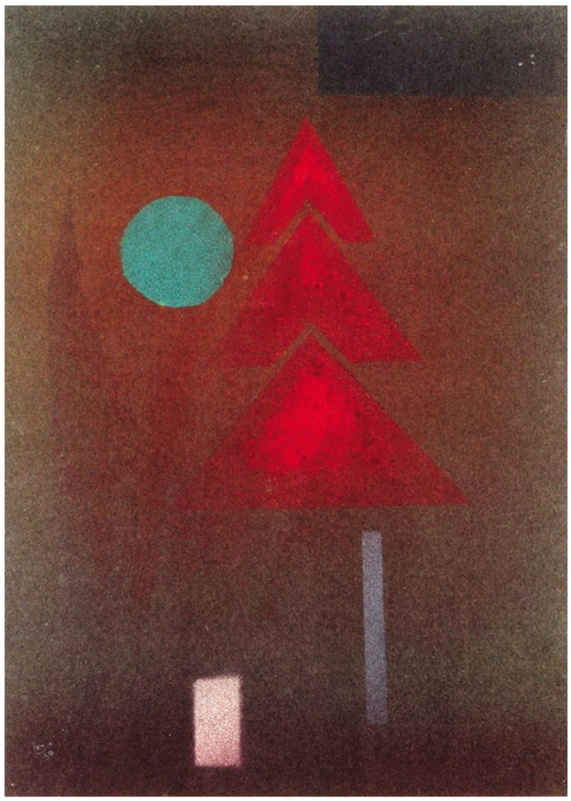
syntax

Syntax

A Generative Introduction

Fourth Edition



Andrew Carnie

WILEY Blackwell

Syntax

Introducing Linguistics

This outstanding series is an indispensable resource for students and teachers – a concise and engaging introduction to the central subjects of contemporary linguistics. Presupposing no prior knowledge on the part of the reader, each volume sets out the fundamental skills and knowledge of the field, and so provides the ideal educational platform for further study in linguistics.

1	Andrew Spencer	*Phonology*
2	John Saeed	*Semantics*, 4th Edition
3	Barbara Johnstone	*Discourse Analysis*, 3rd Edition
4	Andrew Carnie	*Syntax*, 4th Edition
5.	Anne Baker and Kees Hengevelt	*Linguistics*
6.	Li Wei	*Applied Linguistics*

Syntax

A Generative Introduction

Fourth Edition

Andrew Carnie

WILEY Blackwell®

This fourth edition first published 2021

Edition History: Blackwell Publishing, Ltd (1e, 2002 and 2e, 2007); John Wiley and Sons (3e, 2013 and 4e, 2021)

Wiley-Blackwell is an imprint of John Wiley & Sons, formed by the merger of Wiley's global Scientific, Technical and Medical business with Blackwell Publishing.

Registered Office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Offices
350 Main Street, Malden, MA 02148-5020, USA
9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, for customer services, and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

Library of Congress Cataloging-in-Publication Data

Names: Carnie, Andrew, 1969- author.
Title: Syntax : a generative introduction / Andrew Carnie.
Description: Fourth edition. | Chichester, West Sussex, UK Wiley-Blackwell, 2021. |
Series: Introducing linguistics | Includes bibliographical references and index. |
Summary: "Almost every preface to every syntax textbook out there starts out by telling the reader how different this book is from every other syntax textbook. On one hand, this is often the truth: each author shows their own particular spin or emphasis. This is certainly true of this textbook. For example, you'll be hard-pressed to find another textbook on Principles and Parameters syntax that uses as many Irish examples as this one does.. On the other hand, let's face facts. The basic material to be covered in an introductory textbook doesn't really vary much. One linguist may prefer a little more on binding theory, and a little less on control, etc. In this text, I've attempted to provide a relatively balanced presentation of most of the major issues and I've tried to do this in a student-friendly way. I've occasionally abstracted away from some of the thornier controversies, when I felt they weren't crucial to a student understanding the basics. This may make the professional syntactician feel that I've cut corners or laid out too rosy a picture. I did this on purpose, however, to give students a chance to absorb the fundamentals before challenging the issues. This was a deliberate pedagogical choice. I'm well aware that sometimes I've glossed over controversies, but I think a student has to learn the basics of how the system works before they can seriously critique and evaluate the model. This is a textbook, not a scholarly tome, so its aim is to reach as many students as possible. The style is deliberately low-key and friendly. This doesn't mean I don't want the students to challenge the material I've presented here. Throughout the book, you'll find grey "textboxes" that contain issues for further discussion or interesting tidbits. Many of the problem sets also invite the student to challenge the black and white presentation I've given in the text. I encourage instructors to assign these, and students to do them, as they form an important part of the textbook. Instructors may note that if a favorite topic is not dealt with in the body of the text, a problem set may very well treat the question."-- Provided by publisher.
Identifiers: LCCN 2020037443 (print) | LCCN 2020037444 (ebook) | ISBN 9781119569237 (paperback) |
ISBN 9781119569183 (pdf) | ISBN 9781119569312 (epub)
Subjects: LCSH: Grammar, Comparative and general--Syntax. | Generative grammar.
Classification: LCC P291 .C33 2021 (print) | LCC P291 (ebook) | DDC

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

Cover Design: Wiley
Cover Image: Wassily Kandinsky, Almost Submerged, 1930

Set in 10 point Palatino LT, Std, 10point Styx Text 2 and 10 point Helvetica LT Std by the author

Dedicated with love to my parents, Robert and Jean, and in memory of my teacher and mentor, Ken Hale

This book is accompanied by a workbook: *The Syntax Workbook: A Companion to Carnie's Syntax*, 2nd Edition available for purchase from http://www.wiley.com. The workbook is optional and serves as a supplement for students who want additional practice in syntactic analysis.

Additional online material including bonus chapters can be found on the book's website:

http://www.wiley.com/go/carnie

Contents

Preface and Acknowledgments xiii

Part 1 Preliminaries 1

1 Generative Grammar 3
0. Preliminaries 3
1. Syntax as Science – the Scientific Method 4
2. Syntax as a Cognitive Science 12
3. Models of Syntax 13
4. Competence vs. Performance 13
5. A Clarification on the Word "Language" 15
6. Where Do the Rules Come From? 16
7. Choosing among Theories about Syntax 25
8. The Scientific Method and the Structure of this Textbook 25
9. Conclusion 26
Ideas, Rules, and Constraints Introduced in this Chapter 27
General Problem Sets 29
Challenge Problem Sets 35

2 Parts of Speech 43
0. Words and Why They Matter to Syntax 43
1. Determining Part of Speech 44
2. The Major Parts of Speech: N, V, Adj, and Adv 47
3. Open vs. Closed; Lexical vs. Functional 50
4. Subcategories and Features 52
5. Conclusion 58
Ideas, Rules, and Constraints Introduced in this Chapter 58
General Problem Sets 59
Challenge Problem Sets 63

3 Constituency, Trees, and Rules 67
0. Introduction 67
1. Rules and Trees 70
2. How to Draw a Tree 84
3. Modification and Ambiguity 90
4. Constituency Tests 91
5. Constituency in Other Languages 93
6. Conclusion 98

Ideas, Rules, and Constraints Introduced in this Chapter 98
General Problem Sets 99
Challenge Problem Sets 107

4 Structural Relations 113
0. Introduction 113
1. The Parts of a Tree 114
2. Dominance 116
3. Precedence 119
4. C-command 122
5. Grammatical Relations 126
6. Conclusions 129
Ideas, Rules, and Constraints Introduced in this Chapter 129
General Problem Sets 131
Challenge Problem Sets 138

5 Binding Theory 141
0. Introduction 141
1. The Notions *Coindex* and *Antecedent* 143
2. Binding 144
3. Locality Conditions on the Binding of Anaphors 146
4. The Distribution of Pronouns 148
5. The Distribution of R-expressions 148
6. Why Does Binding Theory Matter to Syntacticians 149
7. Conclusion 151
Ideas, Rules, and Constraints Introduced in this Chapter 151
General Problem Sets 152
Challenge Problem Sets 154

Part 2 The Base 159

6 X-bar Theory 161
0. Introduction 161
1. Bar-level Projections 163
2. Generalizing the Rules: The X-bar Schema 168
3. Complements, Adjuncts, and Specifiers 170
4. Some Definitional Housekeeping 179
5. Parameters of Word Order 180
6. Drawing Trees in X-bar Notation 182
7. Conclusion 199
Ideas, Rules, and Constraints Introduced in this Chapter 200
General Problem Sets 201
Challenge Problem Sets 209

7 Extending X-bar Theory to Functional Categories 213
0. Introduction 213
1. Determiner Phrases (DPs) 214
2. A Descriptive Tangent into Clause Types 217
3. Complementizer Phrases (CPs) 221
4. Tense, Perfect, Progressive and Voice Phrases 223
Ideas, Rules, and Constraints Introduced in this Chapter 234
General Problem Sets 236
Challenge Problem Sets 238

8 Constraining X-bar: Theta Theory 243
0. Introduction 243
1. Some Basic Terminology 244
2. Thematic Relations and Theta Roles 245
3. The Lexicon 250
4. Expletives and the Extended Projection Principle 251
5. Conclusion 253
Ideas, Rules, and Constraints Introduced in this Chapter 254
General Problem Sets 255
Challenge Problem Sets 259

9 Theta Grids and Functional Categories 265
0. Introduction 265
1. Complementizers 266
2. Determiners 269
3. Using Theta Grids for English Auxiliaries 273
4. Main verbs vs. Auxiliaries 284
5. Conclusion 287
Ideas, Rules, and Constraints Introduced in this Chapter 287
General Problem Sets 288
Challenge Problem Sets 292

Part 3 Movement 295

10 Head-to-Head Movement 297
0. Introduction 297
1. Verb Movement (V $\rightarrow$ T) 300
2. T Movement (T $\rightarrow$ C) 314
3. *Do*-support 317
Appendix: Determining if a Language has V $\rightarrow$ T Movement 318
Ideas, Rules, and Constraints Introduced in this Chapter 318
General Problem Sets 319
Challenge Problem Sets 324

11 DP Movement 331
0. Introduction 331
1. A Puzzle for the Theory of Theta Roles 332
2. Passives 337
3. Case 341
4. Raising: Reprise 345
5. Passives: Reprise 347
6. Inherently Passive Verbs: Unaccusatives 348
7. DP Movement in SVO vs. VSO Languages 351
8. Conclusion 353
Ideas, Rules, and Constraints Introduced in this Chapter 353
General Problem Sets 354
Challenge Problem Sets 360

12 *Wh*-movement and Locality Constraints 365
0. Introduction 365
1. Movement in *Wh*-questions 366
2. Relative Clauses 375
3. Islands 379
4. The Minimal Link Condition 382
5. Echo Questions (*Wh*-in-situ) in English 386
6. Conclusion 387
Ideas, Rules, and Constraints Introduced in this Chapter 387
General Problem Sets 388
Challenge Problem Sets 392

13 A Unified Theory of Movement 395
0. Introduction 395
1. Move 397
2. Explaining Cross-linguistic Differences 400
3. Scope, Covert Movement, and the MLC 403
4. Conclusion 406
Ideas, Rules, and Constraints Introduced in this Chapter 407
General Problem Sets 407
Challenge Problem Sets 408

Part 4 Advanced Topics 411

14 Ditransitives 413
0. Introduction 413
1. The Problem of Ditransitive Verbs 414
2. The Active Voice Head 415
3. Object Shift 417
4. Ditransitives: Reprise 422

Ideas, Rules, and Constraints Introduced in this Chapter 426
General Problem Sets 426
Challenge Problem Sets 428

15 Raising, Control, and Empty Categories 429
0. Introduction 429
1. Raising vs. Control 430
2. Two Kinds of Raising, Two Kinds of Control 438
3. Control Theory 443
4. Another Kind of Null Subject: "Little" *pro* 446
5. Conclusion 447
Ideas, Rules, and Constraints Introduced in this Chapter 447
General Problem Sets 448
Challenge Problem Sets 450

16 Ellipsis 453
0. Ellipsis 453
1. LF-copying or PF-deletion 455
2. Antecedent-Contained Deletion and Pseudogapping 461
3. Conclusion 464
Ideas, Rules, and Constraints Introduced in this Chapter 464
General Problem Sets 466
Challenge Problem Sets 468

17 Advanced Topics in Binding Theory 471
0. Introduction 471
1. Levels of Representation 472
2. The Definition of Binding Domain 473
Ideas, Rules, and Constraints Introduced in this Chapter 479
General Problem Sets 479
Challenge Problem Sets 481

18 Polysynthesis, Incorporation, and Non-configurationality 483
0. Introduction 483
1. Polysynthesis 484
2. Incorporation 486
3. Scrambling and Non-configurationality 487
4. Conclusion 495
Ideas, Rules, and Constraints Introduced in this Chapter 495
General Problem Sets 496
Challenge Problem Sets 497

19 Merge 499
0. Introduction 499
1. External Merge 500
2. Internal Merge 505
3. Conclusion 506
Ideas, Rules, and Constraints Introduced in this Chapter 507
General Problem Sets 507
Challenge Problem Sets 508

Conclusions and Directions for Further Study 509
References 511
Index 521

Go to

www.wiley.com/go/carnie

for bonus chapters on Lexical-Functional Grammar (LFG) and Head-driven Phrase Structure Grammar (HPSG), as well as a glossary and other supplementary materials.

Preface and Acknowledgments

Almost every preface to every syntax textbook out there starts out by telling the reader how different this book is from every other syntax textbook. On one hand, this is often the truth: each author shows their own particular spin or emphasis. This is certainly true of this textbook. For example, you'll be hard-pressed to find another textbook on Principles and Parameters syntax that uses as many Irish examples as this one does.. On the other hand, let's face facts. The basic material to be covered in an introductory textbook doesn't really vary much. One linguist may prefer a little more on binding theory, and a little less on control, etc. In this text, I've attempted to provide a relatively balanced presentation of most of the major issues and I've tried to do this in a student-friendly way. I've occasionally abstracted away from some of the thornier controversies, when I felt they weren't crucial to a student understanding the basics. This may make the professional syntactician feel that I've cut corners or laid out too rosy a picture. I did this on purpose, however, to give students a chance to absorb the fundamentals before challenging the issues. This was a deliberate pedagogical choice. I'm well aware that sometimes I've glossed over controversies, but I think a student has to learn the basics of how the system works before they can seriously critique and evaluate the model. This is a textbook, not a scholarly tome, so its aim is to reach as many students as possible. The style is deliberately low-key and friendly. This doesn't mean I don't want the students to challenge the material I've presented here. Throughout the book, you'll find grey "textboxes" that contain issues for further discussion or interesting tidbits. Many of the problem sets also invite the student to challenge the black and white presentation I've given in the text. I encourage instructors to assign these, and students to do them, as they form an important part of the textbook. Instructors may note that if a favorite topic is not dealt with in the body of the text, a problem set may very well treat the question.

A quick word on the level of this textbook: This book is intended as an introduction to syntactic theory. It takes the student through most of the major issues in Principles and Parameters, from tree drawing to constraints on movement. While this book is written as an introduction, some students have reported it to be challenging. I use this text in my upper-division undergraduate introduction to syntax course with success, but I can

Syntax: A Generative Introduction, Fourth Edition. Andrew Carnie.

certainly see it being used in more advanced classes. I hope instructors will flesh out the book, and walk their students through some of the thornier issues.

This textbook has grown out of my lecture notes for my own classes. Needless to say, the form and shape of these notes have been influenced in terms of choice of material and presentation by the textbooks my own students have used. While the book you are reading is entirely my fault, it does owe a particular intellectual debt to the following three textbooks, which I have used in teaching at various times:

Cowper, Elizabeth (1992) *A Concise Introduction to Syntactic Theory: The Government and Binding Approach. Chicago*: Chicago University Press.

Haegeman, Liliane (1994) *Introduction to Government and Binding Theory (2nd edition).* Oxford: Blackwell.

Radford, Andrew (1988) *Transformational Grammar: A First Course*. Cambridge: Cambridge University Press.

I'd like to thank the authors of these books for breaking ground in presenting a complicated and integrated theory to the beginner. Writing this book has given me new appreciation for the difficulty of this task and their presentation of the material has undoubtedly influenced mine.

Sadly, during the final stages of putting the first edition of this text together, my dissertation director, teacher, mentor, and academic hero, Ken Hale, passed away after a long illness. Ken always pushed the idea that theoretical syntax is best informed by cross-linguistic research, while at the same time the accurate documentation of languages requires a sophisticated understanding of grammatical theory. These were important lessons that I learned from Ken and I hope students will glean the significance of both by reading this text. While I was writing this book (and much other work) Ken gave me many comments and his unfettered support. He was a great man and I will miss him terribly.

It's hard to believe that I began writing the first drafts of this book twenty two years ago, and now we're going into the 4th edition. A whole generation of syntacticians began their careers with this book and I'm utterly humbled that it has been so useful to people. I hope this new edition, and the 2nd edition of the accompanying workbook will continue to guide people into the world of syntax. The 4th edition has some important changes from the 3rd. I brought back affix lowering for those who loved it. That's now at the end of chapter 7. But I've also kept the selection-based analysis of English Auxiliaries in chapter 9. Instructors can safely do one, the other, or both. I've changed the names of some of the functional categories to reflect current practice. There's a new, albeit cursory, chapter on Merge at the end. There's new exercises and problem sets in every chapter both in the main book and in the workbook.

Perhaps the most important change to the book is more subtle and may not be initially apparent. While example sentences are just examples and aren't meant to make profound statement with their content, I've learned through the years that they can impact people nonetheless by perpetuating cultural bias. The subconscious messages example sentences can send cannot help but impact readers. Macaulay and Brice (1997) and Pabst et al (2018) have shown that syntax textbooks and journals often have example

sentences that show bias towards male actors and female patients. In reviewing the example sentences in the third edition, I discovered that not only was there a gender bias in the examples, but there was an unconscious bias towards white anglophone names and a failure to represent ethnic, racial, sexuality, and gender diversity. So I made a conscious effort to address those biases. Honestly, I probably haven't entirely succeeded but I hope the book is now a little more welcoming and will encourage more women, more gender-diverse people, more people of color, more indigenous people and generally broader range of students to seriously consider syntactic theory as their life's work.

I hope that instructors and students will find these revisions helpful. I have attempted where possible to take into account all the many comments and suggestions I received from people using the previous editions, although of course, in order to maintain consistency, I was unable to implement them all.

Acknowledgments:
I'd like to thank the many people who taught me syntax through the years: Barb Brunson, Noam Chomsky, Elizabeth Cowper, Ken Hale, Alec Marantz, Diane Massam, Jim McCloskey, Shigeru Miyagawa, and David Pesetsky. A number of people have read through this book or the previous editions and have given me helpful comments; others have helped on smaller issues but have had no less of an impact on the work and still others have contributed problem sets or editorial advice. This long list includes: Dong-Hwan An, David Adger, William Alexander, Dean Allemang, Gabriel Amores, Diana Archangeli, Ash Asudeh, Dali Balti, Brett Baker, Uldis Balodis, Mark Baltin, Luis Barragan, Andy Barss, Michael Bauer, Dane Bell, Emily Bender, Abbas Benmamoun, Jeff Berry, Tom Bever, Bronwyn Bjorkman, Laura Blumenthal, Claire Bowern, Joan Bresnan, Aaron Broadwell, Dirk Bury, Ivano Caponigo, Roy Chan, Ronald Charles, Danny Chen, Deborah Chen-Pichler, Jaehoon Choi, Barbara Citko, Ian Clayton, Peter Cole, Chris Collins, Jennifer Columbus, Richard Compton, Andrew Comrie, Robert Coren, Dick Demers, Lorenzo Demery, Sheila Dooley, Rebecca Drinkall, Joe Dupris, Yehuda Falk, Muriel Fisher, Megan Figueroa, Sandiway Fong, Leslie Ford, Amy Fountain, Stefan Frisch, Alexandra Galani, Andrew Garrett, Jila Ghomeshi, David Gil, Carrie Gillion, Erin Good-Ament, Anthony Green, Andrea Haber, Paul Hagstrom, Ken Hale, John Halle, Mike Hammond, Daniel Harbour, Jack Hardy, Heidi Harley, Josh Harrison, Rachel Hayes-Harb, David Heap, Bernhard Heigl, One-Soon Her, Caroline Heycock, Nicky Hoover, Stephan Hurtubise, John Ivens, Eloise Jelinek, Ling Jiang, Alana Johns, Mark Johnson, Hyun Kyoung Jung, Arsalan Kahnemuyipour, Dalina Kallulli, Simin Karimi, Dan Karvonen, Andreas Kathol, Chris Kennedy, Greg Key, Amy LaCross, Erwin Lares, Richard Larson, Péter Lazar, Carlos Gelormini Lezama, Jeff Lidz, Anne Lobeck, Leila Lomashivili, Pen Long, Sarah Longstaff, Alicia Lopez, Ahmad Reza Lotfi, Ricardo Mairal, Joan Maling, Jack Martin, Diane Massam, Jeffrey Maxwell, Martha McGinnis-Archibald, Nathan McWhorter, Dave Medeiros, Jason Merchant, Mirjana Miskovic-Lukovic, Tel Monks, Kumiko Murasugi, Alan Munn, MaryLou Myers, Jian Gang Ngui, Chris Nicholas, Janet Nicol, Jon Nissenbaum, Peter Norquest, Diane Ohala, Kazutoshi Ohno, Heidi Orcutt-Gachiri, Hiroyuki Oshita, Panayiotis Pappas, Jaime Parchment, Hyeson

Park, Barbara Partee, Matt Pearson, David Pesetsky, Colin Phillips, Massimo Piatelli-Palmarini, Carl Pollard, Bill Poser, Kristen Pruett, Jeff Punske, Mike Putnam, Sevren Quijada, Eric Randall, Janet Randall, Marlita Reddy-Hjelmfelt, Jodi Reich, Norvin Richards, Frank Richter, Bob Ritchie, Betsy Ritter, Alexander Robertson, Sharon Rose, Ed Rubin, Jeff Runner, Ivan Sag, Nathan Sanders, Yosuke Sato and his students, Theresa Satterfield, Leslie Saxon, Sylvia Schreiner, Kevin Schluter, Carson Schütze, Jim Scobbie, Deborah Shapiro, Leah Shocket, Dan Siddiqi, Echo Ki Sihui, Peter Slomanson, Kyle Smith, Ryan Walter Smith, Norvel Smith, Nick Sobin, Peggy Speas, Megan Stone, Tania Strahan, Joshua Strauss, Dana Sussman,Maggie Tallerman, Takashi Tanaka, Chris Tancredi, Deniz Tat, Brian ten Eyck, Ariel Theisen, Lisa deMena Travis, Alex Trueman, Adam Ussishkin, Huseyin Uysal, Sakari Vaelma , Robert Van Valin, Martin Walkow, Enwei Wang, Shan Wang, Natasha Warner, Andy Wedel, Jennifer Wees, Jerry Weltman, Mary Ann Willie, Marian Wiseley, Dainon Woudstra, Susi Wurmbrand, Alper Yavuz, Kimberley Young, Kim Youngroung, J.R. Yu, James Yuen, my Facebook friends who I regularly victimized as testers for the problem sets in the book, and several anonymous Blackwell and Wiley reviewers. I'm absolutely convinced I've left someone off this large list. If it is you many apologies – I really did appreciate the help you gave me. The students in my Introduction to Syntax classes in Michigan in 1997, and in Arizona in 1998–2019, have used all or parts of this textbook. Glynis Baguley, Ada Brunstein, Sarah Coleman, Danielle Descoteaux, Lisa Eaton, Simon Eckley, Rachel Greenberg, Charlotte Frost, Graham Frankland, Tami Kaplan, Becky Kennison, Julia Kirk, Meryl Le Roux, Hannah Lee, Tanya McMullin, Leah Morin, Allison Medoff, Anna Oxbury, Rhonda Pearce, Clelia Petracca, Iain Potter, Venkatnadhan Rajagopalan, Beth Remmes, Jennifer Seward, and Steve Smith of Wiley-Blackwell and their subcontractors all deserve many thanks for help getting this and the previous three editions to press. My family (my mother Jean, my late father Bob, Morag, Fiona, my lost love Pangur, and my new fur kids Crònan, Nechtan and Aoife) were all incredible in their support and love. Go raibh maith agaibh agus tapadh leibh!

The artwork in chapters 3 and 6 was created by Dane Bell for this book and is used with permission.

Part 1

Preliminaries

chapter 1

Generative Grammar

Learning Objectives

After reading chapter 1, you should walk away having mastered the following ideas and skills:

1. Explain why language is a psychological property of humans.
2. Distinguish between prescriptive and descriptive rules.
3. Explain the scientific method as it applies to syntax.
4. Explain the differences between the kinds of data gathering, including corpora and linguistic judgments.
5. Explain the difference between competence and performance.
6. Explain the difference between i-language and e-language
7. Provide at least three arguments for Universal Grammar.
8. Explain the logical problem of language acquisition.
9. Distinguish between learning and acquisition.
10. Distinguish among observational, descriptive, and explanatory adequacy.

0. Preliminaries

Although we use it every day, and although we all have strong opinions about its proper form and appropriate use, we rarely stop to think about the wonder of language. So-called language "experts" tell us about the misuse of *hopefully* or lecture us about the origins of the word *boondoggle*, but surprisingly, they never get at the true wonder of language: how it actually works as a complex machine. Think about it for a minute. You are reading this and understanding it, but you have no conscious knowledge of how you

Syntax: A Generative Introduction, Fourth Edition. Andrew Carnie.

are doing it. The study of this mystery is the science of ***linguistics***. This book is about one aspect of how language works: how sentences are structured, or the study of ***syntax*** and the people who study syntax are called ***syntacticians***.

There are many perspectives on studying linguistics. One could study language looking at languages across time, or one could study how language is used as a social too. But syntacticians typically take a different view. They look at language as a psychological or cognitive property of humans. That is, my mind contains certain principles that allow me to sit here and produce this set of letters, words and sentences, and you use similar principles that allow you to translate these squiggles back into coherent ideas and thoughts. At least I hope you can translate them back into coherent ideas!

There are several subsystems at work in when we use language. If you were listening to me speak, I would be producing sound waves with my vocal cords and articulating particular speech sounds with my tongue, lips, and vocal cords. On the other end of things, you'd be hearing those sound waves and translating them into speech sounds using your auditory apparatus. The study of the acoustics and articulation of speech is called ***phonetics***. Once you've translated the waves of sound into mental representations of speech sounds, you analyze them into syllables and pattern them appropriately. For example, speakers of English know that the made-up word *bluve* is a possible word of English, but the word *bnuck* is not. This is part of the science called ***phonology***. Then you take these groups of sounds and organize them into meaningful units (called morphemes) and words. For example, the word *dancer* is made up of two meaningful bits: *dance* and the suffix *-er*. The study of this level of language is called ***morphology***. Next you organize the words into phrases and sentences. One usage of the term ***syntax*** is the cover term for studies at this level of language. Finally, you take the sentences and phrases you hear and translate them into thoughts and ideas. This last step is what we refer to as the ***semantic*** level of language.

Syntax as a discipline studies the part of language knowledge that lies between words and the meaning of utterances: sentences. It is the level that mediates between sounds that someone produces (organized into words) and what they intend to say.

Perhaps one of the truly amazing aspects of the study of language is not the origins of the word *demerit,* or how to properly punctuate a quote inside parentheses, or how kids have, like, destroyed the English language, eh? Instead it's the question of how we subconsciously get from sounds and words to the meaning of sentences. This is the study of syntax.

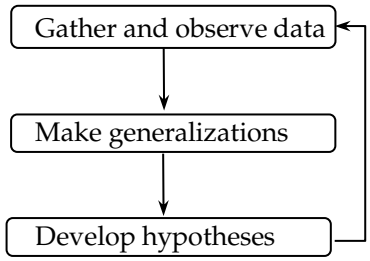
1. Syntax as Science – the Scientific Method

For many people, the study of language properly belongs in the humanities. That is, the study of language is all about the beauty of its usage in fine (and not so fine) literature and its impact on human culture. However, there is no particular reason, other than tradition, that the study of language should be confined to a humanistic approach. It is also possible to approach the study of language from a scientific perspective; this is the domain of linguistics. People who study literature often accuse linguists of abstracting

away from the richness of good prose and obscuring the beauty of language. Nothing could be further from the truth. Most linguists, including the present author, enjoy nothing more than reading a finely crafted piece of fiction, and many linguists often study, as a sideline, the more humanistic aspects of language. This doesn't mean, however, that one can't appreciate and study the formal properties (or rules) of language and do it from a scientific perspective. The two approaches to language study are both valid; they complement each other; and neither takes away from the other[1].

Science is perhaps one of the most poorly defined words of the English language. We regularly talk of scientists as people who study bacteria, particle physics, and the formation of chemical compounds, but ask your average Joe or Jill on the street what science means, and you'll be hard pressed to get a decent definition. But among scientists themselves, *science* typically refers to a particular methodology for study: the deductive scientific method. The scientific method dates back to the ancient Greeks, such as Aristotle, Euclid, and Archimedes. The method involves observing some data, making some generalizations about patterns in the data, developing hypotheses that account for these generalizations, and testing the hypotheses against more data. Finally, the hypotheses are revised to account for any new data and then tested again. A flow chart showing the method is given in (1):

1)



In syntax, we apply this methodology to sentence structure. Syntacticians start[2] by observing data about the language they are studying, then they make generalizations

[1] The notion that science should be the primary means of investigating linguistics has come under significant criticism by some members of indigenous communities (see for example, the opinions expressed by community members as reported in Czaykowska-Higgins 2009, the articles in Bischoff and Jany et al. 2019 and Rosborough, et al. 2017). The central idea is that by exclusively using a deductive scientific methodology in investigating their languages we are prioritizing a western European set of values on their traditions and cultures. When we as linguists work *with* native communities on their languages it is important that we acknowledge this perspective and think carefully about ways that our research can complement, connect to and support indigenous ways of knowing. We must also find ways in which we can be partners with these communities to help further their local agendas rather just observing their communities as objects of study and imposing our ideas upon them.

[2] This is a bit of an oversimplification. We really have a "chicken and the egg" problem here. You can't know what data to study unless you have a hypothesis about what is important, and you can't have a hypothesis unless you have some basic understanding of the data. Fortunately, as

about patterns in the data (e.g., in simple English declarative sentences, the subject precedes the verb). They then generate a hypothesis about these patterns and test the hypothesis against more syntactic data, and if necessary, go back and re-evaluate their hypotheses.

Hypotheses are only useful to the extent that they make ***predictions***. A hypothesis that makes no predictions (or worse yet, predicts everything) is useless from a scientific perspective. In particular, the hypothesis must be ***falsifiable***. That is, we must in principle be able to look for some data, which, if true, show that the hypothesis is wrong. This means that we are often looking for the cases where our hypotheses predict that a sentence will be grammatical (and it is not), or the cases where they predict that the sentence will be ungrammatical (contra to fact).

In syntax, hypotheses are called ***rules***, and the group of hypotheses that describe a language's syntax is called a ***grammar***. The term *grammar* can strike terror into the hearts of people. But you should note that there are two ways to go about writing grammatical rules. One is to tell people how they *should* speak (this is of course the domain of English teachers and copy-editors); we call these kinds of rules ***prescriptive rules*** (as they prescribe how people should speak according to some standard). Some examples of prescriptive rules include "never end a sentence with a preposition", "use *whom* not *who*" and "don't split infinitives". These rules tell us how we are supposed to use our language. The other approach is to write rules that describe how people *actually* speak, whether or not they are speaking "correctly". These are called ***descriptive rules***. Consider for a moment the approach we're taking in this book. Which of the two types (descriptive or prescriptive) is more scientific? Which kind of rule is more likely to give us insight into how the mind uses language? We are going to focus on descriptive rules. This doesn't mean that prescriptive rules aren't important (in fact, in the problem sets section of this chapter you are asked to critically examine the question of descriptive vs. prescriptive rules), but for our purposes descriptive rules are more important.

You now have enough information to answer General Problem Sets GPS1 & 2, as well as Challenge Problem Set CPS1 at the end of this chapter. For practice try Workbook Exercise WBE1 in chapter 1 of The Syntax Workbook, 2nd Edition*, an optional companion book to this text.*

Do Abstract Rules Really Exist?

As discussed in detail later in this chapter, the approach to grammar we are using here is supposed to be part cognitive psychology, so it's reasonable to ask whether formal rules *really* exist in the brain/minds of speakers. After all, a brain is a mass of neurons firing away, so how can abstract rules exist up there? Remember, however, that we are attempting to *model* language; we aren't trying to describe language exactly. This question confuses two disciplines: psychology and neurology. Psychology is concerned with the mind, which represents the output and the abstract organization of the brain.

---

working syntacticians this philosophical conundrum is often irrelevant, as we can just jump feet-first into both the hypothesis-forming and the data-analysis at the same time.

Neurology is concerned with the actual firing of the neurons and the physiology of the brain. Our approach doesn't try to be a theory of neurology. Instead it is a model of the psychology of language. Obviously, the rules per se don't exist in our brains, but they do model the external behavior of the mind. For more discussion of this issue, look at the readings in the further reading section of this chapter.

3.1 An Example of the Scientific Method as Applied to Syntax

Let's turn now to a real-world application of the scientific method to some language data. The following data concern the form of a specific kind of noun, called an ***anaphor*** (plural: ***anaphors***; the phenomenon is called ***anaphora***). These include the nouns that end with *-self* (e.g., *himself, herself, itself*). In chapter 5, we look at the distribution of anaphors in detail; here we'll only consider one superficial aspect of them. In the following sentences, as is standard in the syntactic literature, a sentence that isn't well-formed is marked with an ***asterisk*** (*) before it. For these sentences assume that *Bill* is male and *Sally* is female.

2) a) Bill kissed himself.
   b) *Bill kissed herself.
   c) Sally kissed herself.
   d) *Sally kissed himself.
   e) *Kiss himself.

Under the assumption that Bill is a cisgender male and Sally is a cisgender female, the ill-formed sentences in (2b and d) just look silly. It is obvious that Bill can't kiss herself, because Bill is male. There is a clear generalization about the distribution of anaphors here. In particular, the generalization we can draw about the sentences in (2) is that an anaphor must agree in ***grammatical gender*** with the noun it refers to (its ***antecedent***). So. in (2a & b) we see that the anaphor must agree in gender with *Bill*, its antecedent. The anaphor must take the masculine form *himself*. The situation in (2c & d) is the same; the anaphor must take the form *herself* so that it agrees in gender with the feminine *Sally*. Note further that a sentence like (2e) shows us that anaphors must have an antecedent. An anaphor without an antecedent is unacceptable. A plausible hypothesis (or rule) given the data in (2), then, is stated in (3):

3) An anaphor must (i) have an antecedent and (ii) agree in grammatical gender (masculine, feminine, or neuter) with that antecedent.

The next step in the scientific method is to test this hypothesis against more data. Consider the additional data in (4):

4) a) The robot kissed itself.
   b) She knocked herself on the head with a zucchini.
   c) *She knocked himself on the head with a zucchini.
   d) The snake flattened itself against the rock.
   e) ?The snake flattened himself/herself against the rock.
   f) The Joneses think themselves the best family on the block.
   g) *The Joneses think himself the wealthiest guy on the block.

h) Gary and Kevin ran themselves into exhaustion.
i) *Gary and Kevin ran himself into exhaustion.

Grammatical Gender vs. Sex vs. Personal Gender

Gender can be a politically charged and deeply personal issue for many people. In this chapter, I am talking about primarily about ***grammatical gender***. Grammatical gender is often confused with sex assigned at birth and with the gender identity/expression of the individual. This is because people often use grammatical gender to signal their sex or gender identity to others. But in the context that I'm using it here, it's a purely formal feature of words. In many languages grammatical gender, also called ***noun class***, has nothing to with actual sex or gender identity. For example, in Navajo grammatical gender is determined by shape, consistency and animacy and is quite distinct from their cultural understanding of gender identity. In other languages, grammatical gender does not need to correspond to gender expression – it can even be the opposite. In Modern Irish, for example, the word *cailín* 'girl' is masculine and the word *stail* 'stallion' is feminine.

Despite the objections of prescriptive language gurus, English has long used the pronoun *they* to refer to humans in a gender-neutral way. Recently this usage has been extended more regularly to people whose gender identity is non-binary. This new usage has some really interesting effects on the phenomenon of anaphora – in particular a new anaphor, *themself*, has been added to the grammatical system of many people, particularly younger speakers. General Problem Set GPS3 gives you a chance to explore the interplay of grammatical gender and personal gender with English anaphora and verb agreement.

Sentences (4a, b, & c) are all consistent with our hypothesis that anaphors must agree in gender with their antecedents, which at least confirms that the hypothesis is on the right track. What about the data in (4d & e)? It appears as if any gender is compatible with the antecedent *the snake*. This appears, on the surface, to be a contradiction to our hypothesis. Think about these examples a little more closely, however. Whether sentence (4e) is well-formed or not depends upon your assumptions about the gender of the snake. If you assume (or know) the snake to be male, then *The snake flattened himself against the rock* is perfectly well-formed. But under the same assumption, the sentence *The snake flattened herself against the rock* seems very odd indeed, although it is fine if you assume the snake is female. So, it appears as if this example also meets the generalization in (3); the vagueness about its well-formedness has to do with the fact that we are rarely sure what gender a snake is and not about the actual structure of the sentence.

Now, look at the sentences in (4f–i) above; note that the ill-formedness of (g) and (i) is not predicted by our generalization. In fact, our generalization predicts that sentence (4i) should be perfectly grammatical, since *himself* agrees in gender (masculine) with its antecedents *Gary* and *Kevin*. Yet there is clearly something wrong with this sentence. The hypothesis needs revision. It appears as if the anaphor must agree in gender and ***number*** with the antecedent. Number refers to the quantity of individuals involved in the sentence; English primarily distinguishes singular number from plural number. (5) reflects our revised hypothesis.

5) An anaphor must agree in gender and number with its antecedent.

If there is more than one person or object mentioned in the antecedent, then the anaphor must be plural (i.e., *themselves*).

Testing this against more data, we can see that this partially makes the correct predictions (6a), but it doesn't properly predict the acceptability of sentences (6b–e):

6) a) People from Tucson think very highly of themselves.
   b) *I gave yourself the bucket of ice cream.
   c) I gave myself the bucket of ice cream.
   d) *She kissed myself.
   e) She kissed herself.

Even more revision to our hypothesis is in order. The phenomenon seen in (6b–e) revolves around a grammatical distinction called ***person***. Person refers to the perspective of the speaker with respect to the other participants in the speech act. ***First person*** refers to the speaker. ***Second person*** refers to the addressee. ***Third person*** refers to people being discussed that aren't participating in the conversation. Here are the English pronouns associated with each person: (***Nominative*** refers to the ***case*** form the pronouns take when in subject position like *I* in "*I* love peanut butter"; ***accusative*** refers to the form they take when in object positions like *me* in "John loves *me*". We will look at case in much more detail in chapter 11, so don't worry if you don't understand it right now.)

7)

	Nominative		Accusative		Anaphoric	
	Singular	Plural	Singular	Plural	Singular	Plural
1	I	we	me	us	myself	ourselves
2	you	you	you	you	yourself	yourselves
3 masc	he	they	him	them	himself	themselves
3 fem	she		her		herself	
3 neut	it		it		itself	

As you can see from this chart, the form of the anaphor seems also to agree in person with its antecedent. So once again we revise our hypothesis (rule):

8) An anaphor must agree in person, gender and number with its antecedent.

With this hypothesis, we have a straightforward statement of the distribution of this noun type, derived using the scientific method. In the problem sets below, and in chapter 5, you'll have an opportunity to revise the rule in (8) with even more data.

You now have enough information to try GPS3, WBE2, and CPS2 & CPS3

3.2 Sources of Data

If we are going to apply the scientific method to syntax, it is important to consider the sources of our data. One obvious source is in collections of either spoken or written texts. Such data are called ***corpora*** (singular: ***corpus***). There are many corpora available,

including some searchable through the internet. For languages without a literary tradition or languages spoken by a small group of people, it is often necessary for the linguist to go and gather data and compile a corpus in the field. In the early part of the last century, this was the primary occupation of linguists, and it is proudly carried on today by many researchers.

The linguist Heidi Harley reports in her blog[3] on an example of using search engines to do linguistic analysis on the huge corpus known as the web. Harley notes that to her ear, the expression *half full of something* sounds natural, but *half empty of something* does not. She does a comparison of *half empty* vs. *half full* and of *half empty of* vs. *half full of*. She finds that the ratio of *half full* to *half empty* without the *of* is roughly 1:1. The ratio of *half full of* to *half empty of* is approximately 149:1. This is a surprising difference. Harley was able to use the web to show that a fairly subtle difference in acceptability is reflected in the frequency with which the expressions are used.

But corpus searches aren't always adequate for finding out the information syntacticians need. For the most part corpora only contain grammatical sentences. Sometimes the most illuminating information is our knowledge that a certain sentence is ungrammatical (i.e., not a sentence of normal English), or that two similar sentences have very different meanings. Consider the pair of sentences in (9) as a starting point.

9) a) Marian blew the building up.
   b) Marian blew up the building.

Most native speakers of English will accept both of these sentences as acceptable sentences, with a preference for (9b). They also know that while the first sentence (9a) is unambiguous, the second one has two meanings (He destroyed the building using explosives vs. he blew really hard with his lungs up the stairwell). The second of these meanings is a bit silly, but it's a legitimate interpretation of the sentence.

Now contrast the sentences in (9) with the similar pair in (10). In these forms I've replaced "the building" with the pronoun "it":

10) a) Marian blew it up.
   b) Marian blew up it.

Here we find a different pattern of interpretation. (10a) is unambiguous just the way (9a) is, it refers to an act of explosion and cannot have an interpretation where Marian was blowing hard with her lungs up something. Sentence (10b), however, is a surprise. Unlike (9b), (10b) cannot have anything to do with explosives. It can only have the interpretation where Marian is blowing air up whatever "it" is. Recall that with (9) this "puff of air reading" was the silly or strange one. With a pronoun, however, it's the only available interpretation. This difference in interpretation would never be captured in a corpus, because the specific meanings of expressions and ambiguities are not indicated anywhere in the data source.

While corpora are unquestionably invaluable sources of data, they are only a partial representation of what goes on in the mind. More particularly, corpora often contain

[3] http://heideas.blogspot.com/2005/10/scalar-adjectives-with-arguments.html.

instances of only acceptable (or, more precisely, well-formed) sentences (sentences that sound "OK" to a native speaker). For example, the online *New York Times* contains very few ungrammatical sentences. Even corpora of naturalistic speech complete with the errors every speaker makes don't necessarily contain the data we need to test the falsifiable predictions of our hypotheses. So, corpora are just not enough: there is no way of knowing whether a corpus has *all* possible forms of grammatical sentences. In fact, as we will see in the next few chapters, due to the productive nature of language, a corpus could *never* contain all the grammatical forms of a language, nor could it even contain a representative sample. It also doesn't tell us about what sentences are ambiguous or what sentences are ungrammatical or strange. Those are really important sources of evidence for doing syntax. To really get at what we know about our languages we have to know what sentences are *not* well-formed. That is, in order to know the range of acceptable sentences of English, Italian or Igbo, we *first* have to know what are *not* acceptable sentences in English, Italian or Igbo. This kind of negative information is very rarely available in corpora, which mostly provide grammatical, or well-formed, sentences.

Consider the following sentence:

11) *Who do you wonder what bought?

For most speakers of English, this sentence borders on word salad – it is not a good sentence of English. How do you know that? Were you ever taught in school that you can't say sentences like (11)? Has anyone ever uttered this sentence in your presence before? I seriously doubt it. The fact that a sentence like (11) sounds strange, but similar sentences like (12a and b) *do* sound OK is not reflected anywhere in a corpus:

12) a) Who do you think bought the bread machine?
b) I wonder what Fiona bought.

Instead we have to rely on our knowledge of our native language (or on the knowledge of a native speaker consultant for languages that we don't speak natively). Notice that this is *not* conscious knowledge. I doubt there are many native speakers of English that could tell you why sentence (11) is terrible, but most can tell you that it is. This is subconscious knowledge. The trick is to get at and describe this subconscious knowledge.

The psychological experiment used to get this subconscious kind of knowledge is called the ***acceptability judgment task***. The judgment task involves asking a native speaker to read a sentence, and judge whether it is well-formed (i.e., grammatical), marginally well-formed, or ill-formed (ungrammatical).

There are actually several different kinds of acceptability judgments. Both of the following sentences are ill-formed, but for different reasons:

13) a) #The toothbrush is pregnant.
b) *Toothbrush the is blue.

Sentence (13a) sounds bizarre (cf. *the toothbrush is blue*) because we know that toothbrushes (except in the world of fantasy/science fiction or poetry or a dream) cannot be pregnant. The meaning of the sentence is strange, but the form of the sentence is okay. We call this ***semantic ill-formedness*** and mark the sentence with a #. By contrast, we can

glean the meaning of sentence (13b); it seems semantically reasonable (toothbrushes can be blue), but it is ill-formed from a structural point of view. That is, the determiner *the* is in the wrong place in the sentence. This is a ***syntactically ill-formed*** sentence, which is marked with an *. A native speaker of English will judge both these sentences as ill-formed, but for very different reasons. In this text, we will be concerned primarily with syntactic well-formedness, but both kinds of judgment can help guide our analyses.

You now have enough information to do WBE3 & 4, GPS3 & 4, and CPS4–6.

Judgments as Science?

Many linguists refer to the acceptability judgment task as "drawing upon our native speaker intuitions". The word "intuition" here is slightly misleading. The last thing that pops into our heads when we hear the term "intuition" is science. Generative grammar has been severely criticized by many for relying on "unscientific" intuitions. But this is based primarily on a misunderstanding of the term. To the layperson, the term "intuition" brings to mind guesses and luck. This usage of the term is certainly standard. When a generative grammarian refers to "intuition", however, she is using the term to mean "tapping into our subconscious knowledge". The term "intuition" may have been badly chosen, but in this circumstance, it refers to a real psychological effect. Intuition (as an acceptability judgment) has an entirely scientific basis. It is replicable under strictly controlled experimental conditions (these conditions are rarely applied, but the validity of the task is well established). Other disciplines also use intuitions or judgment tasks. For example, within the study of vision, it has been determined that people can accurately judge differences in light intensity, drawing upon their subconscious knowledge (Bard et al. 1996). To avoid the negative associations with the term intuition, we will use the less loaded term ***judgment*** instead.

2. Syntax as a Cognitive Science

Cognitive science is a cover term for a group of disciplines that all have the same goal: describing and explaining human beings' ability to think (or more particularly, to think about abstract notions like subatomic particles, the possibility of life on other planets or even how many angels can fit on the head of a pin, etc.). One thing that distinguishes us from other animals, even relatively smart ones like chimps and elephants, is our ability to use productive, combinatory syntax. Language plays an important role in how we think about abstract notions, or, at the very least, it appears to be structured in such a way that it allows us to express abstract notions.[4] The discipline of linguistics is thus one of the important subdisciplines of cognitive science.[5] Sentences are how we get at expressing abstract thought processes, so the study of syntax is an important foundation stone for understanding how we communicate and interact with each other as humans.

[4] Whether language constrains what abstract things we can think about (this idea is called the Sapir–Whorf hypothesis) is a matter of great debate and one that lies outside the domain of syntax per se.
[5] Along with psychology, neuroscience, communication, philosophy, and computer science.

3. MODELS OF SYNTAX

One dominant theory of syntax that fits into the cognitive science mold is due to Noam Chomsky and his colleagues, starting in the mid 1950s and continuing to this day. This theory, which has had many different names through its development (Transformational Grammar (TG), Transformational Generative Grammar, Standard Theory, Extended Standard Theory, Government and Binding Theory (GB), Principles and Parameters approach (P&P) and Minimalism (MP)), is often given the blanket name ***generative grammar***. A number of alternate approaches to syntax have also branched off of this research program. These include Lexical-Functional Grammar (LFG) and Head-Driven Phrase Structure Grammar (HPSG). These approaches are also considered part of generative grammar; but we won't cover them extensively in this book. But I have included two additional chapters on these theories in the web resources for this book[6]. The particular version of generative grammar that we will mostly look at here is roughly the ***Principles and Parameters*** approach, although we will occasionally stray from this into the more recent version called ***Minimalism***.

The underlying thesis of generative grammar is that sentences are generated by a subconscious set of procedures (like computer programs). These procedures are part of our minds (or of our cognitive abilities if you prefer). The goal of syntactic theory is to model these procedures. In other words, we are trying to figure out what we subconsciously know about the syntax of our language.

4. COMPETENCE VS. PERFORMANCE

Consider sentences such as (14). Native speakers will have to read this sentence a couple of times to figure out what it means.

14) # Cotton shirts are made from comes from India.

This kind of sentence (called a ***garden path sentence***) is very hard to understand and process. In this example, the problem is that the intended reading has a noun, *cotton*, that is modified by a reduced relative clause: *(that) shirts are made from.* The linear sequence of *cotton* followed by *shirt* is ambiguous with the noun phrase *cotton shirts*. Note that this kind of relative structure is okay in other contexts; compare: *That material is the cotton shirts are made from.* Sentences like (14) get much easier to understand with really clear pauses (where … is meant to indicate a pause): *Cotton … shirts are made from … comes from India*. Or by insertion of a *that* which breaks up the potentially ambiguous *cotton shirts* sequence: *The cotton that shirts are made from comes from India*. What is critical about these garden path sentences is that, once one figures out what the intended meaning is, native speakers can identify them as acceptable sentences or at the very least as sentences that have structures that would otherwise be acceptable in them. The problem for us as

[6] http://www.wiley.com/go/carnie