 329
Distributed Morphology 64, 274
distributive reading (of quantifiers), see quantifier
ditransitive (predicates) 85, 221, 290, 316
dominance, see also Unambiguous Domination:
dominance set 21–3, 26, 29–30
reflexive dominance 26–8, 35, 38
double object construction 33, 126, 304, 312, 315, 318, 322, 329
clicks, see also copies; linearization; Linear Correspondence Axiom (LCA)
complementizer 81, 124, 136, 140, 210–11, 224–5, 239
Complex NP Constraint 228
compositionality 35, 322
computational system 114–17, 119–21, 123–5, 127, 137, 227, 241, 261, 374
Conceptual–Intentional (CI) features, see features
Conceptual–Intentional Interface (CI) 193, 254, 257–8, 260–1, 264, 374
Condition on Extraction Domains (CED) 219, 231, 330
constituent structure set (CSS) 19–20, 24–5, 27–9
Constraint on Sharing (COSH) 94, 107–11, 113
Control (configurations) 118–21, 133–38, 222, 224, 239–40, 267, 367–8, 371–3, 377, 378
coordination 48, 50, 53–5, 58, 106–7, 112, 131, 139, 140–1, 151–2, 155, 159, 165–6, 172, 185, 237
copies, see also chains; linearization; Linear Correspondence Axiom (LCA) phonetic realization of copies 128
pronunciation of lower copies 129, 132
pronunciation of multiple copies 133–4
copy operation 122, 126
copy theory (of movement) 20, 62, 116, 118, 126, 133, 142, 150, 231–2
counter-cyclic Internal Merge, see Merge
counter-cyclic merger, see Merge
covert movement, see movement
c-selection 216
c-selection-related 214, 226–7, 235, 238
cyclic Transfer, see Transfer
deletion, see also ellipsis; Antecedent Contained Deletion:
non-constituent deletion 204
scattered deletion 204, 220, 232–3, 235
VP deletion 189
defective intervention (in Icelandic), see intervention effects
Dependency Grammar 22, 39, 42–3
Dislocation 293, 297–9, 302, 308, 310, 319, 321–2, 329
Distributed Morphology 64, 274
distributive reading (of quantifiers), see quantifier
ditransitive (predicates) 85, 221, 290, 316
dominance, see also Unambiguous Domination:
dominance set 21–3, 26, 29–30
reflexive dominance 26–8, 35, 38
double object construction 33, 126, 304, 312, 315, 318, 322, 329
Earliness Principle 273–4
economy, see also Scope (Economy) 65, 118, 129, 134, 137, 195–6, 256, 306–8
Edge Feature (EF), see feature edge restrictions:
Left Edge Restriction 181, 185, 194, 196–9
Right Edge Restriction (on RNR) 170, 175, 180, 194–6, 198–200
ellipsis, see also deletion; Left-Node-Raising:
Right-Node-Raising:
VP ellipsis 46–54, 56, 63, 95, 105, 189, 306–7
Empty Category Principle (ECP) 202, 379 equidistance 272, 279–82, 296
Extended Projection Principle (EPP) 214, 247–8, 251, 265–70, 382
Extended Projection Principle (EPP) (cont.)
EPP feature, see feature
Extension Condition 115, 118, 120, 124–5, 142, 256, 258
External Merge, see Merge
eternal remerge, see also Parallel Merge;
  multidominance 144–5, 147–8, 151–3, 159–66
extraction, see movement
extraposition 184, 294
  CP-extraposition 216–8, 220–2, 224, 226,
  229, 233–9, 242
features:
  Case features 56, 277, 287, 289, 290, 294,
  296, 297, 298, 299, 300, 301, 302, 303
  Conceptual-Intentional (CI) features 257, 259, 260, 264
discourse (related) features 56
Edge Feature (EF) 253, 258, 263, 265–6, 268,
  270, 269, 270, 271, 278, 281, 286, 287
ellipsis feature (E-feature) 47
EPP feature, see also Edge Feature 227,
  227–8, 228
feature checking 20, 31, 129, 247, 250, 355
Feature-Inheritance 253, 258–9, 263–6,
  268–70, 366, 369
feature matching 265–6
feature movement, see movement
feature valuation 129, 257, 260, 373
focus feature 184, 292
formal features 81
interpretable features 257–60, 273–4
operator feature 31
phi-features (ϕ-features) 253, 255,
  257–9, 263–6, 269–70, 277, 283–4,
  286–7, 294–5, 315, 368–71, 375, 378,
  380, 382–3
phonological feature 258–60, 273, 278,
  370, 374–7
strong features 136–8, 182
uninterpretable features 258–9, 273–4
unvalued features 257, 260, 370, 373, 374, 377
valued features 257–60, 264
wh-feature 203–4, 206
Feature-Splitting Internal Merge 366–8, 370,
  378–9, 384
floating quantifier (FQ), see also quantifier
  float 246, 249–50, 315, 317–23, 325–7,
  331, 333, 335–43, 345–9
focus 31, 56, 133, 184, 249, 292–3
Focus Phrase 56
freezing effect 199, 219, 318
functional category 295, 335
functional head 32–3, 54, 70, 76, 79, 92,
  161, 349
functional projection 65, 71, 76, 79–82, 86,
  90–1, 161, 166, 316–17, 322
fusion 133–4, 141
gapping, see ellipsis
Government and Binding Theory (GB) 114,
  116, 118, 127, 202, 267, 379
head parameter, see also VO language; OV
  language 221
head-final 207–9, 211, 213–17, 221, 224,
  227, 229, 231, 234, 238–9, 241, 243, 327
head-initial 208–9, 211, 213, 215–17, 224–7,
  229, 231–2, 234, 237–9, 241, 243
heaviness 200
heavy NP shift 184, 194
improper movement, see movement
inactivity condition 266
Inclusiveness Condition 116, 259
inherent Case, see Case
initiator, see argument structure
Initiator Phrase 74, 90
instrument, see argument structure
interface conditions, see also PF (output)
  condition 113, 261, 295
Internal Merge, see Merge
intervention effects 350, 384
defective intervention (in Icelandic) 277, 290
inverse linking 304, 318, 320, 322–4, 329
invisibility 144, 155, 157, 162–3, 166, 254, 266
island, see also adjunct island:
  island condition 169–70, 182, 186, 190, 192,
  200–1, 204–5
island (in)sensitivity 202–5
islandhood 219, 227–8, 236, 364–5
strong island 219, 227–8, 231, 233, 236–7, 242
weak island 219

Lasnik Effects 169, 190
Last Resort 117, 121–4, 126, 137, 232, 287, 293
Law of Conservation of Features (LCF) 259
Law of Conservation of Relations (LCR) 256
Left Edge Restriction, see edge restrictions
Left-Node-Raising (LNR), see also Right
Node Raising 169, 179, 181, 183, 191–3, 195–7, 200–1, 205
lexical array 227–8, 244, 254
lexicon 23, 64, 66, 79, 84, 90, 92, 137, 146, 222, 265, 368
Linear Correspondence Axiom (LCA) 19, 93, 96–7, 99, 103, 110–12, 127, 133–4, 183, 206–8, 213, 216, 221, 227, 230–2, 238–9, 243
linearization, see also copies; Linear
Correspondence Axiom (LCA); Order
Preservation Condition; partial order;
locality, see also antilocality; non-locality 58, 152, 163, 165, 169–71, 178–9, 182, 187–90, 195–201, 205, 272, 279, 282–5, 289, 296
locality insensitivity 173–4, 197, 205
Logical Form (LF) 26, 113, 154, 166, 182, 202, 241
long-distance agreement 271, 285–6, 294
look-ahead 123, 240

Maximize Matching Effects (MME) 273–5
Merge, see also remerge; Parallel Merge:
counter-cyclic Internal Merge 253–6, 261, 263, 267–8, 270
counter-cyclic merger 270
External Merge (EM) 44–5, 148, 255, 273
Internal Merge (IM), see also Feature-
Splitting Internal Merge 44–5, 109, 148, 213, 227, 253, 255, 366–70, 378–80, 384
Parenthetical Merge 162, 164, 166
Merge-over-Move 135, 137–8

Minimal Link Condition (MLC) 272, 274–5, 279–82, 285, 287–8, 291, 294, 296
Anti-MLC 272, 281, 289, 294, 296
Minimalist Program 118, 143, 205, 261
Minimality, see also Relativized Minimality 42, 121, 281, 303, 322
morphological merger 141–2
movement, see also copy theory (of
movement); predicate fronting; Vacuous
Movement Hypothesis 214, 219
A’-movement, see also parasitic gap 33, 245, 247, 251, 259, 343, 368–9, 376
A-movement 33, 126, 245, 247, 251, 258, 368, 369, 380
covert movement 112, 182–3, 200–1, 204–5, 297, 312, 315
improper movement 246–7, 366–7, 370–1, 376–9, 381–4
Move-F(eature) 203–4
overt movement 169, 182, 185–7, 189, 200–5, 316
parallel movement 245, 247–52
remnant movement 56, 60, 63, 192, 207
sideward movement 45, 114, 117–37, 144, 148, 165, 183, 345–6
string-vacuous movement 169, 189–95, 197–9, 205
successive cyclic movement 136–7, 152
tough movement 384
multidominance, see also Parallel Merge 20, 24, 29–30, 43–6, 66–7, 74–5, 83, 92, 93, 144, 147–9, 153, 165, 175, 263
multiple Spell-Out 227, 233
multiple wh-questions, see wh-questions
negation 53–4, 106, 194
nominalization 68–73, 75–6, 79–84, 87–92, 223, 236
non-locality 144, 153
No-Tampering Condition (NTC) 253, 369
null operator, see operator
numeration 114, 124, 127, 135, 138, 140, 144, 147, 182, 241
renumeration 232, 236
numeration set 21–3, 25–6, 29–30, 33–4, 40–1, 43
object shift 245, 249–52, 307, 317
operator 35, 36, 37, 83, 153, 158, 321
null operator 367, 383
operator feature, see feature
tense operator 90
Order Preservation Condition 172, 175–7, 180, 184, 192, 194, 196
order-preserving movement 272, 281–2, 289
OV language, see also head parameter 211–12, 217–21, 225–6, 229, 237, 239, 243
overt movement, see movement
Parallel Merge, see also multidominance
44–6, 62–3, 65, 147, 165
parallel movement, see movement
Parallelism Requirement 140–1, 198
parameter, see also head parameter 215–16, 221, 226, 269–70
parasitic gap 46, 118–19, 121, 127, 151, 182
parenthesis (parentheticals) 46, 95, 144, 153–61, 164, 166
Parenthetical Merge, see Merge
partial order 21–3, 27, 40
path, see argument structure
patient, see argument structure
PF Adjacency Condition 203
PF effect 166, 169, 182, 184, 190, 192, 199, 205
PF operation 32, 71
PF (output) condition, see also *-marking
181, 182, 186–7, 189, 203–50
phase, see also spell-out domain 81–2, 92, 123–6, 137, 144, 146, 152, 155, 172, 227–8, 236, 240–1, 244, 250–66, 269–70, 272–3, 279–80, 282, 287–88, 292, 298, 367, 370–1, 380
phase edge 137, 241, 258–60, 264, 287, 298
Phase Impenetrability Condition (PIC) 81, 251–2, 255, 258–9, 263–4, 272, 279, 280, 282, 363–4, 370, 381, 382
phi-agreement, see agreement
phi-features, see features
phonetic/phonological component 116, 126, 128–9, 159, 260
pied-piping 203–4
power set 23, 26, 40
predicate fronting 323–4
P(reposition)-stranding 171–2, 187, 198, 316
PRO, see Control (configurations)
Process Phrase (ProcP) 75, 91–2
pseudogapping, see ellipsis
quantifier, see also distributive reading;
inverse linking; scope:
distributive reading (of quantifiers) 176, 179
negative quantifier 324
quantifier float (Q-float), see also floating quantifier 246, 249–50
quantifier raising (QR) 106–7, 112, 193, 304–7, 312, 315, 324
raising (configurations) 224, 270, 328, 367–8, 378, 383
reconstruction 37, 315, 317, 324–9
anti-reconstruction 157, 326
relational structure 21, 28, 30, 41, 43
Relativized Minimality (RM) 213, 215, 217
remerge 32, 45, 67, 74, 76, 78, 80–2, 144, 146–9, 151–3, 159–66, 300–2, 319–20
remnant movement, see movement
renumeration 232, 236
reprojection 30, 33, 67
Right Edge Restriction, see edge restrictions
Right-Node-Raising (RNR), see also Left Node Raising 151, 169–90, 193–5, 198–205
Right Roof Constraint (RRC) 169, 171, 182, 184, 186–7, 190, 200, 205
root syntactic object 114–18, 121, 123
scope, see also inverse linking; negation;
quantifiers:
intermediate scope 304, 315, 318, 320, 322
inverse scope 316, 324
narrow scope 304, 318, 324–5
Scope Economy 306–7
successive cyclic movement, 193
superiority, see also 271
syntactic component 129, 132, 142

secondary predicate(s):
deictive secondary predicate 332–3, 339, 343
object-oriented secondary-predicate 332–3, 335, 344, 358, 365
subject-oriented secondary-predicate 332, 335, 344–5, 358

semantic component 33, 165, 256–7, 260, 326
Sensory Motor System (SMS) 374
set containment 21, 24, 26, 40
sharing, see multidomiance; Constraint on Sharing (COSH)
sideward movement, see movement
Silverstein hierarchy 296
single cycle hypothesis 169, 182
single output model 312, 315, 323
sluicing, see ellipsis
spell-out domain, see also phase 172–80, 184, 190, 195, 244
string-vacuous movement, see movement
strong features, see features
strong island, see island
Strong Minimalist Thesis 227, 254
subarray 124–5, 263
Subjacency 182
Subject Condition effects 368
Subjektlicke in finite Sätzen (SLF) 94, 97, 105–7, 110, 112
Subset Principle 271
successive cyclic movement, see movement
superiority, see also anti-superiority 272, 279, 280, 282, 293–4
syntactic component 129, 132, 142

telescope structure 20, 22–4, 28–30, 39, 41–3
that-trace effect 219, 247


Topic Phrase (TopP) 191–2
topicalization 36, 154, 249, 250, 325
VP-topicalization 323–5
tough-constructions 366–7, 370–2, 375–6, 378–84
tough movement, see movement
Transcendence 111, 113
Transfer 228, 241, 253, 254–61, 264, 266, 269–70, 370, 375–6
cyclic Transfer 241, 253, 257–9, 261, 266, 269
type driven interpretation 301–2

unaccusative (predicates) 75, 88, 241, 333, 345, 359, 361, 365
Unambiguous Domination 283, 285
unergative (predicates) 241, 333, 345, 359, 361, 364, 365
uninterpretable features, see features
Universal Grammar (UG) 113, 197, 213, 245, 345
unvalued features, see features

Vacuous Movement Hypothesis 202

valuation:
Case-valuation 264–5, 270, 370–1, 376
feature valuation, see features
valued features, see features
variable binding 54, 57, 166
verb second 292–3
VO language, see also head parameter 211–14, 218, 229
voice mismatches 46–50, 52, 57, 60–5
VP deletion, see deletion
VP ellipsis, see ellipsis
VP-topicalization, see topicalization

weak island, see island

wh-questions:
multiple wh-questions 93, 97, 280
wh-in situ 34, 36, 100, 131, 136–7, 182, 245, 247
wh-movement, see movement

word order, see also head parameter;
linearization; OV language; VO language 108, 111–12, 171, 179, 181, 206–9, 212, 229, 231, 235, 243, 324–5
OXFORD STUDIES IN THEORETICAL LINGUISTICS

1 The Syntax of Silence
Sluicing, Islands, and the Theory of Ellipsis
by Jason Merchant

2 Questions and Answers in Embedded Contexts
by Utpal Lahiri

3 Phonetics, Phonology, and Cognition
edited by Jacques Durand and Bernard Laks

4 At the Syntax-Pragmatics Interface
Concept Formation and Verbal Underspecification in Dynamic Syntax
by Lutz Marten

5 The Unaccusativity Puzzle
Explorations of the Syntax-Lexicon Interface
edited by Artemis Alexiadou, Elena Anagnostopoulou, and Martin Everaert

6 Beyond Morphology
Interface Conditions on Word Formation
by Peter Ackema and Ad Neeleman

7 The Logic of Conventional Implicatures
by Christopher Potts

8 Paradigms of Phonological Theory
edited by Laura Downing, T. Alan Hall, and Renate Raffelsieben

9 The Verbal Complex in Romance
by Paola Monachesi

10 The Syntax of Aspect
Deriving Thematic and Aspectual Interpretation
Edited by Nomi Erteschik-Shir and Tova Rapoport

11 Aspects of the Theory of Clitics
by Stephen Anderson

12 Canonical Forms in Prosodic Morphology
by Laura J. Downing

13 Aspect and Reference Time
by Olga Borik

14 Direct Compositionality
edited by Chris Barker and Pauline Jacobson

15 A Natural History of Inflection
by Alan C. L. Yu

16 Phi-Theory
Phi-Features Across Interfaces and Modules
edited by Daniel Harbour, David Adger, and Susana Béjar

17 French Dislocation: Interpretation, Syntax, Acquisition
by Cécile De Cat

18 Inflectional Identity
edited by Asaf Bachrach and Andrew Nevins

19 Lexical Plurals
by Paolo Acquaviva

20 Adjectives and Adverbs
Syntax, Semantics, and Discourse
edited by Louise McNally and Christopher Kennedy

21 InterPhases
Phase-Theoretic Investigations of Linguistic Interfaces
edited by Kleanthes Grohmann

22 Negation in Gapping
by Sophie Repp

23 A Derivational Syntax for Information Structure
by Luis López

24 Quantification, Definiteness, and Nominalization
edited by Anastasia Giannakidou and Monika Rathert

25 The Syntax of Sentential Stress
by Arsalan Kahemnuyipour

26 Tense, Aspect, and Indexicality
by James Higginbotham

27 Lexical Semantics, Syntax, and Event Structure
edited by Malka Rappaport Hovav, Edit Doron, and Ivy Sichel

28 About the Speaker
Towards a Syntax of Indexicality
by Alessandra Giorgi

29 The Sound Patterns of Syntax
edited by Nomi Erteschik-Shir and Lisa Rochman

30 The Complementizer Phase
edited by E. Phoevos Panagiotidis

31 Interfaces in Linguistics
New Research Perspectives
edited by Raffaella Folli and Christiane Ulbrich

32 Negative Indefinites
by Doris Penka

33 Events, Phrases, and Questions
by Robert Truswell

34 Dissolving Binding Theory
by Johan Rooryck and Guido Vanden Wyngaerd

35 The Logic of Pronominal Resumption
by Ash Asudeh

36 Modals and Conditionals
by Angélica Kratzer

37 The Theta System
Argument Structure at the Interface
edited by Martin Everaert, Marijana Mareli, and Tal Siloni