

# THE PSYCHOLOGICAL IMPORT OF SYNTACTIC THEORY

by

DAVID PEREPLYOTCHIK

A dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2012



This manuscript has been read and accepted for the Graduate Faculty in Philosophy in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

John D. Greenwood

\_\_\_\_\_  
Date

\_\_\_\_\_  
Chair of the Examining Committee

Iakovos Vasiliou

\_\_\_\_\_  
Date

\_\_\_\_\_  
Executive Officer

Michael Devitt \_\_\_\_\_

Stephen Neale \_\_\_\_\_

Janet Dean Fodor \_\_\_\_\_

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK  
Abstract

THE PSYCHOLOGICAL IMPORT OF SYNTACTIC THEORY

by

David Pereplyotchik

Adviser: Michael Devitt

My primary goal is to assess whether, and in what sense, the rules or principles of grammar are psychologically real. I begin by casting doubt on a received view in generative linguistics, according to which a true theory of the syntax of natural language would, *ipso facto*, be a theory of a psychological state or mechanism. I argue that a nominalist construal of linguistic theory is a viable alternative to the dominant Chomskyan view that linguistics is a branch of psychology. If this is correct, it follows that there are substantive issues about whether the theoretical constructs of formal linguistics play any role in psychological processes, and, if so, what role they play. To address these issues, I examine a range of behavioral and neurocognitive data from psycholinguistics. The data strongly suggest that the human language processing mechanism constructs mental representations of the syntactic properties of incoming linguistic stimuli. I then survey a number of computational models of human language comprehension. While all such models account for an impressive range of data, they make use of the rules or principles of a grammar in one of two very different ways—either by explicitly *representing* them in a data structure or by *embodying* them in the form of hardwired procedural dispositions. It is reasonable to suppose, then, that grammars are psychologically real in one of these two ways. But which? To answer this question, I go on to sketch a theoretical framework for thinking about represented and embodied rules, distinguishing embodiment from mere conformity to a



rule. I then argue that embodied rules are typically implemented by simpler mechanisms, and that embodiment is, therefore, the more parsimonious hypothesis (*ceteris paribus*). Furthermore, I argue that we have no principled grounds, at present, for asserting that grammars are represented, rather than embodied, in the human brain. From this, I conclude that a common claim in generative linguistics, i.e., that grammars are *represented* in the minds of competent language users, must be seen as either as a conflation of the notions of embodiment and representation, or as an attractive but as-yet-ungrounded hypothesis.

## Acknowledgements

This dissertation is the culmination of a long journey. Over the course of my 12-year engagement with philosophy, I've been influenced in countless ways by my conversations and relationships with the people whose names appear below. Looking at the sheer number of wonderful people with whom I have shared not only ideas, but joyous times, laughter, friendship, and love, I can't help but to feel blessed. I am grateful beyond words to all of you, for making these past years so amazing.

My abiding interest in philosophy took hold at New York University, where Stephen Schiffer, Hartry Field, Roger White, Ken Ehrenberg, Scott Walden, Declan Smithies, and Rohit Parikh introduced me to the rigors of the analytic approach and instilled in me a deep and lasting fascination with language and the mind. It was in Stephen Schiffer's course that I had my first exposure to the works of Frege, Russell, Quine, Grice, and Kripke. I am grateful for his support, encouragement, and advice throughout the years. I should also like to thank Tycho Sleater in the Physics Department, Stanley Boorman in the Music Department, Stephen F. Cohen in the History Department, and Michal Starke in the Linguistics Department for treating me to a wealth of material in fields outside of philosophy.

By the end of my time at NYU, I had developed close friendships with several young philosophers, who were all either in graduate school already or working hard to get admitted: Ray Buchanan, Declan Smithies, Jonathan Simon, Joomi Kim, Benjamin Boudreaux, and Mike Bruno. It was they who inspired me to pursue philosophy at the graduate level. I thank them for

giving my life direction at a time when I badly needed it, and for setting me on what has turned out to be an incredibly rewarding path. I wish them all massive success in their lives and their careers.

Having arrived at the CUNY Graduate Center, I found guidance, support, and high-quality instruction from a great many faculty members, in both the Philosophy and the Linguistics programs. I wish to name a few here and thank them for their time, their help, and their wisdom.

I thank Steven Ross for his guidance through a rough patch in my graduate career, and for inspiring in me a shift in perspective, from an inchoate and ultimately indefensible relativist position in metaethics to an unabashed realism. That shift has had a lasting effect not only on my philosophical views, but also on my personal choices. It was in part on this account that I drastically reduced my meat consumption and, more recently, began donating regularly to charities that combat the ills of global poverty.

I thank Michael Levin for countless engaging, exciting, and helpful conversations over the years, as well as some of the best courses that I took at the Graduate Center. Professor Levin's dedication to his students is extraordinary. I have benefitted enormously from his insightful and impassioned lectures on virtually every area of philosophy and his extensive comments on a significant part of my total written output during my graduate career. His defenses of compatibilism, nominalism, and anti-skeptical deflationary realism have had a powerful influence on my own thinking about free will, ontology, and epistemology. I hope to one day

achieve the disciplined and committed work ethic that he both cultivates in his students and exemplifies in his own research and teaching.

Several faculty members in the Linguistics program have been quite generous with their time and very kind in accepting a philosopher into their midst. For Robert Fiengo, the latter came naturally, for he is as knowledgeable and interested in the philosophy of language as anyone I've encountered in the course of my graduate studies. I will always be grateful for his patience during our office-hour meetings, where he challenged a great many of my presuppositions about pragmatics (no pun intended) and about the foundational goals of linguistic theorizing.

Christina Tortora and Marcel Den Dikken have done a great deal to both stimulate and to foster my interest in formal syntax. Their courses supplied me with an immensely valuable inside look at the history, the methodology, the practice, and the impressive achievements of generative linguistics. Both have been very receptive to my questions and concerns about both the conceptual foundations and the technical details of their subject matter. For this, I am very grateful. Dianne Bradley has done the same with respect to psycholinguistics. Auditing her course on the topic was an eye-opening experience for me, and plunged me headlong into an intensive study of the exciting research on the frontiers of the field.

It's difficult to overestimate the significance of the role that David Rosenthal has played in my life over the past 8 years. David is not only a professor and a mentor; he is also a close friend. (The semicolon in the previous sentence is no accident. ;) David took me under his wing early on in my graduate career. Working with him on his book, *Consciousness and Mind*, was one of

the most intense and rewarding experiences that I've ever had. I didn't know it then, but, ultimately, David would come to have an immense influence on my thinking about virtually every topic in philosophy. His courses in the philosophy of mind and his weekly cognitive science symposia provided me with a fantastically rich perspective on both traditional philosophy and its interface with cognitive science. In his courses on W. V. Quine and Wilfrid Sellars, of which I had the benefit of attending no less than three, David enlarged upon topics in the philosophy of language and metaphysics, tracing deep connections to logic, linguistics, psychology, philosophy of science, and related fields. His occasional forays into the fine points of scholarship on Plato, Aristotle, and Descartes (among others) occasioned in me an interest in the history of philosophy, which I now pursue with delight, explicitly adopting the methodology that David both advocates and exemplifies in his own writings.

David's class sessions were, to my mind, paradigms of how philosophy ought to be done. While charitably presenting his opponents' views, he offered original and compelling arguments for his own, and engaged with objections carefully and fairly. When the sessions ended, David would regularly devote an incredible amount of time and attention to the students that stayed behind. His after-class office hour meetings were valuable not only for their rich philosophical content, but also for the implicit lessons in pedagogical strategy. David's way of relating to his students is at once too subtle and intuitive to be calculated, but too systematic and effective not to be. His ability to target a student's specific intellectual concern is uncanny. And so it was with me. His arguments and his generosity with his time had a profound effect. It's probably no exaggeration to say that I've completely drunk the Kool-Aid. On the pivotal issues, I discern no substantive difference between his views and my own. I've thought a lot, over the years, about what I might

have wound up thinking about consciousness, intentionality, ontology, and the like, had I not studied with him. In a spirit that I suspect he will appreciate, I've come to the following conclusion: If I'd thought anything other than what I think, I would have probably been wrong, because I know the arguments for what I think, and they're really pretty good!

In writing this dissertation, I was fortunate to have the help and support of an extraordinary committee. I thank my lucky stars that I had such an amazing group of kind, supportive, and generous people guiding me along this path.

John G. Greenwood and Gary Ostertag were kind enough to sign on to the committee on very short notice. I am deeply grateful to both of them for being so generous with their time in the weeks prior to the defense. Our meetings before and after the defense were very valuable to me.

I owe special thanks to John for encouraging me to come CUNY in the first place. But his involvement with my graduate career did not stop there. I am deeply grateful to him for his unfailing support over the years, as well as his help and advice in the connection with the job market in recent months.

I have known Gary since my undergraduate days. I recall auditing his philosophy of language course at NYU, learning for the first time about Austin and Searle. It's quite fitting that he should have played a role in the culmination of my graduate work in the philosophy of language. I thank him for his warm attitude toward me, before, during, and after the defense.

Stephen Neale's work has been an inspiration to me ever since I was first introduced to it as an undergraduate in Stephen Schiffer's course. His thorough and compelling explanations and extensions of H. P. Grice's views struck me as a paradigm of clear writing and careful thought. Naturally, I was thrilled when he joined the faculty at CUNY. I eagerly sat in on his courses and was pleasantly surprised to find him approachable and down-to-earth despite his well-deserved fame. His agreeing to be on my prospectus and dissertation committees meant—and continues to mean—a great deal to me. I'm grateful for having gotten the chance to bounce my ideas off him and for his help, guidance, and support, both with the dissertation and the job market. Stephen has been incredibly kind and encouraging to me, despite the fact that my views tend to run counter to his on a great many issues in metaphysics, philosophy of language, and philosophy of mind. I have an optimistic hope that, one day, I'll be able to see my way clearly enough through the issues that the differences will reveal themselves to be nothing more than our grasping different parts of the proverbial elephant. In the meantime, the depth and rigor of Stephen's work will serve as both a lasting source of inspiration and a sobering reminder that I may well be mistaken in some of my core commitments.

To me, as to countless people, Janet Dean Fodor is nothing short of a hero. I respect and admire her a great deal, both as a teacher and as a researcher. Her pathbreaking work in what appears to be literally *every* area of psycholinguistics is of the very highest quality. Her ideas are strikingly original, and always supported by cleverly designed experiment and powerful argumentation. Small wonder that they've held up so well, despite the major shifts in both linguistic theory and psycholinguistic practice. I have studied Janet's work not only with an abiding interest but also with genuine veneration. That she would take the time to guide me through the psycholinguistic

terrain is as amazing to me now as it was when she first signed on to be on my committee. I recall walking into her office for the first time, trying my best to keep calm and not let on that I was utterly star-struck. As we spoke, the sheer confidence that she displayed in her every mannerism was palpable. I felt completely out of my depth. Coming into a new field, with an agenda firmly rooted in another, is a daunting enterprise. But Janet not only exudes confidence; she *inspires* it. Over the years, on account of her patient guidance, careful reading of my work, and countless helpful pieces of advice, I developed a sense of both the history and the cutting-edge developments in psycholinguistics. And, like all of Janet's students, I also had the benefit of her coaching in matters of style, presentation, and professionalism. For, if there's anything that Janet has more of than confidence, it's style. The grace with which she carries herself, both in person and in print, is befitting of someone who has made unparalleled contributions to her field. I am grateful for every minute that I got to spend working with her over these past few years. Her attention and encouragement meant more to me than she could ever know.

Finally, I want to thank my extraordinary dissertation advisor, Michael Devitt. I worry that I'll simply never be able to repay the massive debt that I owe him for all the help and support over the past few years. I took Michael's course, *The Psychological Reality of Language*, in my very first semester as a graduate student. Having been thoroughly indoctrinated into the Chomskyan mode of thought in my undergraduate courses, I was shocked to find Michael voicing opposition to what I had come to regard as core and obvious truths. I have a vivid memory of walking aimlessly around Manhattan after each of his class sessions, so deeply immersed in thought that I could barely see the ground in front of me. It was as though I was talking to him in my head, developing objections to his views and then immediately countering them with the replies that



suddenly seemed to emerge from what he had said in class or written in what was then a rough draft of *Ignorance of Language*. It was a most exhilarating experience. Once in a while, I would stop myself, take a step back, and think: “This! *This* is what it’s like to do philosophy!” What a wonderful thought to have! It alone made going to graduate school worthwhile.

I aspired to emulate Michael’s clarity of thought and decided to take every course that he would offer in the years hence. My experience in his course, *Scientific Realism*, was no less exciting. Here, though, I found that Michael and I agreed on almost everything. I came away from the course more committed to scientific realism than I ever had been—something that has remained a fixed point despite massive overhauls in other areas of my thinking. Other courses followed: *Meaning, Reference, Animal Minds*, and two more instances of *The Psychological Reality of Language*. Then several conferences. (I’ll never forget watching Michael and his daughter Kate sing sports anthems at the dinner table after the PhilMiLCog conference in Ontario. The joy and camaraderie in their voices was a pleasure to behold.)

When it came time to choose a dissertation topic, the decision was a no-brainer. Over the years, I’d developed an interest in a wide variety of issues, but none had as strong a grip on me as the psychological reality of language. Here was an issue whose resolution required familiarity with not only philosophy, but also linguistics, psychology, neuroscience, and computer science—all the fields that I had ever wanted to delve into. Michael’s encouragement in the early stages was crucial. Taking on such a daunting challenge is not something one can do without proper guidance. I trusted Michael to set me on the right path, and he did. His coaching prior to my prospectus defense was immensely helpful. Shortly after that, I took a full-time position at

Baruch College, where the heavy teaching load made it impossible for me to work on the dissertation. I am grateful to Michael for his patience with me throughout those two and half years, as well as his occasional and gentle reminders that there is unfinished business to attend to. When I did finally get to work, Michael diligently read and commented on more than 1000 pages of my writing. I was stunned by the amount of care and thought that he put into this project. I am especially grateful to him for steering me away from fruitless tangents and major pitfalls. Eventually, when I had something to show for myself, Michael invited me to Croatia, to present my work to an audience of philosophers and linguists at the annual conference in Dubrovnik. This led to my first major publication—a condensed early version of chs. III and V. Finally, in the late stages, Michael helped me to resolve issues of tone and style that, in retrospect, may well be as significant as the content itself. For all of these things, and for his help on the job market in recent months, I am grateful beyond words.

The list of fellow graduate students with whom I've had meaningful conversations, beautiful connections, and memorable times is blessedly long. I want to express my deep appreciation to Jennifer R. Corns, Daniel W. Harris, Jacob F. Berger, Kyle E. Ferguson, Rosemary Twomey, Monique Whitaker, Myrto Mylopoulos, Chris Allen Sula, James Dow, Mike Nair-Collins, Dan Shargel, Josh Dulberger, Frank Pupa, Mark Zelcer, Alex Kiefer, Thomas Kivatinos, Ellen Fridland, Andrew Porter, Ryan DeChant, Arina Pismenny, Mark Alfano, Rachel Fedock Golas, Michal Klineciewicz, Wesley Buckwalter, Oliver Marshall, Ben Abelson, Elmar Geir Unnsteinsson, Henry Shevlin, Rachel McKinney, Joshua Livingston, Corey Evans, Mark Robertson, Rebecca Traynor, Sharisse Kanet, Stuart Hal Smith, Svetlana Nagumanova, Tim Czech, Yuval Abrams, Dan Shabbasson, Dan Leafe, Leonard Finkelman, Sergio Gallegos,

Matías Bulnes, Esteban Withrington, David Morrow, Jean-David LaFrance, Benjamin Daniel Young, Brian Robinson, Bana Bashour, Jeremy Ginsburg, Wade Martin, Laura Di Summa, Damien Dupont, Anatoly Nichvoloda, Sophia Bishop, Jon Golin, Jake Quilty-Dunn, Cosim Sayid, Tony H. Y. Cheng, Jared Blank, James G. Snyder, Roblin Meeks, James M. Hitt, Peter Langland-Hassan, Maria Brincker, Russell Marcus, Richard Brown, Douglas Meehan, Joshua Weisberg, Fritz J. McDonald, Andrew Bonfilio, Teresa and Edmund O’Neil, Erin Quirk, Adam Faulkner, Steffi Nickels, Ronit Danya Shaham, Andrew Bonfilio, Emily Zane, Rachel Szekely, Erika Troseth, Lucia Pozzan, Susan Schweitzer, Walter H. Dean, Elliot Michaelson, and Matt Moss.

Of the folks named above, I would like to express special thanks to those who read and commented on various pieces of the dissertation: Daniel Harris, Fritz J. McDonald, Chris Allen Sula, Ellen Fridland, J. M. C. Dow, Michael Nair-Collins, Jacob F. Berger, Elliot Michaelson and, most of all, Jennifer R. Corns, who read word of every chapter and commented on many of them.

I’ve had the pleasure of working with many wonderful people in the past few years. My thanks goes out to the faculty and students at Baruch College and the Brooklyn College Writing Fellows program: Douglas Lackey, Mary Gin, Thomas Teufel, Hagop Sarkissian, Sandeep Sreekumar, Amy Elizabeth Trautwein, Michael Garra, Gerrit Kamperdyk, Carl Hammer, Jonah Goldwater, Josh Dulberger, Pete Graham, William James Earle, Boris Maizel, Florian Lengyel, Corey Frost, Karin Kohlmeier, Jessica Hammerman, Pamela Albanese, Liz Weiss, Lilia Melani, Dov Zamore, Joshua Kaufman, Rebecca Buyanovsky, Sarah Felsenthal, Dmitriy Povazhuk, Eugene Goyfman,

Denis Gostev, Shiri Tamir, Erjon Xholi, Flora Olegovna, Christina Avenoso, Daniella Bondar, Natali Nassimian, Inayat Kaur, Regina Duzanskaya, Maksim Kotlovsky, Eugene Danielyants, Natalia Gayova, Mike Marceante, Stephen De Jesus Frias, Joseph DeSimone, Ilia Barboutev, Kirill Kan, Yash Chavda, Tyler Alterman.

Several professors from other schools have reached out to me at various points in my graduate career, offering advice and encouragement. I wish to thank Willem DeVries, Anthony Dardis, Georges Rey, Steven Gross, Uriah Kriegel, Endre Begby, and Pete Mandik, for their help, and for showing me how rewarding a life in philosophy can be. I admire them all a great deal and can only dream that my contributions to the field turn out to be as rich and well received as theirs have been. My special thanks to Endre Begby and Pete Mandik for their generous help on the job market in the past few months, and for being so warm and kind to me in general. Their friendship means a great deal to me.

My life has been enriched beyond measure by my close friends, who stuck by me when I was a nervous wreck and did their best to help me pull through. My sincere thanks to the following people for the joy, laughter, and special times that we've shared, and will continue to share for years to come: Konstantine Barsky, Igor Bernstein, Alia Rose Connor, Roman Palitsky, Jeff Arancherry, Sarah Poleshuck, Eugene and Elena Kravstsov, Yuri Niyazov, Galia Silva, Erika Hart Mayo, Toni Mayo, Dan Ribaud, Sarah Schumacher, Adam Ribaud, Claire Heitlinger, Christopher Plaskett, Michael Philip Fisher, Rosie Twomey, Eric Weinschenk, Thomas Livezey Laughlin, Allison Lack, Betty Chen, Caroline Brennan, Susie Benton, Renata Ko, Daniel Harris, Margot Otten, Aoife Mooney, Pete and Rachelle Mandik, F. Oriana Horowitz, Jennifer Marie

Dunn, Paul Kim, Dani Leeds-Kim, Ryan Egan, Erika Bruchhausen, Benny Gershman, Mike Marceante, Nick Devyatkin, Lev Michael Barsky-Zorin, Daniel and Ilona Grafman, Kevin F. Mason, Marianna Kosharovsky, Ruthy Robbins, Adele Kudish, Sage Jobsis, Stiles and Jake Najac, Ali Gruber, Louis Torchio, Sarah McNeil Barton, Joshua and Suzanne Cotton, Danielle Brenner, Tamir Novotny, Roman Finkel, Gabe Marin, Steve Zhik, and Ella Bromblin.

The people named above are my dearest friends, my closest colleagues, and have all had a profound influence on my personal and professional life. But one person stands out as playing all of those roles and many more. Jennifer Renee Corns is a truly gifted person—an uncommonly warm, sincere, and kind person. I was blessed to have met her. The times that we spent together in the past several years are some of the most memorable of my entire life. And now, I am deeply honored to be her friend and her colleague.

Jennifer is, without a doubt, the most talented young philosopher I have ever come across—a gift to the discipline. Her approach to philosophy is as strikingly fresh as it is effective. Her independence of thought is both remarkable and admirable. And her achievements in the field are all the more impressive, given the countless obstacles in her path. With the odds stacked against her, Jennifer pursued her dream with tenacity, systematic rigor, and an unrivaled passion.

Jennifer's dissertation, written in parallel with mine, was not just a success; it was a triumph—an important contribution, for which she is now only starting to get the recognition that she so clearly deserves. If there is justice in the world, she will enjoy a meteoric rise in the field.

It must be noted, however, that Jennifer's talents and intelligence extend well beyond the narrow confines of academia. I have heard philosophers debate whether there can be such a thing as moral genius. Jennifer renders the debate otiose. Her moral compass is the most accurate and reliable I have ever encountered. Her devotion to what is true, beautiful, and good is palpable in every choice that she makes. She was, and continues to be, an inspiration—not just to me, but to many others. I am grateful to her for being there for me, for caring about me, for showing me what is good about the world, and what needs to be done for there to more of it. May she encounter success commensurate with my gratitude, both in her professional life and in her efforts to help everyone around her.

Finally, I would like to thank my family, including my folks in the Ukraine—Tanya, Alexander, and Vladyslav Trostyanetsky—and those dearly departed: Katherine Tur, Rimma Trostyanetskaya, and Boris Trostyanetsky.

I owe the biggest debt of gratitude to my wonderful and everloving parents, Aron and Larisa.

My father, Aron Efimovich, is one of the most amazing people I know. He taught me the power of humor, the value of careful thinking, and the importance of selflessness, generosity, and humility. His courage, integrity, and warm character are beyond compare. To know him is to love him and to respect him. He is an inspiration not only to me, but to countless others—a shining beacon of health, balance, and well-being. His powerful intellect and clarity of thought set the standards to which I aspire. I am blessed to have a father whose mind, body, and spirit

work in such harmony. His love and devotion throughout my life have been for me a constant source of confidence and inner peace.

My mother, Larisa, is an extraordinary person, whose keen intuition, warm soul, and unconditional love are a guiding light for me in everything I do. She is a part of me, present in every word I speak, every choice I make, every relationship I form, and every emotion I feel. Without her attention, strength, and dedication, I would be nothing. She has devoted her entire life to me, from the moment she knew that she would be bringing me into the world. Her every action is aimed at helping me, encouraging me, comforting me, strengthening me, and preparing me for the challenges to come. Having lived a very difficult life, she has worked tirelessly to make sure that my experience in this world is as rich, carefree, and joyous as possible. I cannot even begin to imagine how I could possibly repay her.

Together, my parents made incredible sacrifices to ensure a better life for me in the US than I would have had in Ukraine. They worked hard to provide me with every kind of nourishment and intellectual stimulation, fostering every interest and supporting every foray into new terrains.

From the very beginning, they took a keen interest in all aspects of my academic development, making heroic efforts to place me in good schools both here and in Ukraine, attending parent-teacher conferences, making sure that I did all that I could to get myself accepted to Stuyvesant High School, pushing me to go to college (despite vociferous and, in retrospect, quite ridiculous protests on my part), and encouraging me to transfer from SUNY New Paltz to NYU. When I announced that I would be attending graduate school in philosophy they were more supportive

than I could ever have hoped for. Though skeptical about the job prospects in philosophy—not unwisely, as it turns out—they understood that my passion for the subject was the important thing. I can only hope that the accomplishment embodied in the present work serves as tangible validation of their support.

I love my parents more than I can ever express. I dedicate this work—and all my achievements, past and future—to them.



# TABLE OF CONTENTS

---

Abstract iv

Acknowledgements vi

## **Chapter 1    LANGUAGE AND LINGUISTICS: THREE VIEWS ON ONTOLOGY AND METHODOLOGY**

- 1.    Introduction    1
  - 1.1    Broad Aims    1
  - 1.2    Background    2
  - 1.3    Historical Preliminaries    4
- 2.    Against Abstractism    11
  - 2.1    Katz's Argument Against Epistemological Holism    11
  - 2.2    The Infinitude of Language and the Ontology of Linguistics    17
  - 2.3    Concluding Remarks and a Preview of Things to Come    28
- 3.    Chomsky, I-language, and the Cognitivist Conception of Linguistics    30
- 4.    Devitt's Linguistic Conception    49
  - 4.1    Metaphysics    49
  - 4.2    Distinguishing Competence from its Products, and Linguistics from Psycholinguistics    63
- 5.    Summary and Conclusion    66

## **Chapter 2    THE COGNITIVE CONCEPTION AND ITS DISCONTENTS**

- 0.    Introduction    69
- 1.    Does Pursuit of Explanatory Adequacy Require the Cognitive Conception?    71
  - 1.1    Linguistic Universals and The Explanatory Adequacy Condition    71
  - 1.2    Defusing an argument for the cognitivist conception    77
- 2.    E-language and I-language    86
  - 2.1    Motivating the study of E-language    86
  - 2.2    Do linguists really study I-languages?    101
  - 2.3    Individuating I-languages    109
- 3.    Residual Issues and Concluding Remarks    117

### **Chapter 3    MENTAL PHRASE MARKERS IN SENTENCE PROCESSING**

- 1     Introduction    130
2.    The Argument from Neurocognitive Data    135
3.    The Argument from Structural Priming        144
4.    The Argument from Garden-Path Effects    152
  - 4.1    Garden Path Phenomena        152
  - 4.2    Initial Parsing Preferences: On the Use of  
      Non-syntactic Information    164
  - 4.3    Conclusion        170
5.    The Argument from Filler-Gap Processing   171

### **Chapter 4    ON TWO ATTEMPTS TO DO WITHOUT MENTAL PHRASE MARKERS**

1.    Introduction    192
2.    The Integrated Processing Model of Comprehension        196
  - 2.1    Assessing the Motivations for “Integrated Processing” Models    197
  - 2.2    Integrated Processing Models: Inner Workings and  
      Local Achievements        208
  - 2.3    Criticisms of the “Mostly-Semantic” Models        215
  - 2.4    Conclusion        223
3.    The Brute-Causal Model of Language Processing    225
  - 3.1.1   RTM, CTM and the Language of Thought Hypothesis        226
  - 3.1.2   The Standard Arguments for LOTH        228
  - 3.1.3   The Public LOTH vs. the Strengthened LOTH        235
  - 3.1.4   Devitt’s Argument for SLOTH        242
  - 3.1.5   Pinker’s Objections to SLOTH        251
  - 3.1.6   Concluding Remarks About SLOTH        258
  - 3.2.1   The Brute-Casual Model of Language Processing    259
  - 3.2.2   The Relation Between Lexical Items and Concepts   275
  - 3.3.1   A Precise Characterization of Brute-Causal Processing    281
  - 3.3.2   Is Language Comprehension an Associationist Process?    285
  - 3.3.3   Is Language Comprehension Simple, Rigid, or Reflex-Like?   288

|       |                                                                                      |     |
|-------|--------------------------------------------------------------------------------------|-----|
| 3.3.4 | The Distinction Between Responding and Representing                                  | 292 |
| 3.3.5 | The Distinction Between Responding- <i>as</i> and Representing- <i>as</i>            | 298 |
| 3.3.6 | Is the Comprehension Process Sensitive to the Structure of<br>Mental Phrase Markers? | 300 |
| 3.3.7 | What Does All This Tell Us About the Nature of<br>Mental Representation?             | 302 |

## **Chapter 5    A HISTORICAL SURVEY OF COMPUTATIONAL PARSING MODELS**

|     |                                                                          |     |
|-----|--------------------------------------------------------------------------|-----|
| 1   | Introduction                                                             | 307 |
| 1.1 | A Preliminary Look at the Rules that Govern<br>Language Processing       | 309 |
| 1.2 | The Oracle                                                               | 310 |
| 1.3 | The Algorithm                                                            | 316 |
| 1.4 | The Grammar, First Pass: Context-Free<br>Phrase Structure Grammars       | 322 |
| 2.  | Foundational Parsing Algorithms                                          | 327 |
| 2.1 | Basic Terms                                                              | 327 |
| 2.2 | Top-down Parsing                                                         | 329 |
| 2.3 | The Earley Algorithm and Probabilistic Context-Free Grammars             | 333 |
| 2.4 | Bottom-up parsing                                                        | 342 |
| 2.5 | The Cocke-Younger-Kasami (CYK) Algorithm                                 | 347 |
| 2.6 | Left-corner Parsing                                                      | 351 |
| 2.7 | Parsing as Deduction                                                     | 354 |
| 2.8 | Efficiency Considerations                                                | 359 |
| 3.  | The Psychological Reality of Transformational Rules                      | 360 |
| 3.1 | A Primer on Transformational Rules                                       | 360 |
| 3.2 | On the Psychological Reality of Transformations                          | 362 |
| 4.  | Augmented Transition Networks                                            | 373 |
| 4.1 | Parsing with an ATN                                                      | 373 |
| 4.2 | The bearing of ATNs on the psychological reality issue                   | 380 |
| 4.3 | Problems for the ATN framework                                           | 384 |
| 5.  | Principles and Parameters in Syntax and Parsing                          | 392 |
| 5.1 | The First Grammar in the P&P Framework:<br>Government and Binding Theory | 392 |
| 5.2 | Principles-based parsing                                                 | 401 |

|                   |                                                                        |     |
|-------------------|------------------------------------------------------------------------|-----|
| 6.                | Minimalist Grammars                                                    | 410 |
| 6.1               | Minimalism: The Basics (Features, Merge, Move, and Spell-Out)          | 410 |
| 6.2               | Parsing with Minimalist Grammars                                       | 416 |
| 6.2.1             | Berwick and Harkema                                                    | 417 |
| 6.2.2             | Weinberg                                                               | 423 |
| 7.                | Philosophical Issues                                                   | 430 |
| 7.1               | Internalism and Externalism about Representation                       | 430 |
| 7.2               | The Personal/Subpersonal Distinction                                   | 441 |
| 7.3               | Computational Psychology, Hardwired Rules,<br>and Procedural Knowledge | 449 |
| 7.4               | A Useful Analogy                                                       | 463 |
| 7.5               | Summary and Conclusion                                                 | 467 |
| <b>References</b> |                                                                        | 471 |

# I

## LANGUAGE AND LINGUISTICS: THREE VIEWS ON ONTOLOGY AND METHODOLOGY

---

### §1. Introduction

#### §1.1 *Broad Aims*

W. V. Quine effected a groundbreaking shift in philosophy, ushering in a mode of inquiry that aims to clarify and resolve foundational issues, while bearing no pretensions to primacy or ultimate authority. Thus the birth of a naturalistic philosophy—one that takes careful notice of developments in the sciences and brings them to bear on questions that have historically dominated philosophical debate. The present work adopts this methodological stance toward a cluster of issues associated with the claim that the rules and principles of natural language are “psychologically real.” My broader goal is to clarify the conceptual framework within which we theorize about psychological states and their relations to behavior.

In chapters III-V, I will launch a sustained argument for the psychological reality of two kinds of representational state. The first, which I call *mental phrase markers*, play an important role in language comprehension, by tracking the syntactic properties of incoming linguistic stimuli (chs. III-IV). The second, *mental syntactic principles*, comprise a neurally-implemented grammar, which I will argue facilitates language processing by guiding the construction of mental phrase markers (ch. V). Our discussion in those chapters will draw heavily on the results of behavioral and neurocognitive studies, as well as computational simulations of language processing. By contrast, the material in this chapter and the next concerns more traditional philosophical questions, dealing mostly with the ontology and methodology of linguistic theory. I will examine the prospects of three competing approaches to these questions—platonism,

nominalism, and cognitivism—arguing that the first of these is defunct, while the latter two remain viable options, each with its distinctive virtues and flaws.

## §1.2 *Background*

Our interest in the present chapter is the metaphysics of language. Methodological naturalism counsels us to give priority to the following question: What conception of language is playing a role in the working linguist's theories? Unfortunately, the question, so formulated, fails of presupposition. Nothing answers uniquely to the description "*the* working linguist." There are dozens of schools, members of each propounding countless theories, each pitched at any of a range of "levels of analysis." A theorist who takes herself to be "doing linguistics," and is so taken by others in her field, may have any of the following goals in mind:

- To lay bare the phonological, morphological, syntactic, semantic, and prosodic properties of the expressions comprising every human language
- To show how such expressions are used to communicate in various contexts
- To characterize the effects of socio-economic status on the acquisition and use of language
- To map out the perceptual and motor functions involved in the comprehension and production of language (spoken, written, signed, and so on)
- To determine the history of both the structure and use of various aspects of language
- To describe the social and cognitive processes whereby human children acquire language
- To detail the neurophysiological structures and mechanisms that underpin or implement language acquisition, comprehension, and production
- To ascertain the genetic apparatus responsible for the construction of the initial neural structures that make acquisition and processing possible
- To offer an account of the origins and the evolution of our capacity to acquire and use human languages
- To describe the relations between the acquisition and use of language, on the one hand, and the acquisition and use of other cognitive, perceptual, and motor capacities, on the other
- To characterize the language-specific disorders that a person might suffer from, and to isolate their genetic, neurological, perceptual, and articulatory sources
- To devise methods for diagnosing and assisting such persons
- To program artificial machines in such a manner as to make them suitable for useful or engaging interaction with competent human speakers

It's safe to say that no monolithic conception of language will do justice to all of these diverse goals. To take a prominent example, consider "Language is a social art"—the opening line of

Quine's seminal *Word and Object* (1960). Though valuable in its own right, the conception of language that is embodied in this claim can hardly be expected to highlight the concerns of a syntactician who is bent on refining the latest conception of *Merge* and *Move* (ch. V), to say nothing of a neurolinguist striving to ascertain the significance of early left anterior negativity (ch. III).<sup>1</sup> Still, while we cannot reasonably expect a tidy account of what language *is*, careful attention to both the methodology and the results of actual linguistic theorizing can have profound consequences for a wide range of philosophical projects. Or so I hope to show.

In what follows, I restrict my attention to the generative tradition, pioneered in the mid-1950s by Noam Chomsky and others. This is not to suggest that other traditions are not worth exploring, much less that they carry no merit; indeed, it would surprise me if that were the case. Rather, my choice of focus reflects the fact that the generative tradition has much to recommend it as an object of historical and philosophical scrutiny. It is at once a field young enough to admit of open foundational questions and, at the same time, sufficiently developed as to enable one to draw on substantive results. (In philosophy, fixed points are all too rare and always welcome.) Furthermore, the presumptions of generative linguistics are rooted in two widely agreed-upon tenets of contemporary analytic philosophy: (1) the rejection of mind-body dualism and (2) the commitment to supplying an analysis of linguistic structures that goes well beyond the acoustic and orthographical superficialities, yielding (one hopes) some insight into the nature of logical form and inferential practice. These presumptions, in turn, place the generative enterprise in a prime position to make deep contributions to the broader field of cognitive science—another longstanding interest of naturalistically inclined philosophers.

---

<sup>1</sup> The second line of *Word and Object* will fare no better with contemporary acquisition theorists in the generative tradition: “In acquiring [language] we have to depend *entirely* on intersubjectively available cues as to what to say and when” (1960: ix, emphasis added).

Linguists in the generative tradition are not exactly unified in their views; even at fairly coarse levels of analysis, there is no canonical doctrine. What common ground there is can be stated in the following relatively broad strokes: A precise characterization of the syntactic and semantic structure of linguistic expressions is best stated in a recursive formalism. What is characteristic of the generative enterprise is its well-founded confidence that (i) such a formalism can be arrived at by methods that are, in essence, no different from those at play in other empirical sciences, and (ii) that the statement of this formalism is absolutely central to the resolution of outstanding questions in other branches of linguistics.<sup>2</sup>

While the introductory texts in generative grammar present what initially appears to be a well-articulated and adequately motivated conception of language, a closer look at the history and sociology of the field reveals deep fissures. A useful starting point in examining these issues is an anthology edited by the late Jerry Katz, aptly titled *The Philosophy of Linguistics* (1985).

### *§1.3 Historical Preliminaries*

Katz opens the anthology by pointing out that there have been two major revolutions in recent thinking about language. First, under the influence of towering figures such as Russell, Wittgenstein, Carnap, and Quine, the analysis of linguistic expressions displaced metaphysics and epistemology as the central focus of philosophical theorizing in the first half of the 20<sup>th</sup> century. Accordingly, philosophers became interested in linguistics, in an effort to put their theories on what seemed to be a more secure footing—a project that remains alive in much

---

<sup>2</sup> Needless to say, a commitment to both the existence and innateness of a Universal Grammar is characteristic of many theories in the generative tradition. But it is not, I think, a defining feature of that tradition. There has been no shortage of detractors from the innateness thesis, some of whom are nevertheless plainly generative grammarians. (I have in mind specifically the linguists Geoffrey Pullum, Paul Postal, and Stephen Levinson, though I have no doubt that other prominent examples can be given.)



contemporary work. Second, under Chomsky's influence, generative grammar came to prominence as the most fertile approach to linguistic analysis.

Chomsky's seminal contributions were threefold. First, he introduced a new kind of syntactic formalism—transformational grammar—that was clearly superior to the extant taxonomic grammars and to Carnap's logical syntax. Second, he put linguistics on a rationalist and mentalist footing, claiming that linguistics studies *partially innate knowledge structures*. This required the introduction of the notion of tacit knowledge, which will serve as the primary focus of the present inquiry. The innateness thesis—which, by contrast, receives little attention in the chapters to come—was advanced in order to account for the speed and facility with which children acquire language, in spite of an alleged poverty in the data to which they are exposed in the early years of development. Accordingly, the innate knowledge was<sup>3</sup> to have three components: i) a specification or template of all possible human grammars, ii) a simplicity metric, along which those grammars are ranked with respect to observational data, and iii) a device for taking in the observational data and ranking the grammars along that metric, eventually ranking highest what is sometimes called the “target grammar.”<sup>4</sup>

---

<sup>3</sup> The past tense is important here, for Chomsky's views of the matter have evolved considerably since the publication of *Aspects of the Theory of Syntax* (1965). In his *Lectures on Government and Binding* (1981) and *Knowledge of Language* (1986), Chomsky introduced and elaborated the Principles and Parameters model, according to which the innate endowment differs, in important respects, from what he had previously taken it to be.

<sup>4</sup> It is worth pausing over this common piece of terminology and taking stock of the connotations it carries with respect to the debates surrounding what Chomsky has called E-language and I-language. The term ‘target grammar’ carries the implication that there is a grammar external to the child at the time of acquisition, which the child is struggling to grasp. Chomsky and his followers would surely resist this implication, claiming that the use of the term is appropriate only in the informal presentation of a theory, not in its serious development—a distinction that Chomsky wields with worrying frequency. (See, e.g., his exchange with Rey, his exchange with Rey, in Barber [2000] and Antony and Hornstein [2003].) Still, whether a linguist uses the term or eschews it, we are owed a formal, “serious” statement of the child's “goal” throughout the acquisition process. This would go a long way toward making sense of what acquisition theorists mean when they claim that a child has made a “mistake” in the course of acquisition, or, equally, that no mistake was made. I am not convinced, at present, that an explicit, formal, and “serious” account of this would make no mention of E-language. See ch. II (§3) for further discussion of this issue, as well as of Chomsky's claim that the notion of E-language plays no role in linguistic theory.

Chomsky's third lasting contribution was to decisively discredit the methodology favored by nominalists like Bloomfield, Harris, and Quine—the so-called “discovery procedures” that employ substitution criteria in constructing a taxonomy of linguistic forms. On Chomsky's then-radical view, a theoretical notion was to be considered legitimate for use in linguistic inquiry if it proved necessary for the construction of an optimal grammar for a language—where optimality goes well beyond mere observational adequacy (ch. II). Chomsky argued forcefully that grammars are not to be seen as mere descriptive botanies of linguistic expressions, but are, instead, substantive empirical hypotheses, to be judged by their explanatory success. A grammar was to be evaluated on the basis of the underlying structures it assigns to various linguistic constructions—the requirement of *descriptive* adequacy—as well as the role that those structures play in a child's acquisition of a specific language—the requirement of *explanatory* adequacy.<sup>5</sup>

Katz goes on to note (1985: pp. 10-11) that what he calls “the second linguistic turn” gave rise to the proliferation of a large variety of formal descriptions of linguistic structure:

- Transformational grammar (*Syntactic Structures*)
- The Standard Theory (*An Integrated Theory of Linguistic Descriptions; Aspects*)
- Extended Standard Theory (*Studies on Semantics in Generative Grammar*)
- Revised Extended Standard Theory (*Reflections on Language*)
- Government and Binding Theory (*Lectures on Government and Binding*)
- Lexical-functional grammar (Joan Bresnan, *Lexical-functional grammar*)
- Montague Grammar (Richard Montague, *Formal Philosophy*)
- Arc-pair grammar (Johnson and Postal, *Arc-pair grammar*)
- Abstract Phrase Structure Grammar (Gazdar)<sup>6</sup>

---

<sup>5</sup> For an impressively lucid discussion of the history sketched above, see Fodor, Bever, and Garrett (1974), chs. 2-3. See also Blumenthal (1970) and Townsend & Bever (2001), ch. 2. I discuss the conceptual link between explanatory adequacy and language acquisition in the next chapter.

<sup>6</sup> Though relatively inclusive, this list is still radically incomplete, as Katz is well aware. (Indeed, he notes that he has mentioned “only the most prominent cases.”) Postal and Langendoen (1985: p. 243), provide a more comprehensive, though still incomplete, list. I reproduce their list here, omitting the entries that appear also on Katz's list, as well as the dates of the relevant publications: Finite Grammar (Hockett), Finite State Grammar (Reich), Realistic Grammar (Brame), Stratificational Grammar (Lamb; Lockwood), Tagmemics (Longacre), Natural Generative Grammar (Bartsch and Vennemann), Semantically Based Grammar (Chafe), Functional Grammar (Dik), Daughter Dependency Grammar (Hudson; Schachter), Phrasal Core Grammar (Keenan), Corepresentational Grammar (Kac), Relationally Based Grammar (Johnson), Dependency Grammar (Hays), Categorical Grammar

This staggering assortment presents a daunting challenge to the philosopher who aspires to make linguistics her object of analysis. Katz writes:

The philosopher who now wishes to make use of linguistics faces a bewildering complexity. With so many theories available in linguistics, philosophers unfamiliar with the issues either have to let their desire to apply linguistics to philosophy go unrealized or risk having happen to their applications of linguistics what happened to Quine's application of substitution criteria. One navigational aid would be a robust philosophy of linguistics. It would provide an ongoing examination of theoretical developments in linguistics, classifying the theories that emerge, highlighting the philosophically important differences between them, and putting them in a form that is more accessible to philosophers generally. Specialists in the philosophy of linguistics would present the philosophically significant issues between alternative theories of linguistic structure in the way that philosophers of logic have presented the philosophically significant issues between alternative logics. (1985, p. 11)

Noble as this vision was, it's hard to credit the thought that it was implemented to the fullest by contemporary philosophers. True, philosophical discussions of *semantic* and *pragmatic* theories have been lively and fruitful. But the number of philosophers who closely follow ongoing developments in phonology, syntax, psycholinguistics, neurolinguistics, acquisition theory, or computational linguistics is depressingly small. In connection with the final sentence of the above-quoted passage, compare, for instance, (i) the number of philosophers who can rehearse the difference between S4 and S5 modal logics at the drop of a hat, and (ii) the number of philosophers who can say anything of substance about the difference between the Government and Binding theory of the 80's and 90's, and the Minimalist Program that has slowly replaced it. Though the latter contrast is surely more pronounced, it is rarely discussed in the philosophical literature, while the former is common ground among philosophers of language.

---

(Lambek), Cognitive Grammar (Lakoff and Thompson), Meaning-Text Models (Melcuk), The Abstract System (Harris), Configurational Grammar (Koster), Neostructural Grammar (Langendoen), String Adjunct Grammar (Joshi, Kosaraju, and Yamada), Equatorial Grammar (Sanders), and Systemic Grammar (Hudson).

Katz thought that there was an important reason for philosophers to engage in a concerted study of contemporary linguistic theory: it is here, he claimed, that the infusion of carefully defended *metaphysical* positions can have a striking effect on the direction of empirical research. Indeed, he argued that the *foundational* questions in linguistics are ontological: What sorts of things is linguistics *about*? To what ontological categories do they belong?

Traditionally, there have been three positions on the matter. Nominalists (e.g., Bloomfield, early Quine, and Goodman) claim that linguistics is about token physical objects or events. Cognitivists or “conceptualists” (e.g., Chomsky, Fodor) argued that it is, rather, about mental states and cognitive processes.<sup>7</sup> The philosophers whom Katz calls “realists” (e.g., Frege, Katz, Postal and Langendoen) take the objects of inquiry to be *abstracta*—entities lacking causal powers and locations in spacetime, and which are therefore accessible to the mind only via a special faculty of intuition.

Katz’s use of the term “realism” is unfortunate here, for a number of reasons. First, a coarse taxonomy of the familiar range of positions in the philosophy of science will include realism, instrumentalism, and outright antirealism. Katz’s label has the unwelcome consequence of suggesting that linguists and philosophers who hold competing views, including Devitt, Chomsky, and Fodor, are either instrumentalists or antirealists—a preposterous thought. Second, the label suggests that abstract objects (whatever these might be) are more *real*, in some sense, than either physical objects or mental states. Given the uncontroversial existence of physical objects and the well-established existence of mental states, and keeping in mind the deep skepticism about the existence of *abstracta* in contemporary philosophy, one rather balks at the

---

<sup>7</sup> Katz wavers between the terms ‘conceptualism’ and ‘mentalism’. I will use the label ‘cognitivism’, which is sufficiently inclusive and has the added virtue of avoiding distracting connotations stemming from the technical uses of the term ‘concept’ in contemporary psychology and philosophy of mind.

suggestion.<sup>8</sup> My friends working in ancient philosophy tell me that the term ‘Platonism’ (or even ‘platonism’) is not much better, for it suggests the false view that Plato (or Socrates) held anything like a commitment to what contemporary philosophers think of as *abstracta*.<sup>9</sup> I recommend the label ‘abstractism’—an atrocious neologism, even by the standards of philosophical jargon—because of the problems noted above, concerning the use of the traditional labels, ‘realism’ and ‘platonism’. Stylistic and historical matters aside, the label reflects, I think, the core ontological commitments of the view. I will employ it henceforth.

Each of the three traditional ontological positions admits of substantive dispute among its adherents. The core nominalist claim, for instance, can be cashed out in the fashion of resemblance nominalism, predicate nominalism, and trope theory (to name just a few positions). Abstractists come in many stripes as well; Katz (2000: ch. 1) lists “classical Platonism,” “contemporary Aristotelianism,” and “naturalized realism” as the going candidates. The same is true of cognitivist approaches, as we will see in some detail in the chapters ahead. Nevertheless, Katz’s coarse-grained tripartite distinction allows him to paint a historical picture that is characterized by what he sees as the two major scientific revolutions in linguistics. On this story, the first revolution consisted in Bloomfield’s revolt against 19<sup>th</sup>-century mentalistic linguistics, which established nominalism as the reigning ontological orthodoxy within the field. The second was Chomsky’s cognitive revolution, against Bloomfield and the behaviorists, which swung the pendulum back in the other direction, replacing nominalism with cognitivism, which remains the received view today. Having sketched this history, Katz writes:

---

<sup>8</sup> Postal (2009) comments: “Katz’s terminological choice was not persuasive definition. Rather (see Quine 1953: 14), *realism* names one of the three main mediaeval points of view regarding universals, that view, often called Platonist.” Given the countervailing pressures noted in the main text, one wonders why adherence to medieval terminological standards should have determined Katz’s choice of nomenclature.

<sup>9</sup> One wonders, though, what Plato’s doctrine of *anamnesis* amounts to when stripped of the apparent ontological commitment to *abstracta*. This is not the place to pursue that issue.

It is not surprising that the ontological issue has had such historical importance when one realizes that all the major questions in the foundations of linguistics depend on how the issue is resolved. What kind of science one takes linguistics to be—whether it is put with empirical sciences like psychology or with non-empirical sciences like mathematics and logic—depends on how one resolves the ontological issue. Similarly, what the nature of grammatical argumentation is, how we are to understand the claim that a statement about sentences or a language is true, what a grammatical fact is, how such facts are known, and what the essence of a language is, are all questions to which we will give different answers depending on whether we are nominalists, conceptualists, or realists. (Katz, 1985: 14-15)

This brief passage sets the agenda for the rest of this chapter and the next. In §2, I take issue with Katz’s distinction between empirical and nonempirical (or “intuitional”) sciences. Even with respect to the hard cases of mathematics and logic, I follow Quine (1986) in holding that Katz’s position rests on an untenable dualism.<sup>10</sup> I show that Katz’s master argument against Quine’s epistemological holism does not work (§2.1) and that his considerations regarding the infinitude of natural language do not militate in favor of abstractism (§2.2). In the remaining sections of this chapter (§§3-4), I introduce the cognitivist position of Chomsky (1965, 1986, 2000) and the nominalist view recently developed by Devitt (2006, 2008, 2010a). I sketch the main tenets of these positions, reserving critical discussion for ch. II.

---

<sup>10</sup> Though I favor Quinean epistemological holism, I hasten to distance myself from the behaviorist views with which Quine is nowadays associated. I take this to be an instance of the strategy—which I quite recommend—of proceeding in full accordance with Quine’s methodological naturalism (Quine, 1969) while casting aside the more dubious aspects of his philosophical psychology (Quine, 1974) and his commitment to the use of substitutional criteria in linguistics (Quine, 1953, 1956, 1970.) For what it’s worth, my own (perhaps heterodox) view is that Quine’s adherence to the kind of behaviorism championed by Pavlov, Watson, and Skinner is not as nearly strong as many recent commentators contend. Certainly, it is absurd to see Quine as holding the Skinnerian doctrine that goes under the label “analytic behaviorism,” according to which statements about the mind can all be *translated, one-by-one and without loss of information*, into statements about overt behavior or dispositions to behave. Quine rejected the notion of analyticity on which this doctrine crucially relies. For a discussion of this point, see Dennett (1969: p. 31; 1978: ch. 4). In all likelihood, Quine, like many of his contemporaries, was reacting to an uncritical reliance on the introspectionist methodology of taking first-person reports as direct insights into the nature of the mind. (See, e.g., Quine’s discussion of Pierce’s “behaviorism,” in Quine, 1981: p. 36). This is of a piece with Quine’s doubts concerning what he dubbed “the museum myth of meaning.” Quine’s behaviorism is, I think, closer to the quite reasonable epistemological position that Sellars (1963a) calls “methodological behaviorism,” according to which our only evidence for the ascription of mental states consists in data garnered from the third-person point of view. As for Quine’s ontological position, there is a case to be made that it was not at all a kind of behaviorism, but, rather, a kind of proto-functionalism, complete with a token-token identity theory and an acceptance of Davidson’s “anomalous monism.” (See, e.g., Quine, 1975b). This, again, points to a kinship with Sellars, who is regarded by some—rightly, in my view—as the father of functionalist theories in the philosophy of mind. (See ch. V: §7.)

## §2. Against Abstractism

### §2.1 Katz's argument against epistemological holism

Recall that Katz writes, “What kind of science one takes linguistics to be—whether it is put with empirical sciences like psychology or with non-empirical sciences like mathematics and logic—depends on how one resolves the ontological issue.” Katz gives two examples of sciences that are allegedly non-empirical: mathematics and logic. The claim that these disciplines are non-empirical encompasses a number of related epistemological and metaphysical theses. One is that their subject matter consists of so-called “abstract objects”—e.g., numbers, functions, and propositions. The mathematician and the logician are taken to be studying the “essences” of such objects, as well as the “metaphysically necessary” (as opposed to contingent) relations that hold between them. Another such claim is that the truths in these fields are discovered in a way that does not rely, in any significant sense, upon perception, observation, or experiment. Rather, they are established by the methods of deductive proof or conceptual analysis. The body of knowledge that they deliver is, thus, *a priori*. And, on a traditional conception of a priori knowledge, it follows that the results in these fields—unlike the results in, say, physics—are immune to revision on the basis of empirical and pragmatic pressures.

Abstractism is the conjunction of these claims—i.e., the view that some fields provide us with *a priori*, unrevisable knowledge of the metaphysically necessary truths concerning a realm of nonphysical entities that transcend spacetime and enter into no causal relations. Although abstractism is most often concerned with logic and mathematics, Katz (2000) holds that it is also the correct view of the ontology and methodology of linguistics.

Though common in the philosophical literature, abstractism has its share of opponents. A particularly influential view that denies the main tenets of abstractism has been championed by W. V. Quine and his followers. Here is a characteristic statement of Quine's position.

Mathematics and logic are supported by observation only in the indirect way that those aspects of natural science are supported by observation; namely, as participating in an organized whole which, way up at its empirical edges, squares with observation. I am concerned to urge the empirical character of logic and mathematics no more than the unempirical character of theoretical physics; it is rather their kinship that I am urging, and a doctrine of gradualism. ... A case in point [is] the proposal to change logic to help quantum mechanics. The merits of the proposal may be dubious, but what is relevant just now is that such proposals have been made. Logic is in principle no less open to revision than quantum mechanics or the theory of relativity. The goal is, in each, a world system—in Newton's phrase—that is as smooth and simple as may be and that nicely accommodates observations around the edges. If revisions are seldom proposed that cut so deep as to touch logic, there is a clear enough reason for that: the maxim of minimum mutilation. The maxim suffices to explain the air of necessity that attaches to logical and mathematical truth. ... This much can be said for the linguistic theory of logical truth: we learn logic in learning language. But this circumstance does not distinguish logic from vast tracts of common-sense knowledge that would generally be called empirical. There is no clear way of separating our knowledge into one part that consists merely in knowing the language and another part that goes beyond. (Quine, 1986: p. 100)

In this passage, Quine challenges the epistemological claims of abstractism, leaving aside its metaphysical claims. Quine's attitude toward abstract entities evolved over the course of his career. Having defended nominalism in his youth (Goodman and Quine, 1947), he later went on to reluctantly admit the existence of *some* abstract entities—sets, individuated extensionally—on account of their alleged indispensability in the natural sciences.<sup>11</sup> Whatever the case about that, Quine has always been unambiguous in his rejection of propositions and other abstract entities for the purposes of linguistics. In what follows, I rebut an argument due to Katz (2000), to the effect that Quine's epistemological views cannot be sustained. We turn to questions of ontology in §2.2.

---

<sup>11</sup> Orenstein (2002) calls Quine a "reluctant Platonist, admitting only as many abstract objects, such as sets, as are indispensable for the business of science" (p. 86).



In his book, *Realistic Rationalism* (2000), Katz explicitly discusses the Quinean epistemology that I favor. Though he presents Quine's views with palpable veneration, he nevertheless thinks that there is a knock-down argument against the Quinean approach.

Full appreciation of the power of Quine's empiricist account of logic and mathematics must compel admiration from even the staunchest rationalist. It is quite surprising that an account based on so extreme a form of empiricism can steer contemporary empiricism past the Scylla of Millian inductivism and the Charybdis of logical empiricism and come so close to capturing the special certainty of mathematics and logic. Nonetheless, Quine's holistic conception of knowledge does not in the final analysis enable contemporary empiricists to provide a satisfactory account of the special certainty of logical and mathematical truth, because the conception is inconsistent. (Katz, 2000: p. 72)

He casts the alleged inconsistency—his sole objection to the Quinean approach in this book—in the form of a “revisability paradox,” which runs as follows:

On Quine's epistemology, noncontradiction, universal revisability, and simplicity are different from other principles in our system of beliefs in this respect: they are constitutive of the epistemology of the system. The epistemology is a belief-revision epistemology and those principles comprise the basic mechanism of belief revision. They thus serve as essential premises in every argument for reevaluating a belief. Every such argument has to assume the principle of noncontradiction as a rationale for departing from an assignment of truth to logically conflicting statements. The principle is required to initiate the process of revising presently accepted statements, otherwise we have to tolerate a radically laissez-faire epistemology on which anything goes. Further, every argument has to assume that the class of revisable statements is the class of all statements of the system, otherwise the epistemology will no longer be the uncompromising empiricism it was intended to be. Finally, every argument has to assume simplicity or something like it to narrow down the class of potentially revisable statements. ... Here is the paradox of revisability. Since the constitutive principles are premises of every argument for belief revision, it is impossible for an argument for belief revision to revise any of them because revising any one of them saws off the limb on which the argument rests. Any argument for changing the truth value of one of the constitutive principles must have a conclusion that contradicts a premise of the argument, and hence must be an unsound argument for revising the constitutive principle. (Katz, 2000: p. 73)

In sum: Quine's epistemology is supposed to underwrite the claim that *every* part of our belief system is revisable. But the very principles that are *constitutive* of that epistemology are themselves unrevisable, on pain of internal inconsistency. Hence, Quine's view is self-

contradictory. Despite his frank admiration of Quine's epistemology, Katz sees this argument as a fatal blow, sufficient to disqualify Quine's view from further consideration. He writes:

The revisability paradox shows that no form of uncompromising empiricism, at least none that we are presently aware of, can meet the epistemic challenge to antirealism. The paradox also undercuts the Quinean explanation of how truths of mathematics and truths of logic can be taken to be about natural objects in the Quinean (1961c, 44) sense of being part of "a device for working a manageable structure into the flux of experience." (Katz, 2000: p. 74)

In contrast to Katz, I believe that Quine's view of the matter is essentially on target, that there is no revisability paradox, and that the illusion of paradox comes from failing to appreciate the status of the principles of noncontradiction, universal revisability, and simplicity. This failure is, in turn, the result of a broader misconception about the nature of epistemology. I will elaborate on these claims in the remainder of this section.

Katz's mistake comes in precisely when he claims that the principles of noncontradiction, universal revisability, and simplicity "serve as essential premises in every argument for reevaluating a belief" and that "every such argument has to assume the principle of noncontradiction as a rationale for departing from an assignment of truth to logically conflicting statements." Katz believes that the principle of noncontradiction "is required to initiate the process of revising presently accepted statements." These are all erroneous statements. The abovementioned principles play no such role in Quine's epistemology, nor, I claim, in *any* viable epistemology. They are *not* premises that we ever have occasion to sincerely marshal in open debate, and they do *not* figure explicitly in our reasoning. Rather, they are useful *descriptions of* our reasoning practices. They are *true of* our belief-maintenance and revision processes in roughly the way that the principles of celestial mechanics are *true of* our solar system.

Let us take these principles in turn. The principle of noncontradiction is not "required to initiate the process of revising presently accepted statements." Rather, there is a true

psychological generalization, according to which, *as a matter of fact*, when two beliefs come into direct conflict with one another, one of the beliefs is revised. No supervisory faculty is needed for this to occur. That is, reasoners need not ask themselves, “Can I revise this belief?”—all the while holding the belief in their cognitive grip, as it were—and then answer in the affirmative after having consulted their explicit commitment to the principle of universal revisability. Rather, it’s that rational beings are *in fact* prepared—indeed, in a certain sense, *programmed*—to revise any belief, given a suitable range of conflicting evidence. Indeed, it’s their ability to do this that qualifies them for the honorific label ‘rational being’. Finally, when revision does take place, reasoners do not begin by restating anew their strong desire to achieve a maximally simple overall worldview—not even tacitly. Rather, the belief revision process *in fact* takes the path of least resistance, according to a metric that, admittedly, no one has yet fully elucidated, and perhaps never will. All in all, it appears that the old distinction between *fitting* a rule and *being guided* by it—a distinction that we will examine closely in later chapters—has never been more apt.<sup>12</sup>

Why does Katz fall prey to such an elementary error, quite uncharacteristically? There are two interpretive possibilities. First, he may be in the grip of a model of reasoning according to which there is a “self” or an “overseer” who supervises and directs the belief-revision process, taking epistemic principles as explicit premises in his or her practical decision-making. This

---

<sup>12</sup> The fitting/guiding distinction is, I believe, not fine-grained enough to capture differences that will become important later. In ch. V (§7), I will follow Devitt (2006) in drawing a tripartite distinction between (1) merely *fitting* a rule, (2) *embodying* a rule, and (3) *representing* a rule—i.e., using it “as a premise” or “as data” in the course of cognitive or computational operations. My claim in the main text is that, *pace* Katz, the principles of universal revision, noncontradiction, and minimal mutilation do *not* belong in class (3). Whether they belong in class (1) or class (2) is, I take it, an open empirical issue. To settle it, we would have to determine whether, in the brain, there is a *common causal mechanism* (Davies, 1989, 1995) that is responsible for *all* of the cognitive operations that conform to each individual principle of reasoning. I suspect that there are no such mechanisms. Hence, my money is on class (1). A bit of foreshadowing: the burden of ch. V (§7) will be to clarify the distinctions among these three classes and to apply the tripartite framework to a discussion of the rules or principles of grammar. I will argue that we have no strong grounds, at present, for deciding between the strong hypothesis that the grammatical rules or principles belong to class (3) and the weaker hypothesis that they belong to class (2).

quasi-voluntarist model of reason is, I believe, very difficult to sustain. Its main strength lies in the fact that various quirks of introspection make it a tempting option for generation after generation of philosophers.<sup>13</sup> The second possibility is that Katz has failed to appreciate the force of Quine's naturalist approach to epistemology. According to Quine, the epistemologist does not furnish us with "first principles," on which to base our inferences and arguments, nor with normative injunctions to achieve certain epistemic ends. Rather, a naturalized epistemology shores up *descriptive*, *explanatory*, and *predictive* theories of psychological functioning—specifically, the cognitive processes of perception and inference.<sup>14</sup> It is plain that Katz does not subscribe to this approach. But Quine does. And the charges of inconsistency and paradox cannot be made to stick as long as that is the case.

Whatever the source of Katz's error, I think it's safe to say that we've effectively defused his alleged paradox.<sup>15</sup> To elaborate, there is no contradiction whatsoever in a theorist's coming to reject explicit statements of the principles of noncontradiction, simplicity, and universal revisability, in the course of reasoning *in accordance with those very principles*. All this requires is that the theorist be factually mistaken about his or her own psychological processes—a position that many psychologists have had the misfortune of occupying. Nor is there any

---

<sup>13</sup> See Dennett (1991) for an excision of the "central executive" and "central meaner"—vestiges of the Cartesian philosophy that retains its grip on the field, despite facing seemingly insurmountable problems. For a more recent attack on the view that state-consciousness serves an important epistemic function, see Rosenthal (2005, 2008).

<sup>14</sup> Michal Devitt and Hilary Kornblith have argued that empirical discoveries could show us not only that our theory of our own inferences was wrong, but also that the inferences we in fact make are not ones that we *should* be making. Indeed, Devitt points out that results like those of Kahneman and Tversky play precisely that role. This sort of view would allow us to see naturalized epistemology as having room for a normative element that is not merely a "chapter of psychology." However, it would not, I think, provide Katz with the resources for a cogent reply to the argument in the main text.

<sup>15</sup> On the supposition that the argument of this section hits the mark, Katz's case (2000: pp. 17-19) against the position taken by Penelope Maddy in her book *Realism in Mathematics* (OUP, 1990) is correspondingly weakened. I welcome this result, as it may be that Maddy's version of "realism" differs only terminologically from the Quinean position that I endorse.

reason to subscribe to a weaker version of the unrevisability claim, according to which the principles that correctly describe our reasoning today will *always* correctly describe it. For all that we presently know, reasoning practices may very well evolve over time, under any number of conceivable pressures—environmental, social, or biological. Indeed, if the change were to take place quickly enough, a theorist who rejected some explicit principle of reasoning may start out being mistaken about her own psychology but, by a curious kind of fortune, *end up* being correct.

## §2.2 *The Infinitude of Language and the Ontology of Linguistics*

As noted above, Katz conceives of the history of linguistics as being characterized by two completed revolutions—Bloomfield’s temporary victory over the mentalists and, eventually, Chomsky’s triumph over the nominalists. He suggests, however, that the evolution of linguistic theory is not yet at an end, for there is one final revolution left to go—a revolution that will overturn the dominant Chomskyan paradigm and herald in the reign of a superior alternative. Wielding the misleading label ‘realism’ (fn. 8) for what I have been calling ‘abstractism’, he takes some then-current work in linguistics to be moving in precisely this direction.

The interest of the ontological issue in linguistics is increased by the fact that there is now a number of linguists and philosophers who are trying to formulate a realist view of language and argue that it should replace the traditional conceptualist views that are presently accepted widely. [DP: In a footnote, Katz lists the following people: R. Montague, E. Itkonen, J. Ringen, J. Katz, D. T. Langendoen and P. Postal.] By making the realist position in linguistics more than a logical possibility, these linguists and philosophers have raised the question of whether linguistics might take the next step from conceptualism to realism.

As we have seen in the historical development from Bloomfield to Chomsky, there has been a correlation between the ontological position that linguists accept and the methodology and kind of grammar they propose. *A different ontology might thus open up a new approach to the study of language. The formulation of a realist position for linguistics give linguists some reason to think that new insights about grammatical structure may be forthcoming.* (1985: p. 15, emphasis added)

In a footnote, Katz cites Langendoen and Postal's book, *The Vastness of Language* (1984), in which the authors argue for the absence of a size constraint on natural languages and, on this basis, for the impossibility of describing language with certain types of grammars.

Before turning to Langendoen and Postal's arguments, I would like to first register a disagreement with Katz's vision of the historical and theoretical dialectic in linguistics. As is evident from the italicized passage in the quotation above, Katz believes that paying particular attention to ontological issues—indeed, according them pride of place, in some sense—may well give rise to his envisioned revolution in both philosophy and linguistics, provided of course that the abstractist option prevails. There are, I believe, two problems with the suggestion that formulating and defending a distinctly philosophical position about the metaphysics of language will yield “new insights about grammatical structure.”

First, Katz's suggestion implies that ontological issues are clear enough for us to be confident in challenging hard-won syntactic analyses on the basis of antecedently established ontological claims. That's a bizarre commitment to hold in the face of the stunning lack of convergence on the part of metaphysicians regarding ontological issues. As compared with the healthy progress and steady evolution that we find in syntactic theorizing, the “progress” in metaphysics, especially in ontology, is virtually nonexistent. I am not aware of a single *result* in metaphysics that is comparable in explanatory power to, say, the shift from pre-GB-era syntax to the Principles and Parameters theory (Chomsky, 1986) and the associated improvements in computational parsing models (Berwick and Fong, 1995).<sup>16</sup>

Second, Katz's claim seems to be out of sync with the way in which the linguists whom he cites actually proceed. Postal and Langendoen, for instance, first launch *empirical* arguments

---

<sup>16</sup> I offer an extended discussion of this historical shift and its bearing on the theory of sentence processing in ch. V.

for their grammars, and only *then* draw ontological conclusions. Indeed, Katz's claim is out of sync with how he characterizes *his own* thinking in a later publication:

Various considerations made a uniform realist interpretation seem preferable. Chomskyan linguistics distinguished between competence, the speaker's knowledge of the language, and performance, the speaker's exercise of that knowledge, but, in defining competence as the speaker's knowledge of the language, it also distinguished between the speaker's knowledge and what the knowledge is knowledge of. Furthermore, although generative linguists frequently preached on the theme of linguistics as a natural science, as working linguists, they formulated principles about the structure of sentence types. Since what makes such principles true or false is the structure they are about, the structure of sentences rather than the structure of our knowledge of them, it seemed more in line with linguistic practice to take grammars to be theories of sentence types rather than theories of the speaker's psychology. (Katz, 2000: p. xxv)

Even more clearly:

[I]f the concerns of a partisan viewpoint in the philosophy of mathematics are allowed to decide questions of grammatical structure, distinctions reflecting no grammatical differences will be made over a wide segment of the language. Since such distinctions are only philosophically motivated, a multiform semantics would compromise the autonomy of linguistics. Linguistic argumentation would degenerate into philosophical debate—Why should the linguist let the concerns of antirealism decide? Why not the concerns of realism? And so on. To preserve the autonomy of linguistics and the integrity of its argumentation, only linguistic considerations can be allowed to determine the description of sentences. (Katz, 2000: p. 31)

In both of these passages, Katz is insisting that issues having to do with the description of sentences—what I regard as *empirical* issues—are to be settled *prior* to the defense of ontological theses.

Putting aside the apparent exegetical inconsistencies in Katz's oeuvre, as well as the implausibility of his prioritization of ontological issues in linguistics, let us focus briefly on whether Postal and Langendoen's argument for an abstractist grammar in fact succeeds. I believe it does not.

A natural place to begin the discussion is with two background assumptions that are commonplace in generative grammar. The first assumption is that any sentence of any natural

language is finite in length, in the sense that it is constituted by finitely many phonemes or morphemes (themselves drawn from a finite stock). The second assumption is that there is no sentence,  $S$ , such that operating on  $S$  with a recursive rule of the grammar would not generate a longer sentence,  $S^*$ . Put more simply, the claim is that there is no longest sentence. These two assumptions—i.e., that all sentences are finite in length, and that they nevertheless unbounded—can be used to prove that the set of sentences in any natural language is *countably*, or, in a different terminology, *denumerably* infinite; in other words, the cardinality of this set is  $\aleph_0$ .

These assumptions are what Postal and Langendoen refer to as “size laws.” Sufficiently formalized grammars of a natural language will typically include one or another such law. What Postal and Langendoen pointed out was that the size laws stated in the previous paragraph—the laws that are standard in generative grammar—are not the only possible laws that we might adopt. For instance, we can imagine replacing the law that the length of natural language sentences is unbounded with a law which states that there is some fixed, finite number,  $N$ , such that all sentences of all natural languages are composed of no more than  $N$  phonemes or morphemes. This, of course, is *ad hoc* and unmotivated. The reasoning behind the unboundedness assumption is similar to (though not the same as) what we encounter in the case of natural numbers. There is no largest positive integer because, for any integer we pick, we can always add 1 to it, thereby yielding a larger integer. Similarly, there is no longest sentence, because you can always add ‘Jennifer knows that’ to any (declarative) sentence, thereby generating a longer sentence. (More on this below.)

Postal and Langendoen’s argument, however, is not aimed at replacing the standard size laws with more *restrictive* ones, like the law just described. Rather, they seek to *abolish* the size laws altogether! Their suggestion—argued at length in *The Vastness of Language*—is that, on



the grounds of simplicity, linguists should adopt a grammar that places *no limit at all* on the length of sentences, thereby allowing the grammar to output sentences of literally *infinite* length. They go on to prove that such grammars, when encompassing plausible ancillary principles, yield the consequence that there are *transfinitely* many sentences—i.e., that the set of sentences is uncountable or nondenumerable, or, in other words, that its cardinality is *greater* than  $\aleph_0$ .

Before evaluating Postal and Langendoen's claims, it is worth pointing out that, contrary to the conclusions that the authors draw in the book—the very conclusions that Katz is seeking to promote in the passages quoted above—the claim that there are transfinitely many sentences in fact has no significant consequences whatsoever for the ontological dispute between abstractists and their opponents. In subsequent remarks, Langendoen has written:

Paul and I spilled a lot of ink trying to clarify Chomsky's "conceptualist" view of natural languages, and having clarified it, to show that it is incorrect. But it now strikes me as quite easy to work out a coherent conceptualist theory of natural languages that does not incorporate a size law. In fact, any principles and parameters theory that does not do so, but that does include what strikes me as the patently correct principle of coordinate compounding, will yield the result that natural languages (albeit, E-languages, in Chomsky's terminology) are of transfinite size. (Langendoen, T., "Is language infinite? Replies to Manaster-Ramer," in *Linguist List* 2.602, Sept. 30<sup>th</sup>, 1991.)

Notably, although Langendoen has changed his mind, Postal continues to attack the dominant Chomskyan paradigm.<sup>17</sup> In Postal (2009) he writes:

The incoherence of Chomsky's position is worth putting slightly differently. To say a collection is discretely infinite is to say its members can be put in one to one correspondence with the members of one of its own sub-collections and also with the positive integers. What Chomsky's ontology then asserts is that the human brain component he calls FL [i.e., faculty of language] or a particular state of it defining a particular NL [i.e., natural language] embodies as aspects a collection of things which can be put in such a correspondence with the integers 1, 2, 3,... But there can't even

---

<sup>17</sup> It is worth remarking that Postal's claims are tinged with such venom and hostility that one finds it hard to believe that his motives are entirely academic. A typical sample from his 2009 paper, "The Incoherence of Chomsky's 'Biolinguistic' Ontology," reads as follows: "[M]uch of the persuasive force of Chomsky's linguistics has been achieved only via a mixture of intellectual and scholarly corruption" (p. 104).

begin to be enough of *anything* in a human brain or its functioning to ground such a correspondence. It is no accident then that across the decades of adumbrating his biolinguistic view, Chomsky has never even sought to specify the nature of supposedly biological entities which manage to have a transfinite cardinality. (p. 110)

Postal goes on to quote a passage from Langendoen & Postal (1984).

Even for an attested NL like English, the claim that a grammar, even a psychogrammar, generates mental representations immediately creates otherwise unnecessary fundamental problems. Evidently, either standard or radical conceptualist must minimally assume that any actual human mind or brain is finite, and thus that its very nature limits the objects which are in fact representable therein, for trivial non-linguistic reasons. Consequently, if psychogrammars generate mental representations of sentences and mental representations are, as the term suggests, things actually present in real minds, in something like the sense in which, say, data or computations are present in real computers, the radical conceptualist position claims that NL grammars have a finite output, one containing no representation of cardinality greater than some finite  $k$ . This is inconsistent not only with the traditional generative position but with claims in the very works where radical conceptualism is advocated that the domain of grammar is infinite. (pp. 131–132)

What both of these passages seem to miss is that the phenomena that the linguist seeks to describe, predict, and explain include not only *past* cognitive and verbal episodes—acts of production, comprehension, and perhaps metalinguistic judgment—but also *future* ones. And while the number of these—both past and future—is finite, there is no way of telling in advance *which* particular such episodes we will come to encounter. What we know from our observations thus far, however, is that such episodes have been *continually novel*. And we have no reason to expect that they will cease to exhibit *that* pattern.

Now, in the long run, advances in the biochemical and bioinformatic sciences may increase the human life span, alter our motivational states, and enlarge our memory resources. As a result, the sentences that we produce and understand may become not only continually novel (as they have been to date), but also significantly longer. And just as there is no telling in advance what the limits are on the linguistic creativity that gives rise, continually, to novel sentences, there is likewise no telling in advance just how long the longest sentence a human will

ever utter might be.<sup>18</sup> *This*, I submit, is the underlying justification for adopting the standard size law of sentence unboundedness.

Note, however, that there is *no* such justification for abolishing the size laws altogether. For, while we do not know how long the longest sentence ever might be, we *do* know this: No such sentence will ever be infinite! Whatever advances are made in life-extending, memory-enhancing, and motivation-boosting technologies, it remains the case that *a human being will never utter an infinitely long sentence*. This we can be sure of. Therefore, there is no empirical pressure for us to posit such sentences, *not even as occupying the status of the merely potential*.<sup>19</sup>

Postal and Langendoen share with Katz the view that the infinitude of language is a *datum*—one that can only be captured by the abstractist conception. (After all, in contrast with *real* objects, there’s always room for more *abstracta*!) The reply that I favor is nicely articulated by Devitt (2006).

Katz has another objection to nominalism: grammars are about an infinite number of sentences but there cannot be an infinite number of tokens. If there were a problem for my sort of nominalism it would lie in its apparent commitment to nonactual

---

<sup>18</sup> My claim, then, is that the following remarks from Rohde (2002) are based on a failure of imagination: “Language is *pseudo-context-free* or *pseudo-context-sensitive*, but infinite recursion is an idealization ungrounded in any observable reality. Some like to say that there are an infinite number of possible sentences in a language, meaning that there are an infinite number of utterances that conform to the rules of the grammars that linguists design. But if we limit the definition of a sentence, as I choose to do, to those utterances that could possibly convey a useful and consistent meaning to at least a few fellow speakers of a language, even under optimal conditions, the claim of an infinite variety of sentences is simply not true. There is a finite bound to the length and complexity of sentences that humans can comprehend, and thus a finite bound to the possible sentences we can create, short of inventing new words. If I were to link all of the sentences in this thesis together with *and* or some other suitable conjunction, the result would not be another sentence. It would be an abomination. I’m not arguing that there aren’t a vast number of possible sentences in English, just that there aren’t an infinite number” (pp. 3-4).

<sup>19</sup> The position developed here is shared by Collins (2008), who writes: “Note, the datum here is not that we understand an *infinity* of sentences. Given the vagaries of the notion of ‘understanding’ and metaphysical suspicions over the infinite, such a claim would needlessly attract irrelevant objections. The datum, rather, is that our competent use of language is ‘indefinitely new’ or ‘continuously novel’; far from finding ourselves stumped at novel sentences, the vast majority of sentences we produce and consume are entirely novel to us... If our competence were somehow finitely bounded, then our apparent ability to project ‘acceptable’ or ‘grammatical’ across an indefinite range would be a kind of fortunate accident, as if there were a finite upper bound, but one sufficiently great to be unnoticed. Such a speculation is irrelevant and, anyhow, false” (p. 32).

possible sentences, a problem that would arise even if we were dealing with a finite language (e.g. English with a limit of one million words to a sentence). The only significance of any apparent commitment to an infinite number of sentences is that it would guarantee that some were nonactual. But talk of there being nonactual possible outputs of a competence can be a mere manner of speaking (notes 6 and 8). So too can talk of there being an infinite number of such outputs. The truth behind the talk of the nonactual can be simply that the grammar is lawlike. [Note 8: So if something were a sentence, a *wh*-question, a passive, or whatever, it would have the properties specified for such items by the theory.] And the truth behind the talk of the infinite can be simply that there is no limit to the number of different sentence tokens that might be governed by the rules the grammar describes. (p. 26)

The idea here, I take it, is that the infinitude of language is *not* a datum. Indeed, it's not even *true* that there are, in existence, infinitely many linguistic tokens or types (nor that there have ever been, nor that there will ever be).<sup>20</sup> Rather, the infinitude is an artifact of theory construction—specifically, a reflection of the fact that we seek maximally wide generalizations in our effort to accommodate *any* potential datum that we might encounter. Chomsky (2001) puts the point as follows:

We also speak freely of derivations, or expression Exp generated by L, and the set of such Exps—the set that is called “the structure of L” in Chomsky (1986), where the I/E terminology is introduced. Evidently, these entities are not “internal.” That has led to the belief that some externalist concepts of “E-linguistics” are being introduced. But that is a misconception. These are not entities with some ontological status; they are introduced to simplify talk about properties of FL and L, and they can be eliminated in favor of internalist notions. One of the properties of Peano's axioms PA is that PA generates the proof P of “ $2 + 2 = 4$ ” but not the proof P' of “ $2 + 2 = 7$ ” (in suitable notation). We can speak freely of the property “generable by PA,” holding of P but not P', and derivatively of lines of generable proofs (theorems) and the set of theorems, without postulating any entities beyond PA and its properties. Similarly, we may speak of the property “generable by L,” which holds of certain derivations D and not others, and holding derivatively of an expression Exp formed by D and the set {Exp} of those expressions. No new

---

<sup>20</sup> Collins (2009) puts the point this way: “Controversy on this issue appears to arise from the common talk in linguistics and philosophy of natural languages being infinite. The infinity of English, as it might be, however, is not a phenomenon. The phenomenon is that particular organisms are continuously novel in their speech and understanding, which we theorize in terms of unbounded generation” (fn. 4). It is worth noting that, in other contexts, Collins is one of Devitt's fiercest opponents. The fact that both Collins and Devitt—the strangest of bedfellows—have converged on the same solution is, in and of itself, impressive. Perhaps this is not sufficient to suggest that the solution has merit, but it certainly inspires confidence.

entities are postulated in these usages beyond FL, its states L, and their properties. Similarly, a study of the solar system could introduce the notion  $HT = \{\text{possible trajectories of Halley's comet within the solar system}\}$ , and studies of motor organization or visual perception could introduce notions  $\{\text{plans for moving the arm}\}$  or  $\{\text{visual images for cats (vs. bees)}\}$ . But these studies do not postulate weird entities apart from planets, comets, neurons, cats, and the like. No “Platonism” is introduced, and no “E-linguistic” notions: only biological entities and their properties. (“Derivation by Phase,” pp. 41-42.)

As Chomsky makes clear, the point holds not only in the realm of linguistics—conceived as a mode of empirical inquiry—but also for the case of mathematics, geometry, and all of the so-called “formal sciences.” Indeed, the point holds with respect to *all* serious systematic inquiry, on account of its being a virtue of *any* descriptive-explanatory theory that its generalizations be maximally broad. (This point will re-emerge with a novel significance in ch. II, §2.)

John Stuart Mill famously held that our confidence in the claims of arithmetic is based on induction from past instances.<sup>21</sup> Although Mill did not put the view as carefully as he might have—appealing to induction where abduction would have served better—I think he was quite correct to insist that the epistemology of arithmetic is no different, at bottom, from that of any other area of empirical inquiry. One begins with claims that facilitate one’s encounters with the world—claims about the quantities of, e.g., apples, miles, dollars, and hours—and later formalizes one’s theory, when necessary, by selecting as “basic” those statements that entail the very claims that one has found serviceable. Crucially, the formalization inevitably involves an element of idealization. One’s navigation of the natural and social environment—e.g., the building of bridges and stock portfolios—rarely requires thinking about *very* large, let alone *infinite*, quantities. But, in the course of formalizing one’s commonsense theory of quantity, there appears to be no principled reason to set an upper bound on the number of successors a given number might have. This is, however, nothing more than a convenience, of the same kind

---

<sup>21</sup> Mill, John Stuart (1874). *A system of logic*. 8<sup>th</sup> edition. New York: Harper & Brothers.

as we find in other cases of scientific idealization. It does *not* signal that the subject matter is now a set of abstract objects. This is what Chomsky (2001) means when he writes:

One of the properties of Peano's axioms PA is that PA generates the proof P of " $2 + 2 = 4$ " but not the proof P' of " $2 + 2 = 7$ " (in suitable notation). We can speak freely of the property "generable by PA," holding of P but not P', and derivatively of lines of generable proofs (theorems) and the set of theorems, without postulating any entities beyond PA and its properties. (p. 42)

Making a similar point about idealization and "indefinitely large" quantities in linguistics, Devitt (2006) says:

[Talk of indefinitely large quantities] may appear to commit theories of outputs to the existence of unactualized possibilities, but it can be, and in my view should be, a mere manner of speaking. It is a convenient way of capturing that these theories, like all interesting ones, are lawlike. Strictly speaking, the theories quantify only over actual entities but the theories are, in some sense, necessary. So the talk captures the modal fact that if something *were* a horseshoe, a chess move, a *wff*, a bee's dance, or whatever, then it *would have* the properties specified by the appropriate theory of outputs. (How are we to explain modal facts? I don't know but, *pace* David Lewis, surely *not* in terms of unactualized possibilities.) (Devitt, 2006: p. 21, fn. 6)

The larger moral to draw from all of this, I think, is that debates about the cardinality of the set of natural language sentences are not debates about ontology—for they have no serious consequences in that direction—but, rather, concern the nature of the idealizations that linguists should make in pursuing their explanatory goals. Given that we have no idea how long a sentence can get—nor any grounds for advancing a principled estimate—and given that we have good reason to believe that the length of sentences is a function of life span, motivation, memory and the like, the natural idealization to make in constructing a grammar is that sentence length is unbounded but nevertheless finite. Postal and Langendoen's argument to the contrary misses the mark.

The best diagnosis of their failure is, I believe, that they misconstrue the aims and aspirations of linguistic theory. The construction of a grammar is not a formal game. In one way

or another, it will have to enter into a theory of natural phenomena such as speech and comprehension. Hence, whatever one's prior "metaphysical commitments" might be, one's grammar will have to one day be integrated with a psychological theory of acquisition and use. This is not to say that it's a foregone conclusion that grammars are, in some sense, "in the mind." That thesis must be both clarified argued for—the dual burden of ch. V in the present work. Rather, it's to say that whatever entities a grammar describes are going to have to enter *somehow* into an explanation of how flesh-and-blood creatures manage to acquire, and to put to use, their ability to produce and comprehend language.

The exact mode of integration of linguistics and psychology has, of course, not yet been specified. (Again, see ch. V). However, we can rule out in advance some extremely implausible suggestions that, to my genuine surprise, have actually been endorsed in the literature. For instance, Postal (2009) says the following: "Everett (2005) claimed *rightly* (*sic*) that if natural languages were biological, syntactic trees should be visible in CAT scans" (fn. 12, p.111, my emphasis). It's clear, I think, that Postal and Everett's claim is either factually mistaken or simply confused.

In ch. III, I will survey the extensive evidence that we have for thinking that the human sentence processing mechanism builds mental representations of the syntactic structure of incoming linguistic stimuli. One of my arguments—though by no means the only one—will draw on recent EEG data, i.e., recordings of real-time neuronal activity during comprehension tasks. Given that mental representations are ultimately patterns of neuronal activity, my argument will commit me to the claim that syntactic representations do, in a sense, "show up" in devices that are functionally similar to CAT scanners. However, it should go without saying that such representations do not *look* (visually) at all like hierarchical tree-structures, and cannot

simply be “read” by mere casual observation, even by a trained linguist. The interpretation of neural activity—i.e., the assignment of representational content to various brain processes—requires painstaking functional analysis. We can conclude that Postal’s claim is either false—because syntactic properties and relations *are*, in fact, neurally encoded—or else it rests on a naïve assumption to the effect that the brain uses *the very same notational format* that one finds in contemporary linguistics texts.

### §2.3 *Concluding remarks and a preview of things to come*

Considered as a package of views concerning the ontology and methodology of linguistics, abstractism offers us a mistaken picture of the field, of its aspirations, and of language itself. In §2.1, I sketched an alternative position, composed of the naturalism and epistemological holism of W. V. Quine. Given that Quine avoids commitment to an extravagant ontology of abstract entities and doesn’t posit a suspect faculty of “rational intuition,” I take his to be the default position—the one we should adopt, pending a powerful argument to the contrary.

We went on to examine Katz’s argument, according to which the Quinean view suffers from a fatal internal inconsistency—what Katz (2000) calls “the revision paradox” (pp. 72-74). I pointed out that Katz fails to distinguish between (i) our descriptive theory of our own psychological practices, and (ii) our actual psychological practices. Although Quine took (i) to be revisable—as all theories are—this does not commit him to the claim that (ii) is revisable in the relevant sense. Hence, no paradox arises. The Quinean position remains unscathed.

In §2.2, I argued that abstractism derives no support from the infinitude of natural language. Indeed, following Chomsky (2001), Devitt (2006), and Collins (2009), I urged that this infinitude be seen not as a datum, but rather an artifact of theory construction. As such, talk



of there being infinitely many sentences in natural language can and should be paraphrased as a commitment to the modal force of syntactic theory—i.e., its counterfactual-supporting character, or “lawlikeness,” in the sense of Goodman (1954/1983). The upshot is that neither the nominalist position nor Chomsky’s “biolinguistic ontology” face a threat. Certainly, Postal (2009) is mistaken in his dogmatic insistence that the latter is “incoherent.”

Having argued against abstractism, I propose to examine the two alternative views mentioned above—the sophisticated nominalist position developed by Devitt (2006) and its dominant rival, the cognitivism of Chomsky (2000).<sup>22</sup> These are the topics of §§3-4.

In §3, I sketch the main features of Chomsky’s influential perspective on language and linguistic theory. I discuss his equation of an individual’s idiolect with a psychological structure that he calls an “I-language,” and contrast this with the public, conventional “E-language” familiar from folk theory and structuralist linguistics. Next, I attempt to pin down Chomsky’s commitments regarding the nature of mental representation, as well as the relation he envisages between an I-language and the performance systems that are involved in the comprehension and production of language. This gives rise to a discussion of the competence-performance distinction, which in turn raises issues about the empirical methodology and naturalistic credentials of Chomsky’s cognitive conception of language.

Although Chomsky’s philosophical views form the foundation of contemporary linguistics (at least in the generative tradition), many theorists have raised substantive questions about the relation that he attempts to forge between grammar and the mind. Chomsky has characterized the competent speaker/hearer as “cognizing” her I-language and as having “tacit

---

<sup>22</sup> As noted in §1, there are several names for the view that Chomsky has developed. Traditionally, the view goes under the label ‘conceptualism’. However, in light of the role that the term ‘concept’ has taken on in cognitive science, this usage is almost certainly misleading. I will use the more suitable label, ‘cognitivism’.

knowledge” in the form of a “mental representation” of various syntactic rules or principles. Just what do these claims amount to, and what grounds are there for believing them? Such questions have become urgent in part because an alternative position in linguistic metatheory has recently been formulated and defended, viz., Michael Devitt’s “Linguistic Conception”—a contemporary descendent of nominalism that stands in sharp contrast to Chomsky’s cognitivism.

In §4, I discuss Devitt’s view, according to which language consists of token inscriptions, acoustic blasts, and other spatiotemporally localizable entities.<sup>23</sup> Devitt argues that the task of linguistic theory consists in accurately describing certain properties of these entities—e.g., grammaticality, constituency, and entailment—which allow such entities to play the role that they do in linguistic communities.<sup>24</sup> In defending the coherence of this view, it is crucial to say something about the hard cases—viz. the empty categories *wh*-trace, NP-trace, PRO, *pro*,  $\emptyset$ , *Op*, and others. I spell out Devitt’s notion of a higher-order relational property and discuss the import of his appeal to linguistic conventions, which I take to be at the core of his account of the metaphysics of empty categories. Finally, I follow Devitt in distinguishing sharply between a competence and its outputs, which in turn underwrites a robust and theoretically significant distinction between linguistics and psycholinguistics.

### **§3. Chomsky, I-language, and the Cognitivist Conception of Linguistics**

Over the course of more than five decades, in a series of groundbreaking works, Noam Chomsky has developed and continually refined a profound and highly influential conception of

---

<sup>23</sup> In using the term ‘entities’, I slur over the differences between objects, events, and processes. Such distinctions will be explicitly noted only when relevant.

<sup>24</sup> Talk of properties does not fit well with Devitt’s nominalist leanings. He takes it to be a mere *façon de parler*—harmless shorthand that can be regimented away when the chips are down. In §4.1, I point to a case where it is not clear what a suitable regimentation might be.

language. Central to his view is the idea that human beings are innately endowed with a dedicated neural structure for acquiring and using language. This “mental organ,” which Chomsky calls the Language Faculty (FL),<sup>25</sup> houses representations of various rules and principles for assigning internal structure to linguistic expressions. On this view, the object of linguistic inquiry is, in the first instance, a *psychological* entity—i.e., a person’s “linguistic competence.” Chomsky identifies this with what he calls an “I-language”—a body of domain-specific *knowledge* or *information* that every competent speaker has at his or her disposal.<sup>26</sup>

Although Chomsky’s specific hypotheses regarding what is represented within FL have evolved significantly over the course of his career—constituting, to his great credit, a breathtakingly rich body of technical work—his basic ideological assumptions have remained largely intact. The following passage, taken from Chomsky (1995), serves as a paradigmatic statement of his view, which overlaps substantially with the position espoused in his early publications (e.g., Chomsky 1965, 1975, 1980, 1986).

We are concerned, then, with the language faculty, which we understand to be some array of cognitive traits and capacities, a particular component of the human mind/brain. ... We call the theory of the state attained [by the language faculty] its *grammar* and the theory of the initial state [of the language faculty] *Universal Grammar* (UG). ... We distinguish between Jones’s *competence* (knowledge and understanding) and his *performance* (what he does with that knowledge and understanding). ... We may think of the language, then, as a finitely specified generative procedure (function) that enumerates an infinite set of structural descriptions (SDs). Each SD, in turn, specifies the full array of phonetic, semantic, and syntactic properties of a particular linguistic expression. (*The Minimalist Program*, pp. 14-15)

---

<sup>25</sup> Following Chomsky, I use the abbreviation ‘FL’ to avoid introducing yet further ambiguity into the abbreviation ‘LF’, which is commonly used to denote a level of formal representation in the Principles-and-Parameters (P&P) approach to syntax. The term ‘LF’ is already bedeviled by difficulties, which arise from the differences between its role in P&P theories and the role of ‘Logical Form’ in traditional logic and philosophy of language.

<sup>26</sup> As is customary, I use ‘speech’ as a cover-all term for the many types of linguistic signal—vocalizations, inscriptions, hand signs, etc.—and ‘speaker’ to refer to both the producers and the consumers of such signals. Where it matters to the discussion, I will, of course, distinguish between spoken, written, and signed linguistic signals, as well as between speakers, writers, signers, and their audiences.

To fix ideas about what structural descriptions are, fig. 1 below provides an example.

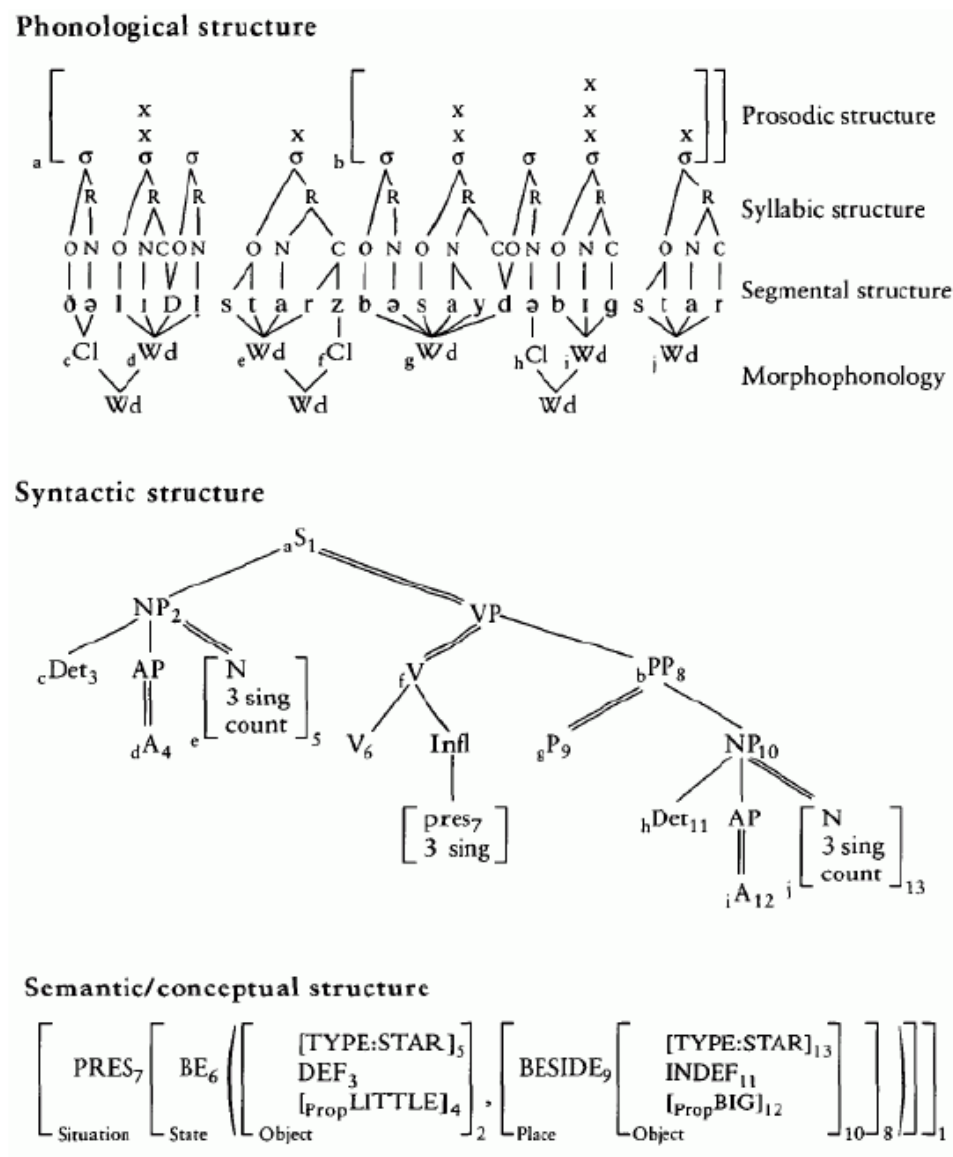


Figure 1: A partial structural description (SD) of the sentence ‘The little star’s beside a big star’ (taken from Jackendoff, 2002a). Chomsky (1995) characterizes SDs as “symbolic objects ... [each of which] specifies the full array of phonetic, semantic, and syntactic properties of a particular linguistic expression” (pp. 14-15).

A notable aspect of the passage quoted above is Chomsky’s stress on the claim that competence in a language involves “knowledge and understanding”—states that are widely

believed to be *representational* in character.<sup>27</sup> As noted above, the knowledge that the representations in the language faculty jointly constitute is what Chomsky refers to as the speaker's "I-language." In Chomsky's usage, an I-language is an explicitly psychologized grammar of a particular speaker's idiolect.

To distinguish this concept of language from others, let us refer to it as an *I-language*, where *I* is to suggest 'internal,' 'individual,' and 'intensional'. The concept of language is internal, in that it deals with an inner state of Jones's mind/brain, independent of other elements of the world. It is individual in the sense that it deals with Jones, and with language communities only derivatively, as groups of people with similar I-languages. It is intensional in the technical sense that the I-language is a function specified in intension, not extension: its extension in the set of SDs (what we might call the *structure* of the I-language). (*The Minimalist Program*, p. 15).

Importantly, I-languages must be distinguished from E-languages, which are conceived of as systems of signs common to a community of speakers. The 'E' in 'E-language' is meant to suggest 'external to the mind/brain', which is what such signs would have been if they were the physical tokens of the nominalist or the abstract types of the "platonist." By contrast, Chomsky's view is that linguistic entities are *in the mind*, rather than out in the world. He argues, furthermore, that constructing a notion of E-language inevitably requires drawing arbitrary boundaries between linguistic communities, or making reference to socio-political constructs that are strictly irrelevant to a scientific study of language. The 'E' in E-language is also intended to suggest 'extensional'. Chomsky (1986) rejects the view, which he associates

---

<sup>27</sup> The term 'knowledge' is a notoriously slippery one. Chomsky does not insist on it, and it is debatable whether he intends the term to denote full-blooded *propositional attitudes*. Many philosophers, including Fodor (1983, pp. 4-5) have taken him to hold this commitment and Devitt (2006) says that it is "natural" to do so (though his official view is noncommittal about this). Collins (2008, ch. 5) argues against this interpretation. In an attempt to shed the unwelcome connotations of the term 'know', Chomsky (1980, pp. 69-70) introduced the technical neologism 'cognize'. Determining the import of this notion is a crucial part of getting clear on Chomsky's view of the relation between a competent speaker and a grammar (i.e., "the generative procedure"). It is also important for determining whether his view entails that the states of the language faculty constitute what philosophers have called "knowledge-that," as against "knowledge-how." Chomsky (2003) provides a less-than-helpful attempt at clarification. In ch. V, I will argue that the states of the human sentence processing mechanism (HSPM) are best seen as subpersonal states rather than personal-level propositional attitudes, though nevertheless representational. While the relation between the FL and the HSPM is a complex one, I believe that my argument applies to both.

with Quine and Lewis, that grammars are functions-in-extension from sounds to meanings. All in all, his view is that E-language plays no serious role in linguistic theorizing.

[E-language] has no known status in the study of language. One might define E-language in one or another way, but it does not seem to matter how this is done; there is no known gap in linguistic theory, no explanatory function, that would be filled were such a concept presented” (*The Minimalist Program*, p. 16).

On the basis of this cognitive conception of language, Chomsky claims that there is no difference between a grammar’s being “psychologically real” and its being simply *true*. For, if a grammar is about a psychological entity—viz., I-language—then the distinction between the truth of a grammar and its psychological reality simply evaporates. This line of reasoning has been a recurring theme in Chomsky’s work, as the passages below illustrate.

We attribute psychological reality to the postulated representations. ... In short, we propose ... that our theory is true ... and give substantial independent evidence for the theoretical constructions, showing that the postulated principles explain many other facts of a similar nature, withstand empirical tests in English and other languages...” (Chomsky, 1978: pp. 206-208).

[W]hat is ‘psychological reality’ as distinct from ‘truth, in a certain domain’? ... I am not convinced that there is any such distinction. [...] the truth of the theories we are constructing; or if one prefers ... their ‘psychological reality’, though this term is best abandoned, as seriously misleading” (Chomsky, 1979: pp. 46-48)

We observe what people say and do, how they react and respond, often in situations contrived so that this behavior will provide some evidence (we hope) concerning the operative mechanisms. We then try, as best we can, to devise a theory of some depth and significance with regard to these mechanisms, testing our theory in by its success in providing explanations for selected phenomena. Challenged to show that the constructions postulated in that theory have “psychological reality,” we can do no more than repeat the evidence and the proposed explanations that involve these constructions. Or...we can search for more conclusive evidence...” (Chomsky, 1980a: p. 191)

As we will see, Devitt (2006) forcefully opposes this identification of a grammar’s truth with its psychological reality. We will examine Devitt’s alternative conception of grammars in §4.

According to Chomsky, the production and comprehension of speech requires not only the possession of an I-language, but also its participation in real-time, on-line computational

processes. In the course of such processes, the representations that constitute a speaker's I-language generate *instructions* to what Chomsky calls the "performance systems" of the mind/brain. More specifically, the representations generated by a person's I-language instruct her *articulatory system* to engage in certain muscle contractions and her *conceptual-intentional* system to interpret incoming linguistic stimuli in highly specific ways.<sup>28</sup>

Jones makes use of [her] competence to express [her] thoughts, to refer, to produce signals, to interpret what [she] hears, and so on. The language faculty is embedded in performance systems, which access the generative procedure. ... We might think of the SDs as providing *instructions* to the performance systems that enable Jones to carry out these actions [i.e., to express her thoughts, to refer, to produce signals, to interpret what one hears, and so on]. When we say that Jones has the language L, we now mean that Jones's language faculty is in the state L, which we identify with a generative procedure, embedded in performance systems. (*The Minimalist Program*, p. 15, emphasis added.)

We note, once again, the plausibility of interpreting Chomsky as claiming that states of the language faculty are "representational"—though, admittedly, in a sense that has yet to be defined. Indeed, in accordance with his slogan "No Computation without Representation," Jerry Fodor (1983, pp. 4-8) has argued that this interpretation of Chomsky is the *only* one on which talk of performance systems "*access[ing]* the generative procedure" and structural descriptions "providing *instructions* to the performance systems" (emphases mine) makes any sense. However, while Fodor's argument is compelling, it should be stressed that Chomsky's remarks, quoted above, are compatible with a broad range of perspectives on representation.

---

<sup>28</sup> Chomsky often suggests that an individual's language faculty assigns interpretations to languages that, intuitively, the individual does not know, so that "one might well learn about the languages of Jones and Wang [monolingual speakers of English and Chinese, respectively] by studying their reactions to utterances of Swahili" (*The Minimalist Program*, p. 16). Indeed, the suggestion is sometimes taken farther, as, for instance in the following passage: "The I-language ... assigns status to a vast range of physical events, including the utterance 'John seems to be sleeping', the utterance 'John seems sleeping', a sentence of Hindi, and probably the squeaking of a door, if we could do careful enough experiments to show how speakers of English and Japanese might differ in the way they 'hear' the noise" (Chomsky, 1988: p. 585).

Fodor's view of representation, which is surely the dominant view in contemporary philosophy, has it that representation is kind of *relation* between the mind/brain and various objects, properties, and events in a creature's extra-mental environment. However, alternative theories of representation have been proposed and defended. In particular, Chomsky (2000) argues forcefully that linguistics is an "internalist" or "individualist" inquiry. By this, he means that linguistics—which he conceives as a branch of psychology—deals solely with the properties of the mind/brain *irrespective* of the relations that it bears to anything in the environment. Thus, while it may be convenient to speak of various internal states as "representations," such talk should not, according to Chomsky, be interpreted as a commitment to the existence of a representation *relation*. Given the centrality of this issue, I think it is instructive to quote Chomsky at length.

Is there a problem with internalist (or individualist) approaches to other domains of psychology? So it is widely claimed, but on dubious grounds, I think. ... [T]he study of visual perception along lines pioneered by David Marr (1982), ... has been much discussed in this connection. This work is mostly concerned with operations carried out by the retina or, loosely put, the mapping of retinal images to the visual cortex. Marr's famous three levels of analysis—computational, algorithmic, and implementation—have to do with ways of construing such mappings. Again, the theory applies to a brain in a vat exactly as it does to a person seeing an object in motion. The latter case has indeed been studied, in work of Marr's collaborator Shimon Ullman (1979). His studies of the determination of structure from motion used tachistoscopic presentations that caused the subject to see a rotating cube, though there was no such thing in the environment; "see," here, is used in its normal sense, not as an achievement verb. If Ullman could have stimulated the retina directly, he would have done that; or the optic nerve. The investigation, Ullman writes, "concerns the nature of the internal representations used by the visual system and the processes by which they are derived." The account is completely internalist. There is no meaningful question about the "content" of the internal representations of a person seeing a cube under the conditions of the experiments, or if the retina is stimulated by a rotating cube, or by a video of a rotating cube; or about the content of a frog's "representation of" a fly or of a moving dot in the standard experimental studies of frog vision. No notion like "content," or "representation of" figures within the theory, so there are no answers to be given as to their nature. The same is true when Marr writes that he is studying vision as "a mapping from one representation to another, and in the case of human vision, the initial representation is in no doubt—it



consists of arrays of image intensity values as detected by the photoreceptors in the retina” (Marr 1982: 31)—where “representation” is not to be understood relationally, as “representation of.” ... The internalist study of language also speaks of “representations” of various kinds, including phonetic and semantic representations at the “interface” with other systems. But here too we need not ponder what is represented, seeking some objective construction from sounds or things. The representations are postulated mental entities, to be understood in the manner of a mental image of a rotating cube, whether it is the consequence of tachistoscopic presentations or a real rotating cube, or stimulation of the retina in some other way; or imagined, for that matter. Accessed by performance systems, the internal representations of language enter into interpretation, thought, and action, but there is no reason to seek any other relation to the world, as might be suggested by a well-known philosophical tradition and inappropriate analogies from informal usage. (Chomsky, 2000: pp. 158-160)

In chapter V, I will discuss the position that Chomsky endorses in this passage and argue that his internalism faces important and potentially insuperable difficulties. I will suggest that the theory of representation that does the most justice to psycholinguistic practice is what might be called ‘functional-role semantics’—an extension of the theory that Wilfrid Sellars proposed in some of his classic works of the 1960s and 70s. Sellars’s functional-role theory *does* make reference to mind-world relations—it is, in that sense, a theory of “wide” functional roles—but it *does not* require that there be a object or event in the environment to which a representation is related. Indeed, like Chomsky, the functional role theorist holds that no causal, informational, or teleological relation can be identified with “the representation relation.” But I digress. The issues here are large and they ramify quickly; we cannot hope to settle them in the scope of a few brief remarks. I postpone a fuller discussion until ch. V.<sup>29</sup>

Against the background sketched thus far, Chomsky formulates several guiding questions, the answers to which would jointly constitute a comprehensive theory of language.

---

<sup>29</sup> See also Devitt’s discussion, in ch. 4 of *Ignorance*, of other ways to interpret Chomsky’s remarks on representation and the psychological reality of grammar.

- a. What does Jones know when he has a particular language?
- b. How did Jones acquire this knowledge?
- c. How does Jones put this knowledge to use?
- d. How did these properties of the mind/brain evolve in the species?
- e. How are these properties realized in mechanisms of the brain?

(*The Minimalist Program*, p. 17)

Let us focus in on questions (a) and (c), which reflect the primary concerns of the present work.

Chomsky's answer to question (a) is this: In virtue of the fact that Jones' I-language generates structural descriptions for an infinite number of strings, Jones knows an infinite number of facts about the linguistic properties of expressions—properties such as rhyme (e.g., that 'pin' rhymes with 'bin'), grammaticality (e.g., that 'the child seems sleeping' is not completely well-formed), entailment (e.g., that the sentence 'Mary persuaded Bill to go to college' entails that Bill came to intend to go to college, though not that Mary did), ambiguity (e.g., that 'Mary is too angry to run the meeting' can mean either that *Mary* can't run the meeting, or that Mary is so angry that *we* can't run the meeting), and so forth.

Turn now to question (c), regarding the ways in which competent speakers put their "knowledge of language" to use. Chomsky writes that this question "calls for the development of performance theories, among them theories of production and interpretation" (*MP*, p. 18). Though he espouses a striking skepticism concerning the viability of this project,<sup>30</sup> Chomsky does say that "highly idealized aspects of the problem are amenable to study ... on the standard empirical hypothesis ... that one component of the mind/brain is a *parser*, which assigns a percept to a signal (abstracting from other circumstances relevant to interpretation)" (*ibid*). Chomsky then reveals what he takes to be the relationship between the parser and the mental representations that constitute an I-language: "The parser presumably *incorporates* the

---

<sup>30</sup> E.g., "Put generally, the problems are beyond reach: it would be unreasonable to pose the problem of how Jones decides to say what he does, or how he interprets what he hears in particular circumstances" (1995, p. 18).

[I-]language *and much else*, and the hypothesis is that interpretation involves such a system, embedded in others” (p. 18, emphases mine).<sup>31</sup> Note that Chomsky does *not* take the linguist’s formal grammar—i.e., the theory of the generative procedure that constitutes an I-language—to be a theory of the computational processes that are effected in comprehension and production. In other words, Chomsky does not think of an I-language as a parsing or production mechanism. Rather, “the parser ... *incorporates* the [I-]language” in some manner.

To clarify the issue, consider first a passage from Chomsky (1980), in which he discusses a well-known syntactic phenomenon, namely, the impossibility of extracting material from *wh*-clauses. Chomsky argues for an explanation of this phenomenon that, in effect, makes reference to three abstract levels of analysis—D-structure, S-structure, and LF. The phrase markers that belong at each of those levels are related by the formal operation of movement. While the details are fascinating, what matters more for present purposes is Chomsky’s commentary on the character of his proposed explanation—particularly its bearing on psychology.

Tentatively accepting this explanation, we impute existence to certain mental representations and to the mental computations that apply in a specific way to these mental representations [and] [w]e attribute psychological reality to the postulated representations and mental computations. Have we gone beyond the bounds of what is legitimate and proper, in so doing? I think not. ... [T]he argument sketched seems to me analogous in relevant respects to that of a physicist postulating certain processes in the interior of the sun. Of course, there are differences; the physicist is actually postulating physical entities and processes, while we are keeping to abstract conditions that unknown mechanisms must meet. We might go on to suggest actual mechanisms, but we know that it would be pointless to do so at the present stage of our ignorance concerning the functioning of the brain” (*Rules and Representations*, 1980: pp. 196-197).

The crucial point here is that the grammar does no more than impose “abstract conditions” on

---

<sup>31</sup> The phrase ‘and much else’ suggests that, on Chomsky’s view, the parser may have access to information that goes well beyond what is contained in SDs, or in the syntactic rules and principles that “generate” SDs. In particular, the hint seems to be that the parser has access to the encyclopedic information required for “pragmatic” inferences regarding the speaker’s message or communicative intent. Spelled out fully, the suggestion seems to be that the parser is not modular, in the sense of Fodor (1983). I shall explore this possibility in some detail in ch. III.

unknown mechanisms. We might ask, of course, how strong these conditions are. Specifically, must the mechanism construct separate D-structure, S-structure, and LF representations in the course of on-line processing? Must a parsing algorithm literally *move* parts of some of these representations in generating the others? Chomsky's remarks in an earlier publication, *Language and Mind*, make it clear that this *not* what he intends.

[I]t is important to distinguish between the function and the properties of the perceptual model PM and the competence model G that it incorporates... Although we may describe the grammar G as a system of processes and rules that apply in a certain order to relate sounds and meaning, we are not entitled to take this as a description of the successive acts of a performance models. In fact it would be absurd to do so.... If these simple distinctions are overlooked, great confusion must result" (*Language and Mind*, 1968/1972: p. 117).

In a similar vein, Jackendoff (2000) writes:

As has been remarked for decades, this framework of grammatical theory cannot serve directly as a model of processing. We were all taught as graduate students that one should not think of a syntactic derivation as modeling the course of processing: one doesn't think of the initial symbol *S* first, then gradually expand it till one chooses the words, then push the pieces around until one finally decides what the sentence means and how to pronounce it. So the notion of derivation has been distanced from processing by calling it "metaphorical"; this term has been applied especially to processes of syntactic "movement." (pp. 21-22)

Jackendoff's remarks in this passage echo an important footnote in *The Minimalist Program*, where Chomsky points out that "the ordering of [syntactic] operations is abstract ... with no temporal interpretation implied. In this respect, the terms *output* and *input* [to a movement operation] have a metaphorical flavor..." (p. 380, fn. 3).

The distinction between theories of linguistic structure and theories of psychological processing mechanisms is an immediate consequence of another distinction that Chomsky drew in *Aspects of the Theory of Syntax* (1965), between *competence* and *performance*. As noted earlier, a speaker's linguistic competence is identified, in Chomsky's work, with her knowledge

of the syntactic rules or principles that constitute her I-language.<sup>32</sup> While these are often characterized as components of a “generative procedure,” the passages quoted above should make it clear that the “generation” or “derivation” of structural descriptions *need not* be a process that takes place in the mind/brain in real time.<sup>33</sup> Rather, Chomsky believes that the I-language is somehow *incorporated* or *recruited* by the psychological mechanisms that effect comprehension and production. Such performance mechanisms, in contrast to the I-language that they incorporate, *do* perform computations in real time. Their inputs in comprehension<sup>34</sup> consist of both linguistic stimuli *and* the I-language, which can be seen (for these purposes) as a static data structure—in effect, a subpersonal analogue of a theory of the inputs.<sup>35</sup> The

---

<sup>32</sup> One might think of linguistic competence as an ability of some sort—an instance of what many philosophers, following Ryle, have called “knowledge-*how*.” In his early writings, Chomsky described it in these terms. For instance, in his famous review of Skinner’s book, *Verbal Behavior*, Chomsky wrote: “The construction of a grammar which enumerates sentences in such a way that a meaningful structural description can be determined for each sentence does not in itself provide an account of ... actual behavior. It merely characterizes abstractly the *ability* of one who has mastered the language to distinguish sentences from nonsentences, to understand new sentences (in part), to note certain ambiguities, etc.” (emphasis added). More recently, however, Chomsky has resisted this way of construing the nature of competence, insisting on an alternative construal that emphasizes what philosophers call “knowledge-*that*”. To my mind, his reasons for doing so are not compelling. For a lucid discussion of this issue, see Devitt (2006), ch. 6. I return briefly to this topic in ch. II.

<sup>33</sup> It must be noted that, in his early work, Chomsky flirted with the idea of treating a grammar as a model of processing. Berwick & Weinberg (1984) write: “Miller and Chomsky (1963) identified rules of the grammar with computational operations of the parser in a one-to-one fashion. This identification led to specific behavioral predictions, collapsing grammatical with processing complexity... Again, this simple first attempt was the natural one. If it had been correct, we would have learned a lot about the parsing device...” (pp. 38-9). The suggestion in Miller and Chomsky (1963) is, however, quite tentative. “The psychological plausibility of a transformational model of the language user would be strengthened, of course, if it could be shown that our performance on tasks requiring the appreciation of the structure of transformed sentences is some function of the nature, number, and complexity of the grammatical transformations involved” (p. 481; quoted in Berwick & Weinberg, 1984: p. 39)

<sup>34</sup> Reflecting a bias that pervades the present work—indeed, the entire field of psycholinguistics—I focus here solely on comprehension. In production, the inputs to real-time computations are presumably what philosophers call “propositional attitudes”—judgments, desires, intentions, and so forth. Chomsky often refers to the inputs as “states of the conceptual-intentional system,” leaving open the relation between these states and propositional attitudes. At present, the issues surrounding linguistic production are relatively murky, though work by Kay Bock and others has yielded important advances in our understanding. It is, moreover, an open question what exactly constitutes the *output* of comprehension. In general, empirical issues concerning non-peripheral cognitive states—the propositional attitudes or states of the “conceptual-intentional system”—seem to be far less tractable than one would have hoped.

<sup>35</sup> This corresponds to what Devitt (2006) calls “position (ii)” on the psychological reality of language. Devitt discusses other ways of interpreting Chomsky in ch. 4 of *Ignorance*.

psychological and behavioral phenomena that arise from the interaction between the I-language and these real-time computations fall under the heading of *performance*.

The competence-performance distinction has been the source of a great deal of confusion, and the topic of much heated debate, both in linguistics and philosophy. One set of issues that frequently arises in this connection has to do with whether, and to what extent, linguistic theory—conceived of as the study of what Chomsky calls “I-language”—is subject to confirmation or refutation by results in psychology and neuroscience. Consider, for instance the following remarks from Eddington (2008).

[S]ome linguists hold that their analyses reflect the language of an ideal speaker/hearer and not actual speakers. In like manner, many linguistic analyses claim to reflect a speaker’s competence—that is, the system of abstract mechanisms that are thought to underlie the speaker’s ability to produce and understand language—and not the speaker’s performance, which consists of actual utterances and other behaviors (Chomsky 1980: 205). Accordingly, hypotheses about ideal speaker/hearers or competence are irrefutable; they do not relate to observable, real-word entities. No tangible evidence of any sort could contradict them because they are hypotheses about fictional entities. The grammar of an abstract speaker/hearer may be an interesting topic of study just as a study of the psychology of the character Jean Valjean in the novel *Les misérables*, or a philosophical treatise on the inherent goodness of man; however, they are not scientific. If all observable manifestations of language relate to performance, they can never be directly relevant to the study of abstract competence that many linguistic studies purport to deal with. Derwing (1983: 66) demonstrates how the competence/ performance distinction serves to insulate a theory from possible falsification: “Suppose we find some child who is quite adept at basic arithmetic. One possible hypothesis about the ‘competence’ thought to underlie this skill might be to attribute the child [sic], not with something so mundane as a learned, laborious, step-by-step procedure for carrying out simple arithmetic operations, but rather with knowledge of number theory. And what if experimental results are found that seem to fly in the face of this hypothesis? Just chalk them up as ‘performance errors’ and the well-formed theory remains inviolate.” (p. 4)<sup>36</sup>

Chomsky would surely reply that Eddington has missed an important distinction, between

---

<sup>36</sup> See also de Beaugrande’s claim: “All data that the grammar cannot treat are shunted off into the domain of ‘performance’ and are excluded from consideration as nonissues.” *Text, Discourse, and Process*, Norwood, NJ: Ablex, 1980. (Quoted in Bever, Carroll, and Miller [eds.], 1984: p. 135.)

studying idealizations—e.g., the paradigmatic frictionless plane of basic physics—and studying outright fictions, like the character Jean Valjean. Pure fiction involves the description of *nonexistent* entities, whereas scientific idealizations are deliberately simplified descriptions of *actual* entities. True, in neither case do the descriptions *accurately* portray any real entity. But this resemblance is superficial. Idealizations are inaccurate because they are *incomplete*; fictions are inaccurate because there is nothing for them to be accurate *about*.<sup>37</sup>

In my view, Derwing’s contribution to the debate, reflected in the passage quoted above, likewise rests on an error. The line between competence and performance is admittedly quite blurry. (Indeed, I will capitalize on this fact in discussing the identity conditions on I-languages in ch. II.) But, *pace* Derwing, it does *not* follow from this that theorists have *no grounds at all* for rejecting certain proposals about the nature of competence. To take an extreme example, no linguist would be tempted “chalk up as ‘performance errors’ [*sic*]” an English speaker’s inability to comprehend German. This is *plainly* a competence issue. More subtly, no linguist has ever been tempted to propose that English speakers’ judgments about the unacceptability of (1) are cases of performance error, despite the acceptability of (2) in a *pro*-drop language like Italian.

- |                  |                               |
|------------------|-------------------------------|
| (1) *Is raining. | [Intended: It is raining.]    |
| (2) Piove.       | [Italian for ‘It is raining’] |

Other theorists have argued that, while results from psychology and neuroscience *can*, in principle, bear on linguistic theory, they have not, as yet, been *brought* to bear. This is the position adopted by Edelman and Christiansen (2003), who argue that the psychological reality of *Merge* and *Move*—the central formal operations in Minimalist syntax (ch. V, §6.1)—has not been empirically addressed, let alone demonstrated.

Unfortunately, to our knowledge, no experimental evidence has been offered to date

---

<sup>37</sup> I will discuss Chomsky’s specific idealizations in more detail in ch. II.

that suggests that merge and move are real (in the same sense that the spatial frequency channels in human vision are). Generative linguists typically respond to calls for evidence for the reality of their theoretical constructs by claiming that no evidence is needed over and above the theory's ability to account for patterns of grammaticality judgments elicited from native speakers. This response is unsatisfactory, on two accounts. First, such judgments are inherently unreliable because of their unavoidable meta-cognitive overtones, because grammaticality is better described as a graded quantity, and for a host of other reasons [Schuetze, 1996]. Second, the outcome of a judgment (or the analysis of an elicited utterance) is invariably brought to bear on some distinction between variants of the current generative theory, never on its foundational assumptions. Of the latter, the reality of merge and move is but one example; the full list includes assumptions about language being a 'computationally perfect' system, the copy theory of traces, the existence of Logical Form (LF) structures, and 'innate general principles of economy'. Unfortunately, these foundational issues have not been subjected to psychological investigations, in part because it is not clear how to turn the assumptions into testable hypotheses. (p. 60)

Edelman and Christiansen share an important theoretical commitment with Eddington and Derwing—namely, the claim that the communicative behaviors and linguistic intuitions of competent speakers are *not legitimate sources of empirical data*. This belief underlies Eddington's striking claim that some conceptions of linguistics render the inquiry in this field "nonempirical."

Some would argue that according to their conception of linguistics there is no need for experimental verification. These researchers would probably agree with Itkonen (1976:15-16) that the nonempirical linguist's goal is "to generate all and only intuitively valid formulae: insofar as they fail to do this, their systems are (non-empirically) falsified ... not by reference to some specific spatiotemporal occurrences, but showing that it does not capture the concept which it tries to capture." (See also Carr 1990:66, Kac 1992:39, Linell 1976:84-85) Linguistic analyses in this sense belong to a metaphysical or philosophical realm of inquiry that deals with axiomatizations about linguistic structure which 'make it possible to deduce all true statements about the system from a small set of prior assumptions about its nature' (Kac 1974:44), much in the same way arguments in philosophy or logic do. However, a serious concern in linguistics is determining whether the claims in an analysis are philosophical or scientific. Of course there are those who do clearly define their position. For example, Marantz (2005:440) believes that 'the categories and operations of generative grammar are hypotheses about the representations and computations in the minds and brains of speakers.' Bromberg and Halle also adopt a realist stance to phonology: 'Do speakers *really* retrieve morphemes from their memory, invoke rules, go through all these labours when speaking? We think they do' (2000:35). According to Marantz and Bromberg and Halle, linguistics belongs in the realm of scientific empiricism. At the opposite end



of the spectrum, Bradley claims that ‘grammars do not (and moreover, are not intended to) dictate the ways in which the computation of speaking and listening proceed ...’ (1980:38). Her definition of linguistics is clearly nonempirical. (pp. 6-7)

This passage suffers from a number of serious misunderstandings. First, as I argued in §2, the reasons that are traditionally given for denying that philosophy and logic are every bit as empirical as physics and chemistry are actually quite weak. Second, the quotation from Dianne Bradley at the end of the passage aims to distinguish theories of the competence grammar from theories of the computational performance mechanisms that draw on that grammar. But, *pace* Eddington, insisting on this distinction is *not* tantamount to endorsing the bizarre view that linguistics is “nonempirical.” At bottom, this is because Eddington’s view about the methodology of gathering linguistic intuitions from reliable informants is simply mistaken. As Lasnik and Phillips (2003) point out in their reply to Edelman and Christiansen, this methodology is not only extremely fruitful, but is *empirical* in a quite straightforward sense.

Gathering of native-speaker judgments is a trivially simple kind of experiment, one that makes it possible to obtain large numbers of highly robust empirical results in a short period of time, from a vast array of languages. Any good linguistics study involves carefully constructed materials, appropriate control items, and robust and replicable results. It is only because the technique is so easy and requires no more than a notebook that it is not usually described as an “experiment”. Note that when 4-year-olds are involved, the same task calls for a quiet room, toys, and various clever ruses, and then everybody agrees that it is an experiment. Outsiders would surely be puzzled by the attitude that seeks to deny the psychological relevance of easy, robust results, while insisting on other, far more subtle measures, such as 20-ms differences in reaction times, or 1-s changes in how quickly babies get bored, or 2% changes in regional cerebral blood flow. The variability that one observes in native-speaker judgments is real, but very small relative both to the agreement among speakers and relative to the variability in other measures... (p. 61)<sup>38</sup>

For his part, Chomsky has likewise endorsed the view that native-speaker judgments serve as an important and fertile source of empirical evidence for linguistic theory, alongside others, including fine-grained behavioral and neurocognitive data from psycholinguistics (ch. III).

---

<sup>38</sup> See Sprouse and Almeida [forthcoming] for detailed studies of the consistency among native speakers.

In the real world of actual research on language, it would be fair to say, I think, that principles based on evidence derived from informant judgments have proved to be deeper and more revealing than those based in evidence from experiments on processing and the like, although the future may be different in this regard. If we accept—as I do...—[the] contention that the rules of the grammar enter into processing mechanisms, then evidence concerning production, recall, and language use in general can be expected (in principle) to have a bearing on the investigation of rules of grammar, on what is sometimes called “grammatical competence” or knowledge of language. (*Rules and Representations*, pp. 200-201)

There is, of course, a great deal more to say about the role of intuitions in linguistics.<sup>39</sup> My purpose here, however, is only to point out that the linguists’ heavy reliance on intuitions as a source of data neither renders their inquiry unempirical nor precludes them from recognizing the legitimacy of other types of data.

The Chomskyan conception of language, sketched above, serves as useful benchmark for the discussion ahead. It is dominant in both linguistics and philosophy, and informs much of the work on the psychological reality issue.<sup>40</sup> Nevertheless, it has seemed to many that Chomsky’s proposals regarding the nature of language are, in one way or another, unclear or mistaken. Others have argued that his views concerning the role of “mental representations” in linguistic theory is open to substantive challenge. In what follows, I outline a number of qualms that Chomsky’s critics have had with his methodological and ontological assumptions—several of which have already received some discussion in the pages above.

To begin with, there is a lack of clarity concerning a host of key notions in Chomsky’s theoretical framework. He holds that competent speaker/hearers have “tacit knowledge” of the generative procedure that “outputs” SDs—i.e., tacit knowledge of a grammar. But in what sense do adult language users *know*, *represent*, or “*cognize*” the grammar of their language? Such

---

<sup>39</sup> See, e.g., Devitt (2006: ch. 7), Culbertson and Gross (2009), Devitt (2010), Culbertson and Gross (forthcoming).

<sup>40</sup> See Devitt (2006: chs. 1 and 4) for a number of quotations from linguists, philosophers, and psychologists who seem to endorse the basic tenets of Chomsky’s ideology.

worries about the notion of tacit knowledge tend to go hand-in-hand with broadly philosophical concerns regarding Chomsky's persistent reliance on the notions of computation, information, and representation, illustrated in the quotations above. What is it, one might wonder, for an I-language—rather than a parsing mechanism—to perform “computations” defined over linguistic rules, principles, or categories? If the I-language is a static knowledge-structure or, even weaker, an “abstract condition on unknown mechanisms,” then why characterize the “generation” of linguistic expressions, or the “derivation” of their associated SDs, as a *computational* procedure? Finally, what conceptions of “information” and “representation” are at play in Chomsky's work?

Underlying all of these questions is the suspicion that there may well be a compelling alternative to viewing linguistic theorizing as an exclusively psychological enterprise. Perhaps, as many philosophers have suggested, it is better to see linguists as engaged in a systematic inquiry into the nature of what Chomsky calls “E-language.” On such a view, the “generation” of SDs by a grammar would not be computational, in the literal sense—a claim that accords with Chomsky's insistence that I-languages need not, and perhaps *ought* not, be seen as algorithms that the brain runs in real time. Perhaps the generation and derivation of SDs should be seen as being somehow *potential*, as when the axioms of a mathematical theory are said to “generate theorems,” though the axioms exist only as static ink-marks on a page.<sup>41</sup> Moreover, perhaps the notion of representation that we find at work in linguistics or psycholinguistics is, *pace* Chomsky, a *relational* notion. If so, then it is not at all clear that competent speakers mentally represent the syntactic rules and principles of their language. Indeed, it would be unclear *what* exactly they would be representing, the ontological status of rules and principles being, at present, an open topic in metaphysics.

---

<sup>41</sup> For two rich and engaging discussions of this analogy, see Levin (1979) and Soames (1984).

One starting point for addressing these issues is a published exchange between Chomsky and the philosopher, Georges Rey.<sup>42</sup> The questions that Rey poses for Chomsky are, in effect, the ones raised above—the very questions to which philosophers of mind have sought answers from the very outset of the generative linguistic enterprise. Importantly, Rey’s line of questioning is agnostic about which theory of representation, computation, information, or intentional content is correct, and which, if any, is applicable to the present case. Rey is also relatively neutral on whether Chomsky’s theory does indeed employ or require any of the notions in question. Finally, in contrast to a number of philosophers who have engaged with Chomsky,<sup>43</sup> Rey does *not* mount an outright attack; his essay constitutes more of an inquiry than a critique.

Chomsky’s argues that the notions Rey and other philosophers employ in characterizing his view—paradigmatically, the notions of “representation” and “intentional content”—are both degenerate and theoretically idle. According to him, philosophers have given no precise meaning to their talk of reference, representation, and intentionality, and insofar as such notions are at all clear, they are in no way relevant to the sort of empirical pursuit that linguists have undertaken in the past several decades. Chomsky also reiterates his contempt for the theoretical construct of an E-language, and insists once more that there is nothing more to the psychological reality of a syntactic rule or principle than its being mentioned in a true linguistic theory. The suggestion, it seems, is that while there is no coherent alternative to a cognitivist conception of linguistics, the notions that render it a psychological inquiry are not those that animate philosophical debates. Those who find it difficult to reconcile these claims with Chomsky’s persistent use of the intentional idiom may find it comforting to note that he takes these uses to

---

<sup>42</sup> See Rey (2003a), and Chomsky’s reply in the same volume. Rey (2003b) follows up, raising largely the same issues, though in a more refined manner.

<sup>43</sup> E.g., Gilbert Harman in Hook (1967), and W. V. Quine in Davidson and Hintikka (1975).

belong to merely “informal” presentations of his theory. In its purest form, he insists, the theory eschews intentional notions altogether.

As previously noted, an opposing view of the relation between a competent speaker and the grammar of her language can be found in Devitt (2006). In contrast to Chomsky’s position, Devitt’s alternative is accepted by a scattered few.<sup>44</sup> Nevertheless, many aspects of it seem to me to be both plausible and illuminating. Below, I give a compressed account of the view, and rehearse some of the reasons for accepting it over its Chomskyan rival.

#### **§4. Devitt’s “Linguistic Conception”**

##### *§4.1 Metaphysics*

Michael Devitt’s book, *Ignorance of Language* (2006), is a study of the galaxy of issues that jointly comprise the debate concerning the psychological reality of language. Devitt begins by criticizing Chomsky’s view that grammars are theories of a psychological faculty and, hence, are to be regarded as psychologically real to the extent that they true. Devitt takes this argument for the psychological reality of grammars to be “not only fast but dirty” (p. 9). He argues that grammars are *not* theories of the mind and that a “powerful psychological assumption” must be made in order to ascribe psychological reality to grammatical rules or principles. One such assumption might be what Devitt calls the Representational Thesis (RT).

**RT** “A speaker of a language stands in an unconscious or tacit propositional attitude to the rules or principles of the language, which are represented in her language faculty” (4).

According to Devitt, Chomsky and others may very well be relying on RT (perhaps tacitly) when they claim that principles of grammar are psychologically real. Consequently, Devitt devotes

---

<sup>44</sup> Scott Soames (1984), Kim Sterelny (1987, co-authored with Devitt), W. V. Quine (1953, 1970), Black and Chiat (1981), Gazdar, Pullum, Klein, and Sag (1985).

much of *Ignorance of Language* to arguing against RT. Chapter V of the present work examines the case for RT. At the moment, however, I would like to explore Devitt's conception of what grammars are theories *of*, if not psychological states or mechanisms.

In contrast to Chomsky's cognitivist conception, Devitt (2003, 2006, 2008) holds what he calls the Linguistic Conception, according to which token linguistic expressions are typically external to the skin of any competent speaker. When spoken, a token expression is an acoustic blast, scattered over time. When written, it is made of ink, lead, chalk and the like.<sup>45</sup> Token expressions have a definite location in space-time, and are concrete, in a quite ordinary sense: we can touch them, burn them, photograph them, and so on.<sup>46</sup> Devitt holds that stating the individuation and persistence conditions of linguistic expressions involves making ineliminable reference to the psychological states of competent speakers.<sup>47</sup> A token physical entity or process counts as belonging to a particular type of linguistic expression in virtue of four factors: (i) the

---

<sup>45</sup> Notice that all the parties to the debate over the Language of Thought Hypothesis (LOTH) are committed to the coherence of this proposition (though, of course, not necessarily its truth). For, they all agree that the debate about is whether straightforwardly physical events—i.e., the neurochemical vehicles of thought—are syntactically structured. Thus, in order to even make sense of the debate, one has to first take on board the view that straightforwardly datable occurrences in a physical medium can have syntactic structure. This gives rise to a dilemma: Either the LOTH debate is strictly incoherent—based on a confused metaphysical presupposition regarding syntactic properties—or it is unproblematic that datable physical events can have syntactic structure. Suppose, then, that an opponent of the Linguistic Conception denies that datable physical events can have syntactic structure. He thereby commits himself to the view that the LOTH debate is strictly incoherent. This will strike most—including myself—as a *reductio*. The LOTH debate cannot be dismissed *that* easily.

<sup>46</sup> In Chomsky's terminology, a set of such expressions constitutes an E-language. It is not clear, however, that Devitt would accept all of the theoretical baggage with which the expression 'E-language' has been loaded. (See his (2003), pp. 120-21.) I discuss the issue at some length in ch. II. It is also worth pointing out that Chomsky, himself, occasionally slips into treating linguistic expressions as concrete, physical events. In his paper, "Explanatory Models in Linguistics" (1962), he writes: "What [the child] accomplishes can fairly be described as theory construction of a non-trivial kind . . . a theory that predicts *the grammatical structure of each of an infinite class of potential physical events*" (p. 528, emphasis mine). Even if Chomsky did not intend, here or elsewhere, to claim that physical events are literally the bearers of grammatical structure, the remark illustrates how natural it is to think and speak of them in this way.

<sup>47</sup> The relevant states, however, need not be metalinguistic intuitions. Such intuitions are, of course, fairly reliable and relatively easy to elicit, given the right setting. But, according to Devitt, they are dispensable in principle. The psychological states that are crucial for the purpose of individuating linguistic expressions are *thoughts* and *intentions*.

communicative intentions of its author, (ii) the psychological state that it is used to convey, and (iii) its relations to other linguistic expressions.<sup>48</sup>

To further flesh out Devitt's view, it will be useful to organize the discussion around four issues concerning the metaphysics of language and the ontological commitments of linguistic theory. These issues concern (i) the individuation of linguistic expressions, (ii) the irreducibility of linguistic properties, (iii) the infinitude of natural language, and (iv) the status of phonologically null constituents ("empty categories"). I will refer to (i)-(iv) as "Significant Metaphysical Issues," to mark the fact that they concern the most hotly debated topics, where the claims that define Devitt's overall position come into sharpest relief and, unsurprisingly, receive the lion's share of his opponents' criticism.

The first Significant Metaphysical Issue has to do with of the individuation and persistence conditions of linguistic expressions. As noted above, Devitt concedes that these conditions are spelled out in terms of the psychological states of competent language users. He points out, however, that this does *not* warrant the identification of linguistic expressions with the psychological entities that Chomsky calls 'structural descriptions' (SDs). The identity and persistence of a linguistic expression surely *supervenes* on psychological states, but the same can be said of any number of things that are plainly *not* psychological entities, e.g., the United States government.<sup>49</sup> Furthermore, recognizing the supervenience relationship between linguistic expressions and psychological states does *not* warrant the methodological reductionism that is encapsulated in the oft-repeated slogan that "linguistics is a branch of psychology" (Chomsky, 1972: p. 1). Extending the analogy with governments, we note that political science is not a

---

<sup>48</sup> There are, of course, difficulties in making the individuation conditions precise in this way. I shall pass over these here, mentioning only that the account of convention in Lewis (1969/2002) may be a good starting point for those wishing to make some headway on this problem. See also Millikan (2004).

<sup>49</sup> See also Rey (2003b), p. 170-71, where essentially the same point emerges.

branch of psychology, though the two fields may be intimately related, and may well provide one another with valuable data. By the same token, biology is not, in any useful sense, a branch of quantum physics, despite the fact that all biological entities, mechanisms, and processes supervene on the activity of sub-atomic particles. Devitt (2006) puts the point as follows:

Even if symbols had their properties in virtue of certain mental facts that would not make the theory of those symbols about those facts and so would not make the theory part of psychology. Indeed, consider the consequences of supposing it would, and then generalizing: every theory—economic, psychological, biological, etc.—would be about physical facts and part of physics because physical facts ultimately determine everything. A special science does not lose its own domain because that domain supervenes on another. (p. 40)

This point is crucial, as it forestalls an all-too-common objection to the view that linguistic expressions are non-psychological, concrete denizens of space-time. Plainly, if there were no people, or no psychological states, then there would be no language. Therefore, the objection goes, language must be a psychological entity. Though certainly tempting, this line of reasoning misses the mark. Supervenience—a relation weaker than both identity and constitution—is all that is needed to account for the connection between psychological states and the persistence or identity of linguistic expressions.<sup>50</sup> If Devitt's view is correct on this point, then a reduction of linguistic expressions to psychological states is simply not in the offing.

The second Significant Metaphysical Issue has to do with another quick and dirty argument against Devitt's view that linguistic expressions are non-psychological entities. This argument appeals to the lack of success that phonologists have had in discovering robust acoustic commonalities between tokens of what is, intuitively, the same phone type. As is well known, there are no strictly acoustic criteria (i.e., necessary and sufficient conditions) for the identity of a phone. This is so even in the case of a single speaker, *even in a fixed acoustic environment*, let

---

<sup>50</sup> It is worth remarking, at this juncture, that no particular account of the supervenience relation is being assumed. To my knowledge, all of the going accounts are consistent with the claims made above.



alone for phones as they occur in a realistic community of speakers. Consequently, the structuralist dream of articulating “objective discovery procedures” for phonological categories has been abandoned. The situation is, moreover, not confined to phonology. Taxonomic linguistics seems to run into decisive difficulties at *every* level of analysis.<sup>51</sup> From this, it has been inferred that phones, morphemes, and phrases must be psychological entities.

Devitt’s Linguistic Conception points to a difficulty with this inference. Though the demand for purely acoustic individuation and persistence conditions cannot be met, the psychological criteria that do the necessary work do *not* substantiate the claim that linguistic expressions are, in the first instance, mental or psychological entities. Rather, phones, morphemes, and phrases are sets of concrete tokens that resemble one another primarily in respect of the psychological states by which they are caused, and the states that they cause in turn. This account lends no support to the *identification* of linguistic expressions with psychological entities. As noted above, such entities *supervene* upon distributions of psychological states, but they are not *identical* to such states.<sup>52</sup>

Devitt’s frank recognition of the fact that the categories of linguistic theory are not reducible to those of physics is just one of the features of his view that separates him from the traditional nominalism of structuralist linguists, such as Bloomfield.

[Jerry Katz] calls Chomsky’s view “conceptualism” and my sort of view “nominalism”. He takes nominalism to have been refuted by Chomsky’s criticisms

---

<sup>51</sup> Fodor, Bever, and Garrett devote two chapters of their classic 1974 text to a seemingly exhaustive survey of the arguments against the structuralist program of taxonomic linguistics.

<sup>52</sup> Black and Chiat (1981) take essentially the same line when they write: “Talking about the grammar or language as abstract systems is no more problematic than talking about society or the economy as abstract systems. The concepts of many social sciences are not reducible to physical reality, and yet they do not have to be interpreted as psychological objects either. We clearly would not want to say that concepts like ‘kinship’ or ‘circulation of money’, which cannot be correlated with particular physical entities, correspond to something ‘in the mind’. Nor does a psychological interpretation of grammar solve any philosophical problems that might arise with respect to its ontological status.” (p. 41)

of Bloomfieldian structuralism. Yet, so far as I can see, these criticisms are not of the *nominalism* of the structuralists but rather of their *taxonomic methodology*, a methodology in the spirit of positivism. According to Chomsky, this methodology imposed “arbitrary and unwarranted” limitations on linguistics: it insisted on defining “lower levels” before “higher levels”; it was inductive instead of explanatory (abductive); its epistemology was localist instead of Quinean holist. Indeed, despite the explicit nominalism of the structuralists, Chomsky is prepared to take the structuralists as implicitly concerned with the psychological reality of language and hence not really nominalist at all (Chomsky 1975c: 30–6). Yet he still thinks his methodological criticisms stand. In any case, Chomsky’s methodological criticisms can be and, in my view, should be embraced by the nominalist. In particular, we should not demand that the linguistic properties of tokens be reduced to “brute-physical” intrinsic properties of the tokens. The linguistic properties that concern us are “high-level” relational properties.

Thus, despite sharing with traditional nominalism a focus on physical tokens—inscriptions, acoustic blasts, bodily movements, and the like—Devitt’s Linguistic Conception is not intended to inherit Bloomfield’s conception of grammar as a taxonomic enterprise, nor the methodology of “discovery procedures” that rely on so-called “substitution criteria.” Rather, Devitt intends his ontological position to be compatible with the deeper *explanatory* goals of contemporary generative grammar, and with much (though perhaps not all) of the methodology that is characteristic of Chomskyan linguistics. The pursuit of these goals, he recognizes, requires linguistic theory to “abstract away... from a range of properties of [physical entities]—for example, form of script and pitch of sound—focusing simply on the syntactic properties that we are interested in.” And, as Devitt makes clear, such properties are typically *relational*, hence difficult to discern without a prior theoretical grip on the relations that constitute the domain of language.

[T]he linguistic structure of an utterance is not obvious and superficial. But this structural property is relational not intrinsic and relational properties are typically not obvious or superficial. Yet objects really have relational properties; for example, some objects really are paperweights, moons, echidnas, Australians, and so on. Sometimes, it is easy to tell that an object has a certain relational property because that property is well correlated with superficial properties. This makes it quite easy to tell an echidna, but not an Australian (if she keeps her mouth shut). And it makes it

quite easy to tell many English adverbs, the ones that end in ‘ly’. It can also be easy to tell that an object has a certain relational property if learning to identify the object involves learning to identify it as having that property. This makes it quite easy to identify the other English adverbs; identification comes with word recognition. One way or another, it is quite easy to tell the explicit structural properties of utterances although sometimes hard to tell the implicit ones. But utterances still really have both. (pp. 185-6)

The third Significant Metaphysical Issue has to do with the infinitude of natural language, which we discussed in §2. Various grounds have been offered for the claim that linguistic theory is concerned with an infinite domain. Chomsky has argued that an observationally adequate grammar for any natural language must be recursive. And, as noted above, we know of no principled bound on the length of natural-language sentences; one can always add ‘Jennifer knows that’ (or a suitable translation) to any well-formed declarative sentence, thereby generating a longer one. Moreover, some theorists hold that there exist sentences that will never be uttered. How then can linguistics be concerned with physical tokens, of which there is only a finite number? Devitt (2006) attributes this sort of argument to Katz, whose abstractist position we have seen reason to reject.

Katz has another objection to nominalism: grammars are about an infinite number of sentences but there cannot be an infinite number of tokens. If there were a problem for my sort of nominalism it would lie in its apparent commitment to nonactual possible sentences, a problem that would arise even if we were dealing with a finite language (e.g. English with a limit of one million words to a sentence). The only significance of any apparent commitment to an infinite number of sentences is that it would guarantee that some were nonactual. But talk of there being nonactual possible outputs of a competence can be a mere manner of speaking (notes 6 and 8). So too can talk of there being an infinite number of such outputs. The truth behind the talk of the nonactual can be simply that the grammar is lawlike. [Note 8: So if something were a sentence, a *wh*-question, a passive, or whatever, it would have the properties specified for such items by the theory.] And the truth behind the talk of the infinite can be simply that there is no limit to the number of different sentence tokens that might be governed by the rules the grammar describes. (p. 26)

As noted in §2, Devitt’s response to Katz is that the infinitude of language is not a *datum*.

Indeed, it’s not even *true* that there are, in reality, infinitely many linguistic tokens or types (nor

that there have ever been, nor that there will ever be). Rather, the infinitude is an artifact of theory construction—specifically, a reflection of the fact that we seek maximally wide generalizations, in our effort to accommodate *any* potential datum that we might encounter. Here, Devitt’s view coincides with that of Chomsky and his followers. For instance, Collins (2009) writes:

Controversy ... arise[s] from the common talk in linguistics and philosophy of natural languages being infinite. The infinity of English, as it might be, however, is not a phenomenon. The phenomenon is that particular organisms are continuously novel in their speech and understanding, which we theorize in terms of unbounded generation. (fn. 4).

To clarify, the phenomena that the linguist seeks to describe, predict, and explain include not only *past* cognitive and verbal episodes—acts of production, comprehension, and perhaps metalinguistic judgment—but also *future* ones. And while the number of these—both past and future—is finite, there is no way of telling in advance *which* particular such episodes we will come to encounter. What we know from our observations thus far, however, is that such episodes have been *continually novel*, and we have no reason to expect that they will cease to exhibit that pattern. Moreover, in the long run, advances in the biochemical and bioinformatic sciences may increase the human life span, alter our motivational states, and enlarge our memory resources. As a result, the sentences that we produce and understand may become not only continually novel (as they have been to date), but also significantly *longer*. And just as there is no telling in advance what the limits are on the linguistic creativity that gives rise, continually, to novel sentences, there is likewise no telling in advance just how long the longest sentence a human will ever utter might be. This, it seems, is the underlying justification for the view that sentence length is unbounded and that natural language consists of denumerably many sentences.

Let us turn now to the fourth Significant Metaphysical Issue, which pertains to the ontological status of so-called “empty categories”—i.e., unvoiced (“phonologically null”) entities that are posited by syntactic theories in the Principles and Parameters (P&P) tradition.<sup>53</sup> We will explore these entities in more detail in ch. III, but, for the moment, we can summarize the motivations that lead syntacticians to posit them as follows: Empty categories allow a grammar to encode various semantically relevant structural relationships between the constituents of a phrase. Let us distinguish four types of empty category. (Following convention, I use the labels ‘trace’ and ‘PRO’.)

- (3) Igor loves<sub>i</sub> Alia and she *verb-trace<sub>i</sub>* him.
- (4) [What instrument]<sub>i</sub> do you play *wh-trace<sub>i</sub>*?
- (5) [<sub>NP</sub> The boat]<sub>i</sub> was carried *NP-trace<sub>i</sub>* by five people.
- (6) Jeffrey<sub>i</sub> tried PRO<sub>i</sub> to record a song.

In virtue of its being co-indexed with the verb ‘loves’, the verb trace in (3) encodes the fact that the referent of ‘she’ loves the referent of ‘he’, rather than bearing some other attitude toward him. The traces in (4) and (5) mark the fact that the phrases with which they are co-indexed are the objects of the verbs ‘play’ and ‘carried’, respectively. The empty category, PRO, in sentence (6) signals that the co-indexed NP, ‘Jeffrey’, is the subject of the infinitival verb ‘to record’.

While the notion of an empty category helps the syntactician capture significant structural relationships and semantic regularities, it creates trouble for both the psycholinguist and the metaphysician. Postponing our discussion of psycholinguistic matters until ch. III, let us focus presently on the metaphysics.

On a standard view in linguistics, which derives from Chomsky’s cognitive conception of

---

<sup>53</sup> In reality, empty categories were introduced some time before the P&P model was fully articulated. Note also that traces are posited only by grammars that rely on movement operations—e.g., the Government and Binding theory. Other types of empty category (e.g., PRO) do not, however, encode the effects of movement. I will not discuss here whether Devitt’s ontology can accommodate the sorts of entities that are posited by formalisms such as LFG and HPSG, which generate syntactic structure without recourse to movement operations.

language, empty categories are a kind of *psychological* construct. The I-language, viewed as a psychological entity, is home to various types of language-specific mental representations, many (perhaps all) of which do not “correspond” to anything in the physical environment. For instance, the I-language contains phonological representations that, as we saw above, do not correspond in any straightforward way to the purely acoustic properties of any actual speech stream. Moreover, contemporary syntactic theory views phrases as having a hierarchical structure that encodes (at a minimum) the relations between the constituents of a phrase. Such relations are not overtly marked in the linear order of a speech stream or an inscription. Proponents of the cognitive conception infer that these relations do not hold between items in the physical environment. And if the relations are not, in this sense, “out in the world,” then they must be in the mind. Less poetically, the conclusion that is often drawn is that a person’s I-language generates mental representations of hierarchical structure. In the course of language comprehension, these representations are imposed on a speech stream or an inscription. This explains how and why these physical entities are perceived by competent speakers as having such properties, despite not actually having them.

Against this background, the posit of empty categories in the syntactic structure of a phrase is not at all puzzling. Empty categories are just another aspect of the mental representations that ultimately constitute language. And because they are on a par with phonological and syntactic structures, it is no surprise that they are not overtly present in the physical form of speech or writing. For, on the Chomskyan view, no interesting linguistic properties *at all* are to be found there. Collins (2008b) sees this as a profound difficulty for Devitt’s view. He writes: “*Prima facie*, perhaps the most serious problem facing Devitt is how the abstractness of syntax might be accommodated in his model of linguistic reality” (p. 20).

In response, Devitt rejects the Chomskyan picture of both language and psychology. He claims that the abovementioned phonological and syntactic properties *are* out in the world, in the sense that they are real properties of the inscriptions and acoustic blasts that, according to him, constitute language. Moreover, he denies that language comprehension requires the deployment of mental representations—metalinguistic descriptions, as it were—in response to linguistic stimuli. Indeed, his break with the cognitive conception is more radical still. For, on the strength of his arguments against the Representational Thesis and the psychological reality of syntax, he denies that there are any such mental representations for a competent speaker to deploy. As we saw above, however, Devitt does recognize that it is impossible to reduce phonological and syntactic properties to the brute-physical features of inscriptions and acoustic blasts. He views them instead as “high-level” relational properties.

[N]o naïve brute-physical account of the relation between sounds and phonemes is possible. Phonology shows that there are many complicated ways in which sounds can instantiate a phoneme, including relations to other sounds; and that a sound may be able to instantiate more than one phoneme. Similarly, there are many complicated ways in which inscriptions can instantiate a letter; and so on for other linguistic media. But this does not show that the sounds, inscriptions, etc. do not instantiate SLEs [Standard Linguistic Entities]. Quite the contrary. The property of being Australian is instantiated by a vast variety of physical forms; for example, the forms of the capitalist Rupert Murdoch, the runner Cathy Freeman, the horse Phar Lap, the city of Sydney, a bottle of Penfolds Grange, and the many forms of the saying “No worries, mate”. The property of being the word ‘cat’ is instantiated by a much smaller variety of physical forms, a variety of sounds, inscriptions in many different fonts and handwritings, and so on. Just as all the former instantiations really are Australian, all the latter really are the word ‘cat’. And note that just as some things do not count as Australian, some things do not count as the word ‘cat’. (p. 186)

With these remarks in a mind, a metaphysician who is puzzled about the phonological and structural properties of phrases may well begin to see, if only in dim outline, how Devitt’s position can stand as a coherent alternative to the Chomskyan view outlined above. Focusing specifically on the case of syntactic properties, she may take comfort in Devitt’s gesture toward

the metaphysical position that was a hallmark of the structuralist tradition.

The outputs of a linguistic competence, physical sentence tokens, are governed by a system of rules.... Something counts as a sentence only if it has a place in the linguistic structure defined by these structure rules. Something counts as a particular sentence, has its particular syntactic structure, in virtue of the particular structure rules that govern it, in virtue of its particular place in the linguistic structure. (p. 24)

Nevertheless, Devitt's picture seems to run into trouble with regard to empty categories. For, in the context of contemporary linguistic theory, an empty category—i.e., an unvoiced (“phonologically null”) constituent of a sentence—is most naturally thought of as a strange kind of *object*, not a high-level relational *property*. For this reason, making clear Devitt's treatment of empty categories is pivotal for the project of defending his overall metaphysical position. It is here that Devitt's appeal to public linguistic conventions is brought into sharpest relief.

The problem [for my view] is thought to be particularly pressing in the case of “non-overt” or “unvoiced” constituents of a sentence. How is it possible for a sentence to have such a constituent? I responded: “The simple answer is: *there is nothing to stop there being a convention of this sort*” (2006b: 599). Consider the string ‘Bob tried to swim’. The idea is, roughly, that each word in the string has a syntactic property by convention (e.g. ‘Bob’ is a noun). Put the words with those syntactic properties together in that order and the whole has certain further syntactic properties largely by convention; these further properties “emerge” by convention from the combination. The most familiar of these properties is that the string is a sentence. A more striking discovery is that it has a “PRO” after the main verb even though PRO has no acoustic realization. There is no mystery here. (Devitt, unpublished)

Thus, on Devitt's view, just as *being a sentence* and *having ‘Bob’ as the first noun* are properties of some physical tokens, so is the property of *having PRO as a constituent c-commanded by the NP*. These high-level relational properties are, metaphysically, all on a par.

Two problems can be raised for this response. The first, due to Collins (2008b) is that it is difficult to see how public-language conventions could “get a grip” on such subtle and fine-grained properties of linguistic expressions. Devitt (2008b) responds:

I don't know. But I don't need to know to sustain linguistic realism. I have shown that it is plausible that a whole lot of sounds and inscriptions that humans produce form representational systems. Those systems are not fully innate and so must be



partly conventional. I have shown how it is *possible* for conventions to yield unvoiced elements. I have indicated in a general way, referring to David Lewis (1969), how linguist conventions, like other conventions (that are not stipulated), arise from regularities together with some sort of “mutual understanding.” The regularities for linguistic conventions are in speaker meanings (pp. 156, 179-80). It would be nice to go much further, giving full explanations of the forming of linguistic conventions—indeed, of the forming of *any* conventions—but the hypothesis that there are such conventions does not depend on giving these. Lewis begins his book by claiming that it is a “platitude that language is ruled by convention” (1969: 1). This is surely right. ... Of course, if it could be shown that the required conventions were *impossible* then my proposal would be in trouble. But giving such an “impossibility proof” would be a mighty tall order. After all, the phenomena that lead linguists to theorize that expressions have certain structures must be phenomena that speakers *could be* sensitive to in forming a language.

Collins will no doubt be left unsatisfied by this reply. Many of his remarks in other publications make it clear that he finds it hard to stomach the lack of detail in Devitt’s characterization of public language conventions. This, however, is not a damning criticism, for two reasons. First, as Devitt points out, Lewis (1969/2002) has supplied a great deal of relevant details. Working out the applications of Lewis’s general picture to the specific case of *syntactic* conventions is a research project whose eventual success we have been given no reason to doubt. Second, Collins’ alternative account of empty categories relies on a fully internalist notion of mental representation, as well as an account of language acquisition in the absence of conventions. It is no stretch, I think, to say that these accounts are presently underdeveloped and in no better shape than contemporary versions of a Lewisian theory of convention.

The second and, in my view, more difficult problem for Devitt’s position emerges when we examine his strategy for dealing with empty categories. The problem, recall, was contemporary linguistic theory treats these as a strange kind of *object*, not a high-level relational *property*. We saw above that Devitt’s reply to this worry consists in recasting claims like (7) as having the form of claims like (8).

- (7) In the sentence ‘David wants to write’, there is an unvoiced entity, PRO, which is c-commanded by the initial NP.
- (8) The sentence ‘David wants to write’ has the *property* of having an unvoiced entity, PRO, as a constituent which is c-commanded by the initial NP.

Devitt’s strategy seems to be one of treating apparent quantification over empty categories (“*there is an unvoiced entity...*”) as garden-variety property attributions (“*has the property of having an unvoiced entity...*”). But there is reason to doubt that Devitt can invoke this strategy, given his leanings toward nominalism.

Devitt (2006) is up-front about the fact that his talk of properties will eventually have to be paraphrased away. He does not, however, specify how exactly the relevant paraphrases would go, relying instead on the results of the quite general paraphrase strategies that nominalists have developed for theoretical discourse in other domains. He writes:

Is my contemplated task appropriately characterized as nominalistic? It takes all the objects that linguistics is about to be concrete tokens, and so to that extent it is nominalistic. Where it stands ultimately on the nominalism issue depends, of course, on what we make of its ascription of meaning *properties* to those objects. However, it seems unlikely that the nominalist would have any *special* difficulty paraphrasing away this property talk. My contemplated task for linguistics is likely to be as nominalistic as tasks in physics, biology, or economics. (p. 30)

Devitt’s focus in this passage is the ascription of “meaning properties,” but the problem of syntactic properties is equally pressing. For, it is not at all clear how one might regiment claims that superficially appear to make reference to an empty-category property—e.g., the property of *having PRO as a constituent c-commanded by the NP*—without employing statements that themselves quantify over empty categories.<sup>54</sup> And if Devitt’s preferred regimentation *does* take

---

<sup>54</sup> There is, of course, a general nominalistic strategy of paraphrasing away apparent references to properties. It involves treating claims of the form ‘*x* has *F*-hood’ as saying no more than that *x* is *F*. In this case, however, the strategy has the effect of treating claims like (8) as saying no more than what (7) says. This is a problem for the nominalist, because (7) makes reference to empty categories. The whole point of moving to (8) was to avoid such

that form, then it seems that his view must, after all, be supplemented with a characterization of these strange, seemingly non-physical entities.

While Chomsky and his followers (e.g., Collins) may see this as a decisive reason to reject Devitt's Linguistic Conception, I myself regard it as no more than outstanding problem in need of a technical solution.<sup>55</sup> Nominalists have always faced issues having to do with regimentation, and no shortage of clever and convincing solutions have been proposed.<sup>56</sup> So, while I do not agree with Devitt's claim that the problems facing the nominalist are the same in linguistics as they are in other domains, I do not see this as cause for despair.

#### §4.2 *Distinguishing Competence from its Products, and Linguistics from Psycholinguistics*

Having set out some of the metaphysical commitments that underlie his position, Devitt (2006) goes on to draw a number of distinctions, which, when applied to objects of our concern, elucidate the relationship between language and the mind, and, correlatively, between linguistics and psychology. For present purposes, the most important of these is the distinction between a competence and its products. Devitt illustrates this distinction with a simple and intuitive example. We are asked to consider a blacksmith who has the competence to produce horseshoes. This competence is, without a doubt, a psychological entity, consisting of myriad beliefs, preferences, action scripts, and so forth. Exercising this competence, the blacksmith creates a horseshoe, which, in contrast to his competence, is decidedly *not* a psychological entity. Of

---

reference. While perhaps not as worrying as apparent reference to numbers ('There is a number between 1 and 3'), or the ascription of properties to properties ('Humility is a virtue'), this should, I think, trouble the nominalist.

<sup>55</sup> One possibility, of course, is to simply reject grammars that posit empty categories, in favor of ones that deal with phenomena like movement and ellipsis in a different way. (See fn. 53 above.) However, this would be to let ontological issues guide theory-choice in linguistics. Though I am quite sympathetic to Devitt's nominalism, I nevertheless hold firm to the methodology of section §2.2, which counsels us to be suspicious of such maneuvers.

<sup>56</sup> See Goodman and Quine (1947), Sellars (1963c).

course, the fact that the blacksmith's creation is a *horseshoe*, rather than, say, a ceiling decoration or a doorstep, may very well supervene on the psychological states of the blacksmith or of those who go on to use the object for their own purposes. But, for reasons now familiar to us, this does *not* make the horseshoe a psychological entity.

Devitt appeals to the distinction between a competence and its products in his argument against the Chomskyan identification of a language with a speaker's competence. Just as horseshoes are the outputs of the blacksmith's competence, linguistic expressions are the outputs of a linguistic competence. The competence to produce and comprehend language, like the competence to make horseshoes, is a psychological entity. Hence, a theory of either sort of competence is *ipso facto* a psychological theory. But, on Devitt's view, one needs to supplement that theory with a *distinct* theory of the *products* of competence—whether those products are horseshoes or linguistic expressions. One must, that is, provide some relatively independent characterization of the domain that the competence engages with.

Devitt argues that the working syntactician is attempting to construct precisely such a characterization when she formulates grammatical principles, e.g., those constitutive of Binding Theory, Case Theory, empty category licensing, and the like. Other branches of linguistics attempt to specify the phonological and the semantic properties of linguistic expressions. From the perspective of the metaphysician, linguistic expressions have all of these properties in virtue of their interactions with a variety of psychological states—the now-familiar supervenience claim. But, from the perspective of the psychologist, these properties may nevertheless have to be “recovered,” “extracted,” or simply *recognized* by the hearer in the course of comprehension. Indeed, it is only in this way that we can make sense of a hearer's *misperceiving*, or being *mistaken* about the structure of a sentence in the course of comprehension. Such cases of

misperception are, of course, all too common in daily life, and often serve as crucial data in psycholinguistic experiments, e.g., those having to do with garden-path sentences (ch. III).

Devitt's conception of the relation between the grammar of a language and the competence to produce and comprehend linguistic expressions is not uncommon. Take, for instance, a classic text in psycholinguistics, *The Psychology of Language* (1974), by J. A. Fodor, T. Bever, and M. Garrett. In criticizing a proposal concerning the mechanisms of language production, Fodor *et al.* write: "We see here a classic example of the problems generated when one fails to distinguish theories of language structure and theories of the computational mechanisms involved in the perception and production of speech" (p. 67). The distinction to which Fodor *et al.* are drawing attention is between what Devitt calls the "structure rules" of a particular language—primarily, the syntactic rules and principles governing the expressions in the language—and the "processing rules" that govern the psychological mechanisms responsible for the perception and production of such expressions. Phillips (1996) notes that this distinction has been implicit in much of the work done in the generative tradition.

Since the 1960's, work in linguistic theory has focused on characterizing linguistic knowledge in terms of *static* mental representations, and on accounting primarily for grammaticality judgments. The main goal of psycholinguistics, on the other hand, has more commonly been to provide a more or less explicitly *procedural* characterization of how speakers perform linguistic tasks such as comprehension or production. It is standardly assumed that mental systems for grammar and processing are separate, and hence that linguists and psycholinguists are not studying the same thing. (p. 3)<sup>57</sup>

---

<sup>57</sup> Though he ultimately rejects this distinction on principled grounds (ch. V), Phillips does provide numerous references to other theorists who endorse it, including notable quotations from Crocker (1996), who writes: "This *grammar as parser* approach is not a rational position given the competence–performance division [...] which clearly separates the declarative properties of the syntactic theory from any procedural notions" (p. 49), and "The suggestion that the grammar *is* the parser is simply not well-conceived" (p. 51). (From Phillips [1994], p. 16.) A standard textbook in computational linguistics, Jurafsky and Martin (2007) expresses the distinction as follows: "Syntactic parsing ... is the task of recognizing a sentence and assigning a syntactic structure to it. This chapter focuses on the kind of structures assigned by context-free grammars ... [S]ince they are based on a purely declarative formalism, context-free grammars don't specify *how* the parse tree for a given sentence should be computed. We'll therefore need to specify algorithms that employ these grammars to produce trees" (p. 431)

Following Chomsky, Phillips places *both* projects in the realm of psychological research. The formal linguists' project is conceived of as a study of the mind, on account of the fact that its domain is a set of "static mental representations."<sup>58</sup> Devitt has provided reasons for being skeptical of this way of construing the work of syntacticians and phonologists. The present point, however, is that, even when viewed in this way, there is *still* reason to distinguish linguistics from psycholinguistics. The domains of the two fields—structure rules and processing rules (*alt.* the grammar and the parser)—are simply not identical. This point holds regardless of whether one takes the structure rules to be psychologically realized as an I-language, or simply a set of rules devised by linguists to capture generalizations concerning what Chomsky might call E-language expressions.

## §5 *Summary and Conclusion*

In the pages above, we surveyed the main features of two competing approaches to linguistic metatheory—Devitt's nominalism and Chomsky's cognitivism. We took note of their metaphysical commitments as well the rich resources that each of them brings to the task of answering foundational questions about language and linguistic theorizing. We also saw that both approaches face non-trivial difficulties. Chomsky's view awaits the articulation of a non-relational theory of representation, or an alternative conception of the relation between grammars and neurophysiological phenomena. Devitt's view requires a more detailed account of public

---

<sup>58</sup> Phillips goes on to develop a proposal according to which the structure rules of a language—i.e. rules of the grammar—do the work traditionally carved out for processing rules. Phillips' view of the matter is certainly interesting, but his proposal rests heavily on a controversial theory of syntax. There is some question as to whether that theory can account for all of the behavioral data that informs psycholinguistic models of parsing and comprehension, while satisfying the constraints that are relevant to pure syntactic theorizing. I cannot explore the matter here.

language conventions and a satisfactory strategy for paraphrasing away apparent references to unvoiced elements of syntactic structure.

Of course, much more can be said—and, indeed, *has* been said—to clarify and defend both approaches. In the sections that follow, I will highlight some of the major disputes between them. For the purposes of the discussion, however, it will be important to note a particularly significant contrast between them. Chomsky’s view, unlike Devitt’s, has the effect of closing off discussion about whether syntactic rules and principles are psychologically real. For, if one begins with Chomsky’s cognitive conception of grammars, then the question of whether a syntactic rule or principle is psychologically real reduces to the question of whether we have good reason to believe the grammar that posits it. And, of course, we typically *do* have such reasons, for it is the working syntactician’s business to provide *prima facie* reasons of just this sort. By contrast, if one begins with Devitt’s linguistic conception, then the ascription of psychological reality to one or another syntactic principle requires a powerful *psychological* assumption—e.g., the Representational Thesis (RT).

In ch. V, I examine the case for RT and for other ways of seeing grammars as being psychologically real. Plainly, this project presupposes that the psychological reality issue is not settled in advance by appeal to Chomsky’s cognitive conception. For this reason, in the next chapter, I will undertake a critical examination of two well-known arguments for the cognitive conception. First, I will cast doubt on an argument due to Collins (2008), according to which only the cognitive conception can rationalize the pursuit of what Chomsky has called “explanatory adequacy.” Second, I’ll tackle issues surrounding the notion of E-language, arguing that E-languages do, in fact, play a role in acquisition theory and that the identity conditions on E-languages are no less secure than those on I-languages. I want to stress that my

goal will *not* to decisively refute Chomsky's cognitive conception, *nor* to establish the superiority of Devitt's "linguistic realism". I aim for nothing more than a Scotch verdict. To substantiate undertaking a sustained inquiry into the psychological reality issue, all that must be established is that the cognitive conception is not "the only game in town."



## II

# THE COGNITIVE CONCEPTION AND ITS DISCONTENTS

---

### Introduction

In chapter I, I argued that abstractism—considered as a package of views concerning the ontology and methodology of linguistics—offers us a radically mistaken picture of the field, of its aspirations, and of language itself. I went on to examine two alternative views—the dominant cognitivist position of Chomsky (1965, 1995), and the sophisticated nominalist rival developed by Devitt (2006, 2008).<sup>1</sup> Having sketched the main lines of the disagreement between Chomsky and Devitt, I proceed in this chapter to defuse two common arguments for cognitivism.

In §1, I examine a key aspect of John Collins’ recent reconstruction of the revolution in linguistics, which he calls “the cognitive turn” (Collins, 2008). My aim is to show that one of his main arguments for taking the cognitive turn is simply not compelling. While Collins makes a very convincing case for aiming to achieve explanatory adequacy in syntactic theory—over and above mere observational and descriptive adequacy—he does not succeed in showing that this requires subscribing to a cognitivist conception of language. Although the pursuit of explanatory adequacy is typically motivated by appeal to considerations regarding language acquisition, I point out that it can just as well be motivated *without* any appeal to those considerations. Explanatory adequacy—i.e., the maximally general, simple, and unified theoretical coverage of all human languages—is desirable *whether or not* the cognitivist conception is correct. Of course, this is not, by itself, a refutation of the cognitivist conception. Nor would I wish to

---

<sup>1</sup> As noted in ch. I, there are several names for the view that Chomsky has developed. Traditionally, the view goes under the label ‘conceptualism’. However, in light of the role that the term ‘concept’ has taken on in cognitive science, this usage is almost certainly misleading. I will use the more suitable label, ‘cognitivism’. It’s also worth pointing out that Devitt may not be entirely comfortable being pigeon-holed as a nominalist or a structuralist. I nevertheless avoid using the more accurate but overly cumbersome label ‘quasi-nominalism’.

suggest that acquisition theory has not contributed a great deal to our understanding of both language and the mind. Rather, my intent in §1 is only to show that one of the standard arguments for the Cognitive Turn does not withstand scrutiny.

Devitt's Linguistic Conception commits him to the reality of public languages—something akin to what Chomsky has called E-language. Devitt appeals, in this connection, to the notion of a convention, which has gained currency in philosophy after the pathbreaking work of Lewis (1969/2002).<sup>2</sup> In §2, I address some common concerns regarding the theoretical utility of the notion of E-language, as well as the difficulties that arise in giving identity conditions for E-languages. First, I suggest that E-languages play a useful role in theories of language acquisition. Second, I bolster Devitt's view that E-language plays a role in formal grammar construction in the straightforward sense that *that is what syntacticians are actually studying*. In later sections of the chapter, I argue that the identity conditions on E-languages are no less secure than the identity conditions on I-languages, contrary to what is often assumed. Along the way, I discuss some reasons to be skeptical of the arguments that Barber (2004), Collins (2008), and Ludlow (2011) give in support of the claim that I-languages, rather than E-languages, are the objects of linguistic inquiry.

Before proceeding to the discussion, I would like to say a word about how the topics covered below fit into the broader goals of the present work. In chs. III-V, I plan to argue that some rules or principles of syntax are psychologically real. However, as both Chomsky and Devitt point out, if one begins with the cognitive conception of grammars, then the question of whether a syntactic rule or principle is psychologically real reduces to the question of whether

---

<sup>2</sup> More recently, in a self-conscious effort to undermine Chomsky's skepticism about the coherence or theoretical utility of the notion 'public language', Millikan (2004) has taken steps to develop an account of public-language conventions, making use of the resources afforded by her well known teleological framework. Devitt's view of such conventions has been criticized by Cain (2010). I will not take a stance here on whether Millikan's or Cain's arguments succeed.

we have grounds to believe the grammar that posits it. And given that contemporary syntactic theories enjoy an abundance of empirical support, we typically *do* have such grounds. So, if Chomsky is right, then the issue of psychological reality has already been settled. By contrast, if one begins with Devitt's Linguistic Conception of grammars, then the ascription of psychological reality to one or another syntactic principle requires, in addition, a powerful *psychological* assumption—e.g., what Devitt calls the Representational Thesis.

Given that my overarching goal in the chapters to come is to defend a substantive thesis concerning the psychological reality of grammar, my project presupposes that the issue has not been settled in advance by appeal to the cognitive conception. For this reason, I shall be casting doubts on the arguments for Chomsky's position. I want to stress, however, that my goal is *not* to refute the cognitive conception, nor to establish that Devitt's alternative is the clearly superior of the two positions. As I see it, in order to substantiate undertaking a sustained inquiry into the psychological reality issue, all that must be established is that the cognitive conception is not, as Jerry Fodor puts it, "the only game in town."

## **§1. Does Pursuit of Explanatory Adequacy Require the Cognitive Conception?**

### *§1.1 Linguistic Universals and The Explanatory Adequacy Condition*

We begin by reviewing some concepts that are now commonplace in linguistic meta-theory. A grammar, considered as a theory of a specific natural language, must meet three distinct adequacy conditions. In order of increasing strength, these are: *observational*, *descriptive*, and *explanatory* adequacy. Observational adequacy consists in the grammar's merely generating each of the surface strings of a language, irrespective of the structures that it

assigns to them.<sup>3</sup> Descriptive adequacy, by contrast, requires the grammar to generate the correct *syntactic structure*—or, in cases of ambiguity, multiple correct structures—for each string. These conditions are typically thought of as “external,” in the sense that they deal with the interface between linguistic theory and empirical data. There is a question about whether descriptive adequacy is really best characterized as an external condition, in this sense. Let us pause to clarify this issue, with an eye toward developing some claims that will assume prominence later in the discussion.

When two distinct grammars,  $G_1$  and  $G_2$  are observationally adequate, it is often the case that one of them—say,  $G_1$ —achieves observational adequacy by positing vastly more rules than the other,  $G_2$ , or by offering distinct explanations of phenomena that  $G_2$  treats in a unified manner (Larson, 2010). In such cases, the choice between  $G_1$  and  $G_2$  is *not* made by appeal to some data that  $G_2$ , but not  $G_1$ , can predict or explain; *ex hypothesi*, no such data exist. Rather, the choice is made on the grounds of what philosophers of science sometimes call “super-empirical virtues,” which include (*inter alia*) formal simplicity, generality of coverage, and explanatory unity (Churchland, 1985/1989).

---

<sup>3</sup> There are important ambiguities in the term ‘surface string’, which bear on issues concerning the coherence and utility of the notion of E-language. On the one hand, a surface string might be a linear ordering of words, morphemes, or phonemes. These categories are picked out by technical notions from linguistic theory and hence might be argued by a Chomskyan theorist to be I-linguistic—that is, psychological—entities. On the other hand, a surface string might be a linear ordering of inscriptions, acoustic waveforms, or muscle contractions. In this case, surface strings are plainly E-linguistic entities. But it is difficult to make sense of the idea that a grammar might *generate* such things. As I see it, the friend of E-language must argue (i) that the notions ‘word’, ‘morpheme’, and ‘phoneme’ actually have their home in E-linguistics and (ii) that an adequate grammar will include some sort of systematic mapping from these theoretical constructs to the observational categories that include inscriptions, acoustic waveforms, and muscle contractions. A formal statement of such a mapping would license talk of a grammar generating observational E-linguistic entities. (Analogously, the cognitivist must specify *psychological* mechanisms that compute phonological, morphological, and syntactic representations on the basis of causal encounters with inscriptions, acoustic waveforms, and muscle contractions.) Note, finally, that even when this ambiguity in ‘surface string’ is resolved, there remains a further issue about whether an observationally adequate grammar must generate only those strings that have actually been observed—e.g., recorded in a corpus—or whether it must also generate strings that either *will* or *might* be observed. See Quine (1953) for reflections on this and related matters.

It is on *these* grounds that Chomsky (1957) argued against pure phrase-structure grammars—what Fodor, Bever, and Garrett (1974) called “Immediate Constituency (IC) grammars”—and in favor of transformational grammar. Similar considerations played a significant role in motivating the shift away from early transformational grammar (the Standard and Extended Standard Theory) toward the Principles and Parameters (P&P) framework, which dispensed with the growing stock of language-specific and construction-specific transformations (ch. V, §5). If pursued further, the early transformational grammars might have turned out to be observationally adequate. But the P&P framework was superior precisely in that it posited fewer formal principles and unified the explanation of phenomena that, from the earlier perspective, seemed heterogeneous. These advantages of P&P grammars are, of course, quite real and significant, but they are, in the end, “theory-internal,” having little to do with specific data that went unexplained by earlier theories. Importantly, they also have little or nothing to do with the psychological status or implications of formal grammars. I develop this point in §1.2.

Characterizing the third adequacy condition—i.e., explanatory adequacy—is a difficult matter, for, as we will see below, it raises a number of issues surrounding the very motivation for a cognitivist conception of language. One mainstay among the battery of arguments in favor of the cognitivist conception relies on the claim that only *this* conception can supply us with principled grounds for aiming at explanatory adequacy. Indeed, it is sometimes suggested that only the cognitivist is able to so much as draw a principled distinction between the “external” adequacy conditions—observational and descriptive adequacy—and the “internal” condition of explanatory adequacy. Let us take a moment to briefly review how it accomplishes this.

On Chomsky’s view, the initial state of the language faculty encodes a Universal Grammar (UG)—roughly, a kind of “template” for the grammars of specific languages, such as

French, Russian, and Hindi. As such, it encodes the structural properties that are *universal* among human languages—i.e., the features common to all of them.<sup>4</sup> Furthermore, Chomsky takes the initial state of the language faculty to be centrally involved in language acquisition, in the sense that it provides the child with the innate resources necessary for structuring and organizing linguistic experience.

In his early writings, Chomsky conceived of the child's innate endowment as having three components: (i) a specification of all possible human grammars, (ii) a simplicity metric, along which those grammars are ranked with respect to observational data, and (iii) a device or a computational mechanism for taking in the observational data and ranking the grammars along that metric, eventually ranking highest what is sometimes called the “target grammar”—the grammar of the “ambient language” in the child's speech community.<sup>5</sup>

Chomsky's view of the matter has evolved considerably since the publication of *Aspects of the Theory of Syntax* (1965). In his *Lectures on Government and Binding* (1981) and *Knowledge of Language* (1986), Chomsky introduced and elaborated the Principles and Parameters (P&P) model, according to which the innate endowment differs, in important respects, from what he took it to be in his earlier work. The P&P model has it that the structure of any possible human language is determined by a handful of core syntactic principles, each of

---

<sup>4</sup> The term ‘feature’ is to be understood in its generic sense, though there is a way in which my claim is true even if the term is taken to have a meaning that is familiar from contemporary theories of phonology and syntax (e.g., Minimalism and HPSG; see ch. V, §6).

<sup>5</sup> As noted in ch. I, the terms ‘target grammar’ and “ambient language” carry the implication that there is a language external to the child at the time of acquisition, whose grammar the child is struggling to grasp. In §2, I lean on this common and seemingly indispensable usage in raising doubts about Chomsky's claim that E-language plays no role in linguistic theory.

which is parameterized.<sup>6</sup> Language acquisition, on this model, is a matter of setting parametric values on innately given principles. One total consistent setting yields an idiolect of, say, English, while another yields an idiolect of, say, Tagalog. The values of the parameters are, in theory, *triggered* by specific inputs, which the learner is assumed to perceive as structured, presumably in accordance with some prior grammar (J. D. Fodor, 1998b).

As is well known, this model of acquisition faces a broad range of difficulties.<sup>7</sup> Whether these can be adequately addressed is not a question that concerns us here. For present purposes, all we need is an understanding of how the P&P framework—considered as an approach to *both* formal syntax *and* acquisition theory—animates the “internal” condition of explanatory adequacy. Within the P&P framework, a grammar of a particular natural language is said to be explanatorily adequate when it can be derived from UG, which is assumed to be encoded in the initial state of the language faculty.

We are now in a position to see how the aim of achieving explanatory adequacy in syntactic theory is thought to underwrite an argument for the cognitivist conception of language. For, on the cognitivist conception, language acquisition is one of the primary explananda of

---

<sup>6</sup> The parameters are often taken to be binary, but this is not a core claim of the P&P model. It is, instead, an additional claim, to be established on independent grounds. In this respect, it is like the claim that the parameters refer solely to the features of functional heads—a condition imposed by grammars in the Minimalist tradition (e.g., Chomsky, 1995).

<sup>7</sup> For a decidedly unsympathetic evaluation, see Tomasello (2005). Tomasello raises a serious objection to the P&P model of acquisition, targeting specifically the assumption, highlighted above, that the linguistic input does not come in the form of a labeled syntactic structure, but, rather, in the form of a messy acoustic and visual stream. For a child to assign *any* syntactic structure to pieces of the primary linguistic data, it must have somehow *already* bridged the gap between an acoustic description of the data and the more abstract description couched in some or other syntactic formalism. Tomasello’s point is that whatever cognitive resources allow the child to do *this*—in his view, sophisticated statistical reasoning, coupled with innate mind-reading abilities—are, in principle, powerful enough to both generate a rudimentary syntactico-semantic framework, and then to refine that framework in accordance with the needs engendered by increasingly complex social interactions. The objection is, I think, a powerful one, but by no means decisive. Impressive efforts to meet it have been made, e.g., by J. D. Fodor (1998b).

linguistic theory.<sup>8</sup> As such, the project of explaining acquisition becomes intimately bound up with the project of finding linguistic universals, or at least patterns of commonality among seemingly diverse languages. Once found, such universals can be assumed to be innate in the mind of the prelinguistic child, and to contribute to his or her acquisition of the local language. With this picture in mind, it's difficult to see how any *other* conception of language can license serious inquiry into linguistic universals or, perhaps equivalently, into the structure of human language *as such*, over and above the structure of particular dialects or idiolects. If language acquisition were *not* an explanandum for linguistic theory, or if the explanans did not make reference to a psychologically real Universal Grammar (UG), then what other role could UG possibly play in linguistics? And if it plays none, then why spend time devising an account of it?

Tempting as it is to let these rhetorical questions stand as an argument for the cognitivist conception, I think the nominalist can provide reasonable answers to them, or at least blunt their rhetorical force. This is what I propose to do below. Adopting the cognitivist conception, I shall argue, is *not* the only way of drawing the distinction between “external” and “internal” adequacy conditions, *nor* the only way of motivating the condition of explanatory adequacy. Put differently, the eminently reasonable aim of constructing grammars that meet an explanatory adequacy condition is quite *neutral* as regards the cognitivist conception and its rivals. As such, the cognitivist conception is—pending further argument—strictly optional.

I hasten to point out that this conclusion does not entail that syntactic principles are not psychologically real. As I go on to show in chs. III-V, an independent argument for the psychological reality of grammar is, in fact, available. That argument rests entirely on well-

---

<sup>8</sup> Needless to say, opponents of the cognitivist conception can, should, and do allow that language acquisition is one of the primary explananda of *some* theory. What they deny is that the theory in question is the formal grammar of the language, or of human language more generally.



documented evidence pertaining to language comprehension in mature, adult speakers, and thus steers clear, as much as possible, from controversial claims about language acquisition, linguistic universals, and the explanatory adequacy condition on grammars. Thus, while I am, on the whole, quite enthusiastic about the psychological reality of syntactic principles, I am skeptical of some of the standard arguments that are used to motivate the cognitivist conception of linguistics, with which the psychological reality claim is typically associated. What I say below is intended to undermine the argument from explanatory adequacy. In §2, I discuss the grounds for rejecting a separate argument, which has to do with the status of E-languages.

### §1.2 *Defusing an argument for the cognitivist conception*

A natural place to begin is with Chomsky's earliest works, *Syntactic Structures* (1957) and *The Logical Structure of Linguistic Theory* (LSLT).<sup>9</sup> Here, the demand for genuinely *explanatory* theories of language is already clearly in place—indeed, serves as one of the hallmarks of the profound shift that Chomsky wrought in linguistics. But, crucially, *this demand is in no way motivated by cognitivist concerns*. In a comprehensive exegesis of Chomsky's work in linguistics, Collins (2008) writes:

Paired with external adequacy conditions is an 'internal' adequacy condition (*Logical Structure of Linguistic Theory*, pp. 80-1, *Syntactic Structures*, p. 50). The external conditions relate a given grammar to the facts (data); the internal condition relates a given grammar to GLT [general linguistic theory—a theory of *all* natural languages, not just a specific one]. This notion is something more than the innocent idea that a given grammar may be evaluated in relation to one's ongoing general conception of a grammar as determined by the set of grammars under construction. General linguistic theory, Chomsky (LSLT, ch. 4) hopes, may provide a simplicity metric by which one could choose the simplest grammar among a set that all meet external conditions of adequacy. (p. 37)

---

<sup>9</sup> Chomsky penned *The Logical Structure of Linguistic Theory* in 1955-1956, prior to publishing *Syntactic Structures* (1957), but the work only appeared in print in 1975.

It is plain, I think, that this approach is vastly superior to the opposing methodology that stressed quasi-botanical “discovery procedures” at the expense of explanation. However, what Collins fails to make clear is how the idea of using General Linguistic Theory (GLT) to provide the relevant kind of simplicity metric is supposed to go *beyond* “the innocent idea that a given grammar may be evaluated in relation to one’s ongoing general conception of a grammar as determined by the set of grammars under construction.” Later in his discussion, Collins motivates this “innocent idea” as follows:

It *might* be that language is essentially heterogeneous, but it would be absurd to assume that constructing grammars for  $L_1$  and  $L_2$  are wholly independent enterprises. If one inquiry can usefully constrain the other, then it would be obtuse to eschew the discipline. Otherwise put, far from *assuming* that all languages are cut from the same cloth, Chomsky is suggesting that it is a matter of empirical inquiry. We construct individual grammars and test them against one another in terms of their shared resources. The extent to which the grammars all conform to the same general conditions and are externally adequate is the extent to which we are discovering something about language as such. Inquiry into particular languages thus goes hand in hand with the general inquiry into language, each constraining the other (Collins, 2008: pp. 83-4, emphases added)

The crucial point to note here is that the motivation for constructing a General Linguistic Theory is, on the hand, quite cogent, but, on the other hand, has nothing to do with language acquisition, nor, indeed, with *any* aspirations of psychological theorizing.

To motivate a shift away from the structuralist methodology, according to which language should be studied as though it were “heterogeneous”—i.e., indefinitely variable from one speech community to the next—one need only stress that, in general, it’s a virtue of *any* theoretical inquiry that it set the bar high, imposing stringent evidential constraints and adequacy conditions. Generality of theoretical coverage is to be valued for its own sake—a point so evident that one wonders how theorists could ever have ever ignored it. But, while we should be grateful to Chomsky, Fodor, and others for bringing it back to its “Moorean” status, we should at the same time guard against attempts to treat it as a proprietary feature of the cognitivist

conception of language. For, again, *anyone*—even the staunchest behaviorist—can fit a demand for theoretical generality into her methodological outlook.<sup>10</sup> It follows that an insistence on explanatory adequacy *cannot* serve as grounds for an argument in favor of the cognitivist conception. In what follows, I elaborate this line of reasoning in response to Collins’ paradigmatic statement of the rationale for taking what he calls the “cognitive turn.”

In what I take to be the most important passage of his book, *Chomsky: A Guide for the Perplexed* (2008), John Collins puts forward precisely the kind of argument that I am here seeking to defuse. It will be instructive to quote him at length.

The cognitive turn, then, essentially arises via a consideration of how language is acquired. The internal justification of general linguistic theory becomes a condition of ‘explanatory adequacy’, where a grammar is adequate only if it is explicable as arising from what all humans share qua an organism that can acquire any human language. Following Chomsky (ATS, ch. 1), we will call whatever this shared property as [sic] *Universal Grammar* (UG). Chomsky (EML, CILT) also refers to UG as a Language Acquisition Device (LAD). This is simply a label for whatever it is about the human mind/brain that maps from exposure to language to a grammar. However it is perfectly natural to construe general linguistic theory as UG, what all humans share. After all, every human acquires a language and so somehow represents a grammar for their language qua creative users of the language. Since, further, general linguistic theory describes the common resources of the grammars and there must be something common to all humans as acquirers of language, it looks as if general linguistic theory was all along an account of a universal human cognitive feature, that is, UG. In line with this cognitive construal of the old methodological notions, ‘external’ justification becomes a desideratum for ‘descriptive adequacy’, that is, a grammar should map a structural description to each sentence such that the set of descriptions explains our intuitions as to the structure within and between sentences. In general then, given a grammar *G* and language *L*, we may take *G* to be descriptively adequate of *L* if it explains the various systematic features of our knowledge or understanding of *L* via an assignment of structural descriptions to *L* sentences. *G* itself cannot be explanatorily adequate, but *G* is

---

<sup>10</sup> In the main text, I have stressed the notion of theoretical generality. But there is also something to be said for the super-empirical virtues of explanatory simplicity and, perhaps more importantly, explanatory *unity*. As noted above, the P&P model was motivated in large part by the fact that it can be used to explain a wide range of seemingly disparate phenomena within a particular language by appeal to a handful of interacting syntactic principles. (See, e.g., Ludlow, 2011: ch. 1 for detailed examples.) Moreover, the model uses a small number of parameters to account for seemingly unrelated differences *between* languages. A prime example of this can be found in the unification of a great many linguistic properties by reference to the *pro-drop* parameter. (See Haegeman, 1994: pp. 19-25 and Berwick, 1991b.) For our purposes, the important point is that these virtues are enjoyed by the P&P model *irrespective* of whether that model is construed as a theory of the cognitive states involved in acquisition.

not adequate unless we are able to show that it is an acquirable grammar from UG and data to which the child is likely to be exposed. Explanatory adequacy relates UG to particular grammars; the condition distinguishes between, as it were, possible languages from ‘arbitrary symbol systems’ (ATS, chapter 1). In sum, linguistic theory becomes an account of human *knowledge of language*. Particular descriptively adequate grammars constitute theories of what speakers know when they know a language, and UG is what speakers know in virtue of them being able to acquire any given language, that is, to represent a given grammar for their language. It should be noted that Chomsky (LSLT, p. 62) has always been happy to treat language as an object of knowledge and so a psychological phenomenon. What is new with the work of the early 1960s is the explicit construal of linguistic theory as having psychological states of knowledge as their object. (pp. 85-6)

I hasten to recognize that Collins may not have intended this passage as a stand-alone *argument* for the cognitive conception of linguistics. Certainly the passage contains a *move* from a seemingly neutral conception to one in which “linguistic theory *becomes* [my emphasis] an account of human *knowledge of language*.” But it may be that this move is simply a description of what Collins sees as the historical facts concerning the cognitive conception’s rise to prominence. In context, however, Collins’ remarks leave one with no doubt that he is an ardent supporter of the cognitive conception—i.e., that he takes the “cognitive turn” to be a turn in precisely the right direction. And, given what he goes on to say in the remainder of the chapter, it is difficult to credit the suggestion that the passage quoted above is intended to carry no argumentative force whatsoever.

On the other hand, if the passage is indeed an argument, then it plainly begs the question against a position like Devitt’s, described in ch. I. For, Devitt takes great pains to argue that the Representational Thesis—arguably the core of the cognitivist conception of language—is unmotivated and probably false. Yet Collins simply *assumes* some version of the Representational Thesis when he writes that “the human mind/brain ... *maps from exposure to language to a grammar*” and, more obviously, that “every human acquires a language *and so*

*somehow represents a grammar for their language*” (emphases mine). As far as I can discern, nothing that Collins says prior to this passage warrants these claims.

It is, of course, possible that all Collins means by ‘represents’ here is something quite weak, like ‘behaves in accordance with’. In this sense, any device that computes a function thereby represents that function. To use Devitt’s terminology, the device *respects* the structure-rules of the grammar.<sup>11</sup> Unfortunately, as Devitt convincingly argues, this wouldn’t be sufficient to show that grammars are theories of the mind, any more than the rules of a card game such as poker are theories of the mind. Devitt would insist that, like the rules of poker, grammars do no more than specify (albeit in exquisite detail) what sorts of inputs and outputs a person will take and emit. Without further argument—specifically, without what Devitt (2006) calls a “powerful psychological assumption” (p. 35)—the rules themselves cannot be taken as specifying the psychological mechanisms by means of which these input-output relations are effected.

Lacking such an assumption, Collins’ claim reduces to the banal observation that whatever “all humans share qua an organism that can acquire any human language” must produce languages that we antecedently know to have various structural features. On this interpretation, the “cognitive turn” is far less exciting than it has been made out to seem, both in academic journals and in the popular press (e.g., Pinker, 1994). For, it is no surprise whatsoever to learn that there is *something* inside of each human being that accounts for whatever universal linguistic features there are (if, indeed, there are any). The claim is even less exciting when, upon reflection, it emerges that this shared trait—what Collins dubs the “trivial” property P<sup>12</sup>—is

---

<sup>11</sup> “[W]e can say that a competence and its processing rules must ‘respect’ the nature of the appropriate output in that, performance errors aside, the processing rules must produce outputs that have that nature” (Devitt, 2006: p. 22)

<sup>12</sup> “Trivially, then, there is some property P of the human cognitive design that allows each of us to acquire any given language. If all of this is so, then it would appear that inquiry into what each language shares is a substantial pursuit, for we must all share something given that we can acquire any language, notwithstanding the many apparent

not (yet) constrained in any independent way. Thus, for all we have been told thus far, P may be a very long way away from a Universal Grammar (UG)—i.e., the rich structure of highly specific, abstract, innate linguistic principles that are the hallmarks of the cognitive turn. For this reason, Collins' equation of P with UG is unmotivated. Specifically, the consequent of the following inference is a *non sequitur*.

Since, ... general linguistic theory describes the common resources of the grammars and there must be something common to all humans as acquirers of language, it looks as if general linguistic theory was all along an account of a universal human cognitive feature, that is, UG. (p. 86)

It's of course true that "general linguistic theory describes the common resources of ... grammars" and it's likewise true that "there must be something common to all humans as acquirers of language." (Devitt [2006: p. 267] calls this the "boring innateness thesis.") But it simply *does not follow*, and it probably is not true, that every linguistic universal that we ever discover—if, indeed, we discover any<sup>13</sup>—must automatically be seen as being encoded in the human genome, represented in the minds of competent speakers, or involved in language acquisition. Regarding the latter claim, Cowie (1997) writes:

[I]t is in general false that theorizing about a thing's essential properties is the same enterprise as theorizing about what we need to know in order successfully to learn about that thing. Biologists may worry about what makes cats cats, but a child's grip on cat-hood predates her excursions into zoology. Philosophers attempt to uncover the essential properties of persons, but it is not your knowledge of that essence that is helping you deal with your boss. Mechanics, perhaps, theorize about the properties shared by all internal combustion engines, but no cognizance of those universals, thankfully, is necessary for our learning to drive. And linguists, finally, theorize about UG, the essence of natural languages. But just as was the case with respect to cats and persons and cars, it is a very substantial additional claim that UG describes

---

differences between any two languages" (Collins, 2008: p. 85). See also Chomsky (1986: p. 17).

<sup>13</sup> It is noteworthy that compelling arguments have recently been advanced for the conclusion that the sustained search for linguistic universals has come up entirely empty-handed. See Evans and Levinson, "The myth of language universals: Language diversity and its importance for cognitive science," in *Behavioral and Brain Sciences* (2009).

not just languages' essential properties, but also what we need to know in order for language acquisition to occur. (pp. 27-28)

The fact is that, for any putative linguistic universal, there are numerous possible explanations.

Methodologically, we must first get clear about what universals there are—or even whether there are any—and then examine each one, case by case, proposing and (dis)confirming competing genetic, environmental, social, and psychological explanations.<sup>14</sup> An *a priori* commitment to the cognitive conception is, as Devitt argues, simply out of place in an empirical discipline.

It is, moreover, important to note in this context that, at present, we don't have any *independent* idea of what exactly it is that “all humans share qua an organism that can acquire any human language.” Pending some independent grip on this property, P, it is vacuous to claim that some grammar “is [or is not] explicable as arising from” P. Consider again Collins' claim:

In general, then, given a grammar G and language L, we may take G to be descriptively adequate of L if it explains the various systematic features of our knowledge or understanding of L via an assignment of structural descriptions to L sentences. G itself cannot be explanatorily adequate, but *G is not adequate unless we are able to show that it is an acquirable grammar from UG and data to which the child is likely to be exposed.* (p. 86; emphasis mine)

If we don't know what P (or UG) *is*, then what sense can we make of the assertion that G<sub>1</sub> is explicable as arising from it, but G<sub>2</sub> is not? Until such time as something substantive is known

---

<sup>14</sup> Collins provides only a couple of examples of actual linguistic universals (p. 85). First, he cites the possibility of ambiguity, and the need for transformations to explain it. Another universal is “creativity”—the capacity for indefinitely novel production and comprehension. These are notably less sexy than the universals whose discovery is sometimes touted as a shining achievement of the Chomskyan approach—e.g. X-bar theory or the Head-Movement Constraint (both of which, incidentally, are now under fire from recent developments in the Minimalist program). The universals that Collins mentions are quite general, and thus impose very modest constraints on grammars. Correspondingly, they provide very weak support for Chomsky's innateness thesis. That these constraints rule out some very anemic grammars is, I suppose, somewhat interesting, but it shouldn't blind us to the fact that the constraints tell us very little about the innate endowment that the child brings to bear on the acquisition task. Indeed, if our cognition of *nonlinguistic* domains can *also* be captured by recursive formalisms—a not wholly implausible proposition (consider our understanding of kinship and other social relations, our mathematical competence, our tonal and rhythmic competence, etc.)—then this particular constraint doesn't even militate in favor of independent *language* faculty. In such a hypothetical-but-not-inconceivable case, the constraints would tell us only about general cognitive development, not language acquisition *per se*. More generally, only the existence of very *specific* universals can be marshaled as evidence for a robust innateness thesis.

*independently* about our innate endowment—in distinctly psychological or computational terms, rather than merely biochemical or anatomic terms—the only way we have of choosing between  $G_1$  and  $G_2$  is by comparing how they fit with the ongoing research into the grammar of *other* languages. This vindicates the aforementioned “innocent” methodological approach of aiming at explanatory adequacy—i.e., at a maximally simple, general, and unified theory of human language *as such*—but it does nothing to warrant a cognitive conception of language.

To put the point slightly differently, consider the fact that the only evidence that one might produce for the claim that UG has structure  $S$  rather than  $S^*$  would be a discovery of some linguistic universal. Consider, furthermore, that we have no way of determining what linguistic universals there are except by constructing grammars for a variety of languages and checking whether the constructs employed by the grammar of  $L$  are applicable to another language,  $L^*$ . Given that there is no evidence for the structure of UG besides what we garner from the innocent methodology of devising and comparing the grammars of various languages, and given that this methodology does *not* presuppose or require a cognitivist conception of language, Collins’s argument for taking the “cognitive turn” does not work. What Collins needs, but does not have, is some *independent* ground—beyond the reasonable aspiration toward explanatory adequacy—for thinking that UG is in the mind. Pending a reason to think that UG is a distinctly psychological faculty, mechanism, process, or state, we are warranted in conceiving of UG as simply the linguist’s working hypothesis about the structure of human language as such.<sup>15</sup>

---

<sup>15</sup> See also Devitt (2006: ch. 12) for a discussion of various arguments for the innateness of a mentally represented UG. On Devitt’s view, an adequate grammar must be *true of* the external representational system that is a language. Moreover, he holds that human languages, like the representational systems of bees, prairie dogs, and other creatures, are acquired on the basis of innate constraints. It follows that an adequate grammar will not ascribe properties to a language that would make it impossible to acquire on this basis. Hence, an independently confirmed theory of the innate constraints can help the linguist narrow down the space of adequate grammars for a language. Devitt claims, however, that our present-day understanding of the innate constraints is poor. I echo this claim in the final pages of this chapter.



I have argued that the “innocent” methodological approach is innocent precisely because it is equally consistent with both the cognitive construal of linguistics and its denial. My point, to reiterate, is that this fact alone is sufficient to undercut Collins’s strategy of presenting the virtues of that methodology as somehow bolstering the cognitive construal of generative grammar against its rivals. Collins cannot make this move without begging the question against those who, like Devitt, are antecedently skeptical of the psychological reality of the linguist’s theoretical constructs. Adopting the aim of explanatory adequacy and the search for linguistic universals does not force upon us a cognitive interpretation of linguistics, precisely because that methodological aim is common ground to both cognitivists and their contemporary opponents.

Before leaving this topic, I must note the inadequacy of Collins’ gesture, at the end of the long passage quoted above, toward precisely the sort of argument that I am claiming he needs.

Explanatory adequacy relates UG to particular grammars; the condition distinguishes between, as it were, possible languages from ‘arbitrary symbol systems’ (*Aspects of the Theory of Syntax*, ch. 1). In sum, linguistic theory becomes an account of human *knowledge of language*. Particular descriptively adequate grammars constitute theories of what speakers know when they know a language, and UG is what speakers know in virtue of them being able to acquire any given language, that is, to represent a given grammar for their language. *It should be noted that Chomsky (LSLT, p. 62) has always been happy to treat language as an object of knowledge and so a psychological phenomenon.* What is new with the work of the early 1960’s is the explicit construal of linguistic theory as having psychological states of knowledge as their [sic] object. (p. 86, emphasis added)

The italicized inference is, I claim, a howler. The fact that something is an “object of knowledge” has *no bearing at all* on whether that thing is a cognitive state, or a psychological faculty, mechanism, or process. If I know a great deli in New York—i.e., if the deli is an object of my knowledge—it simply does not follow that the deli is a psychological entity.

## §2. E-language and I-language

### §2.1 *Motivating the study of E-language*

As noted in ch. I, Devitt (2006) argues that the objects of linguistic theorizing are linguistic expressions, construed as those physical entities—inscriptions, acoustic blasts, and muscle movements—that are distinctive in having phonological, syntactic, and semantic properties, in addition to their brute physical properties. This is the essence of what Devitt has come to call the “Linguistic Conception.” Collins (2007) takes Devitt’s position to be committed to the existence of what Chomsky (1986) has called ‘E-languages’. Devitt is, however, hesitant to commit himself to E-languages outright.

On my view, a language is composed of the outputs of a linguistic competence, symbols that are governed by a system of linguistic structure rules. That is the reality of a language. And the task we have been contemplating, and that I wish to promote, is the study of the nature of this reality. This is not Chomsky’s task (i), the study of the nature of the competence itself. Indeed, at first sight, the contemplated study may seem to be alien to Chomsky’s enterprise. It may even seem to smack of studying an “E-language”, of which Chomsky takes a dim view: “the concept [of an E-language] appears to play no role in the theory of language” (1986: 26); an E-language has “no corresponding real-world object” (p. 27). But it is not obvious that the outputs of linguistic competence fit Chomsky’s description of an E-language. According to him an E-language is “externalized . . . in the sense that the construct is understood independently of the properties of the mind/brain” (1986: 20). And it sometimes seems as if an E-language for Chomsky is essentially Platonic. In any case, the outputs I have identified, physical sentence tokens governed by a system of linguistic rules, are certainly not divorced from the mind/brain since they are the symbolic outputs of the mind/brain. In studying them our object of study is not the mind/brain, of course, but their linguistic properties are surely largely determined by the mind/brain. Finally, the theory of them is as much concerned with real-world objects as the theories of horseshoes, chess moves, bees’ dances, and *wffs*. It is often convenient to talk of the objects posited by these theories as if they were types not tokens, as if they were abstract Platonic objects, but this need be nothing more than a manner of speaking: when the chips are down the objects are parts of the spatio-temporal physical world. (pp. 25-26)

Nevertheless, there is a clear case to be made for seeing Devitt as a friend of E-language. For one thing, Devitt is skeptical of the very reality of I-languages. More importantly, he commits himself (*Ignorance*, §10.5) to the existence of public languages. Indeed, he argues that linguistic

theory is concerned with *these*, rather than with I-languages; his “sixth major conclusion” in *Ignorance* is that “the primary concern in linguistics should not be with idiolects but with linguistic expressions that share meanings in idiolects” (p. 183). Finally, Devitt also rejects a crucial aspect of Chomsky’s characterization of E-language—viz., its association with behaviorism. Barber (2004) writes: “Folk languages could be regarded as E-languages, but Chomsky does not have the folk in mind so much as philosophers who are explicitly or implicitly behaviorist in their assumptions.” As Devitt (2006: ch. 5) makes quite clear, he is no fan of behaviorism. So, while the entities that Devitt is concerned with are not “independent of the mind/brain,” nor Platonic abstracta, nor relics of behaviorist theorizing, it is safe to say that they are E-languages nonetheless.

As Devitt notes in the passage quoted above, Chomsky has strenuously objected to the coherence and the utility of the notion of E-language. Since the introduction of the E-language/I-language distinction in *Knowledge of Language* (1986), Chomsky has repeatedly claimed that defining a notion of E-language is not only pointless, but that he knows of no successful examples.<sup>16</sup> Collins (2008) echoes these concerns in the following passage:

There is a philosophical cottage industry attempting to show that there is a coherent notion of public language or at least a language that is independent of speaker/hearers’ brain states. It might be of some curious interest were such efforts to prove successful, but their relevance to linguistics is wholly obscure, for no one has made any serious suggestion how such a public notion might fruitfully enter into our understanding of the empirical questions of the structure and development of linguistic competence. A metaphysics of our commonsense concept of a language challenges no thesis of linguistics. (p. 20)

---

<sup>16</sup> One might object here to Chomsky’s frequent use of the word ‘define’ in this connection. As Chomsky is surely well aware, the notions employed in empirical pursuits are rarely if ever *defined*, strictly speaking. Rather, they emerge from fruitful theory construction. Chomsky’s own characterizations of I-language are not, I take it, definitions.

Collins and Chomsky are, I think, perfectly right to point out that the friend of E-language owes an account of both the utility and the content of this notion. There are problems, however, with the way that Collins couches this demand in the passage quoted above. For, he limits in advance the interesting questions for linguistics to only those that have to do with “the structure and development of linguistic competence.” Following Chomsky, he suggests that any other concern belongs to mere folk theory. This, I think, begs the question against those who, like Devitt, argue that routine practice in linguistics is concerned precisely with dialects, public languages, and families thereof (e.g., the Romance, Slavic, and Germanic families).

[Public language] classifications seem to be useful in linguistics, for linguistics books and articles are replete with them. The first few pages of Haegeman’s GB textbook (1994) have many uses of ‘English’ to classify shared meanings. Then ‘English’ is compared with ‘Italian’; for example, ‘In Italian a subject of a subordinate clause can be moved to the main clause domain across the overt conjunction *che*, corresponding to *that*; in English this is not possible’ (p. 20). And then with Spanish and French (p. 23). And so on throughout the book. (*Ignorance of Language*, p. 183, fn 23)

Moreover, Devitt points out that a theory of public languages *does* in fact contribute to our understanding of “the structure and development of linguistic competence.” On his view, linguistic competence is an ability to produce and comprehend expressions in a public language. Hence, we can only understand that competence if we have a prior grip on the nature of the language itself.<sup>17</sup>

---

<sup>17</sup> There is another problem with the quoted passage from Collins. His rhetoric suggests that questions of ontology are not empirical, contrary to the conclusions we drew in ch. I. It is important, I think, to remember that there is no one particular kind of inferential consequence that a thesis must have in order to count as empirical. The fact that some thesis about the metaphysics of language does not, by itself, yield predictions about syntactic structure or the acquisition process does not mean that this thesis has no inferential consequences whatsoever. It very well might have significant consequences, either at present, or at some future time, when conjoined with other appropriate premises or theoretical commitments. In any event, making and defending claims about the history and sociology of a field like linguistics is surely an empirical enterprise.

Thus, for Devitt, the point of clarifying the notion of an E-language—or, better, the nature of E-languages—is to show that *this* is what working linguists are actually studying, *pace* Chomsky’s ideology and perhaps even despite their own protests. He writes:

Turn next to Chomsky’s claim [that] conventions have no interesting bearing on the theory of meaning or knowledge of language. Chomsky thinks that this also should be a truism (1996: 48). Yet, if it were true it would seem to be at odds with what most linguists are actually doing. For, what they are mostly doing is theorizing about the largely conventional syntactic and semantic properties of expressions. And they are right to be doing so, in my view. Conventional meaning is important to theory in at least four ways. (p. 181)

Devitt goes on to respond to Chomsky’s demand for a specification of the alleged role that E-language plays in empirical theorizing. His answers have mostly to do with conventional shared meanings and the roles that they play in explaining successful communication and in the ascription of propositional contents to the internal states that explain our behavior.

The members of any group that share a meaning of one linguistic expression tend to share meanings of a vast number of others and it is convenient, on the basis of this, to follow the custom of classifying sets of these expressions with shared meanings as English, Spanish, and so on. The classification is bound to be a bit vague but no more so than many scientifically appropriate ones. And such classifications seem to be useful in linguistics, for linguistics books and articles are replete with them. Are these classifications mere manners of speaking that can be paraphrased away when the serious linguistic work is to be done? I think not: they are necessary for the linguist to identify what she is talking about, to identify the subject matter. For, the subject matter is the shared meanings and syntactic properties of linguistic expressions in a certain group of people. (p. 183)

For reasons that go beyond the scope of the present discussion, I am not convinced that we must appeal to public meanings in explaining communication and the ascription of propositional contents. Nevertheless, Devitt does mention one explanatory project for which reference to conventions and public languages seems to me to be indispensable—viz., the project of giving an account of language acquisition. He writes:

[T]he syntactic differences between public languages show that much syntax is not innate. These differences are captured, on the received Chomskyan view, by different settings of “parametric values”. Very occasionally an idiolect’s parameter

settings may be eccentric but almost always they will be conventional. Thus most people in the USA participate in parameter-setting conventions that lead them to speak an SVO language; most people in Japan participate in parameter-setting conventions that lead them to speak an SOV language. (Devitt, 2006: p. 181)

In what follows, I would like to expand on this claim, making clear its implications for how we should regard both acquisition theory and the famous poverty-of-stimulus arguments that Chomsky has popularized.<sup>18</sup>

In his original discussion of the E-language/I-language distinction, Chomsky (1986) claimed that certain aspects of the folk conception of language would have to be abandoned as the study of language goes scientific. One of these is what he called the “normative-teleological” aspect of the folk concept, which arises in connection with considerations of language acquisition, in both children and adults.

The commonsense notion [of language] also has a normative-teleological element that is eliminated from scientific approaches. I do not refer here to prescriptive grammar, but to something else. Consider the way we describe a child or a foreigner learning English. We have no way of referring directly to what the person knows. It is not English, nor is it some other language that resembles English. We do not, for example, say that a person has a perfect knowledge of some language L, similar to English but still different from it. What we say is that the child or foreigner has a “partial knowledge of English,” or is “on his or her way” toward acquiring

---

<sup>18</sup> In a recent reply to Collins, Devitt (2008b) makes reference specifically to parameter-setting accounts of language acquisition in the course of motivating a commitment to conventionally instituted public languages. “I argue that language acquisition provides one reason for believing in SLEs [Standard Linguistic Entities]: the best explanation of the setting of parametric values in the typical member of a linguistic community is that she comes to participate in the parametric conventions of the community. ... My view is that *the best explanation* of our *actual* acquisition adverts to conventions. So that is a good reason to believe that there are conventions. The child sets his parameters in a certain way because it regularly experiences others who have set them in that way. Conventions explain those regularities. Thus, “conventions explain the child’s experiences which explain its settings. And the conventions explain why all the children in that community usually come to have the same setting” (2006b: 582).

While I myself am happy to endorse parameter-setting accounts of acquisition, I am not sure that such accounts are compatible with Devitt’s rejection of the psychological reality of syntactic principles. As it is ordinarily conceived, the enterprise of setting parameters is a psychological one, and, crucially, involves mentally representing the rules or principles of UG. Moreover, there are parameter-setting models (e.g. J. D. Fodor’s Waiting-STL and Guessing-STL, Charles Yang’s VL, and others) that rely on parsing strategies that Devitt must reject on similar grounds. Perhaps Devitt intends his talk of parameter setting as committing him only to the claim that children “come to participate in the parametric conventions of the community.” Still, the going accounts of the psychological processes that lead children to their eventual success in participating in such conventions traffic heavily in mentally represented rules and principles. Devitt addresses such concerns in ch. 12 of *Ignorance of Language*.

knowledge of English, and if they reach that goal, they will then know English. Whether or not a coherent account can be given of this aspect of commonsense terminology, it does not seem to be one that has any role in an eventual science of language. I will follow standard practice in disregarding these aspects of the commonsense notions of language and the associated notions of rule-following and so forth, although the departure should be noted and one may ask whether it is entirely innocent. (Chomsky, 1986: p. 16)

A glance at the literature on language acquisition, however, reveals that what Chomsky calls the “commonsense notion” actually plays a major and seemingly indispensable role in studies of language acquisition. Consider, for instance, the following passages, which are taken from a recent and authoritative textbook by Maria Teresa Guasti. (Guasti is not only a leading researcher in the field but also, without a doubt, a proponent of the Chomskyan paradigm.) I have underlined what I take to be instances of “normative-teleological” vocabulary, as well as references to public languages.

Human language acquisition is an astonishing process. Let us consider what these children have accomplished in about 3 years. Although their language may still not be perfect, they put words in the correct order. Nina produces quite a complex sentence, putting the complements in the right order (first the direct object and then the prepositional complement) and applying wanna-contraction. Scrambling complements is possible in Italian, and Diana shows that she can take advantage of this option, by putting the prepositional complement (*a Luca* ‘to Luca’) before the direct object (*la bambola* ‘the doll’). Eve places the adjective *funny* before the noun clown, as required in English, while Diana places the adjective *lossi* (*rossi*) ‘red’ after the noun *capelli* ‘hair’, since she speaks Italian. (Guasti, 2002: p. 2, underlined emphases added)

It is well known that babies born into a multilingual environment can easily pick up more than one language. This shows that infants can distinguish not only between utterances in the language of their environment and utterances in other languages, but also between utterances in two or more languages spoken around them. Learning a language requires discovering the rules of that language—for example, how words are ordered in clauses, and how questions are formed. If infants could not distinguish between utterances from different languages, they might make bizarre conjectures concerning the properties of what they hear. Without this ability, how could a child hearing sentences from, say, French and Spanish ever figure out the properties of French? (Guasti, 2002: p. 24, underlined emphases added)

For the purposes of the present discussion, what is striking about these passages is that they are

replete with what appear to be clear and direct references to public languages, their properties, the expressions that belong to them, the rules that govern them, their presence in the child's environment, and the degrees to which various children's usages are "correct," "right," or even "perfect" with respect to them.

Chomsky and his followers would surely resist this interpretation, claiming that these remarks belong to the informal presentation of a theory, not in its serious development—a distinction that Chomsky wields with worrying frequency.<sup>19</sup> But this is difficult to square with the fact that Guasti (2002) *continues* to make use of the "normative-teleological" terminology throughout the text, even in the context of what appear to be quite serious statements of empirical generalizations. Consider, for instance, the following passages, taken from those chapters of Guasti (2002) where she reports a variety of experimental and corpus data. Once again, I have underlined the phrases that are remarkable in the present context.

Italian, Spanish-, and Catalan-speaking children use singular agreement morphemes with the appropriate subject. ... Errors are rare and mostly found with plural subjects. (pp. 120-121)

Agreement errors are rare, about 3-4% in early Italian (Guasti 1993/1994; Pizzuto and Caselli 1992), 1.72% in early Catalan and Spanish (Torrens 1995). These rare errors mostly consist of using a singular third person morpheme with a plural subject or a third person morpheme with a first person subject. These findings have been replicated for early German: Poeppel and Wexler (1993) found that their child subject, Andreas (2;1), used the first and third singular agreement morphemes accurately; the three plural morphemes and the second singular morpheme were rare or absent in his speech; and errors were rare. (p. 121)

Italian learners whose speech has been studied use postverbal subjects in about 30% of their sentences with overt subjects and make no agreement errors... (p. 125)

examples ... found in the CHILDES database show that children do not make mistakes with coordinate subjects. (p. 125)

---

<sup>19</sup> See, e.g., his exchange with Rey, in Barber (2000) and Antony and Hornstein (2003).



Given that Guasti is a card-carrying Chomskyan linguist, it cannot be that her use of the underlined phrases signals an explicit rejection of Chomsky's ideology. One might, of course, resort to the claim that her usage in these passages (and countless others throughout the book) is merely "informal" or perhaps even "sloppy."<sup>20</sup> Putting aside the fact that such claims run a serious risk of insulting Guasti and many other acquisition theorists, one simply finds no warrant for them in the actual text. Indeed, consider the kinds of empirical findings that animate inquiry in acquisition theory. As illustrated in the passages quoted above, these typically concern quantitatively described patterns of error—e.g., the percentage of agreement errors that children make in Italian (apparently, 3-4%). It is extremely difficult to see how such findings might be "formally recast" as claims about the relation between the child's usage and some specific I-language. Against *whose* I-language would a child's usage be quantitatively compared?

It's true, of course, that the acquisition theorist can claim to be comparing the child's grammar to an *idealized* I-language. But it's not at all clear what import the appeal to idealization has in this context. If we press on the notion of an *idealized* I-language, we find that it amounts to no more than a consistent setting of parameters (assuming a P&P grammar). But there are many such settings, most of which fail to match the language of the community. Thus, reference to the grammar of a public language seems unavoidable in singling out the language

---

<sup>20</sup> Collins (2008) avoids this charge by inserting occasional scare quotes around the offending terms, as, for instance, in the following passage: "Children exposed to a degraded language, such as a pigeon or late-acquired sign, do not learn the 'errors' but develop a language that is consistent in its syntax and morphology. ... The development of language is also marked by very little 'error' outside of narrow parameters. Familiarly, children have trouble learning irregularity in number and past tense and they make other notable errors, such as leaving wh-words in medial positions. Viewed from the perspective of what children could get wrong, however, their competence is remarkably consistent with adult understanding. Indeed, adults often stumble over irregularity. Children's competence is so error free that it has been plausibly suggested (Crain and Thornton, 1998) that children's competence matches the adult speakers; the 'errors' there are performance based" (pp. 105-6, underlined emphasis added)." Plainly, though, this typographical change does not settle the substantive issues. It remains unclear whether Collins can avoid commitment to the existence of a public language—an "adult understanding"—to which the child's idiolect might conform or fail to conform.

that the theorist identifies as the child's "target grammar." If there is an idealization in the vicinity, it is one that abstracts away from the variation *within the community*—i.e., the differences between individual speakers—and yields an *idealized speech community* as the theoretical object of interest.

What this shows, I think, is that the opponent of E-languages owes a formal, "serious" account of the child's goal throughout the acquisition process—a characterization of what is typically called "the target language," "the ambient language," or (as above) "the language of [the child's] environment." Providing such an account would go a long way toward making sense of what acquisition theorists mean when they claim that a child has made a "mistake" in the course of acquisition, or, equally, that no mistake was made. I am not convinced that an explicit, formal, and "serious" account of this would make no mention of E-language. Whether it would is, at this time, an open empirical question.<sup>21</sup>

It is important to note, moreover, that the "normative-teleological" terminology is ubiquitous in the various statements of the poverty-of-stimulus argument for the innateness of UG. Consider the following two passages:

[C]hild errors are rare *tout court*. This would appear to demonstrate that children are making 'decisions' about the target language independent of data. If children were being instructed by the data, one would find a preponderance of early mistakes that would gradually decrease as more disconfirming evidence is accumulated. (Collins, 2008: p. 105)

Although children make "errors," they do not make certain errors that would be expected if they generalized from the linguistic input. For example, although children hear sentences like *Who do you wanna invite?* and *Who do you wanna see?*, they do not generalize from these to impossible English sentences like *\*Who do you wanna come?*; although this generalization would seem reasonable, children never

---

<sup>21</sup> Here, I wish to thank Daniel Harris for pointing out that my main purpose in this chapter—i.e., to argue that Chomsky's cognitivist conception is not the only game in town—does *not* require the strong claim that public E-languages actually play a role in acquisition theory. My conclusion is a weaker one—viz., that we do not, at present, have overwhelming grounds for supposing that E-languages will *not* play any such role.

say such sentences. (Guasti, 2002: p. 5)

Collins' use of the term 'target grammar' carries the implication that there is a grammar external to the child at the time of acquisition—a “theory” of a *public* language—which the child is struggling to grasp. Given that he is at pains to reject this implication, it is important that he be able to state the poverty of stimulus argument without using this term or any of its cognates.

Again, I am not persuaded that this is, in fact, possible to achieve. And if it is not, then we have here a clear motivation, supplied by a respectable, up-and-running, empirical research program, for investigating the nature of E-languages. Combined with Devitt's arguments to the same effect, I believe the case for this project is, if not airtight, at the very least non-trivial.

It is easy to see what the outlines of a methodology for such a project might be. We begin with the idea that the child's goal in acquiring a language is to obtain the capacity for fluid linguistic interaction with maximally many members of her linguistic community.<sup>22</sup> One metric along which the fluidity of communication can be measured—though by no means the most important metric—has to do with the extent to which the grammar of the child's idiolect deviates from the grammar of the “ambient language” in her community. The actual process of constructing such a grammar would proceed by first collecting data from the adult members of that community. Suppose, for instance, that the community initially consists of only two competent adult speakers, A and B. Now, if A and B have a child, C, then the set of their utterances  $\{\{u(A)\} \cup \{u(B)\}\} = \{u(A, B)\}$  exhausts C's primary linguistic data. Now take  $\Gamma_{AB}$  to be the grammar, or set of grammars  $\{\gamma_1, \gamma_2, \dots, \gamma_n\}$ , that best covers  $\{u(A, B)\}$ . The abovementioned metric involves making comparisons between  $\Gamma_{AB}$  and the child's grammar  $\Gamma_C$ .

---

<sup>22</sup> Chomsky famously denies that the purpose of language is communication, but we can put this puzzling view to one side for the moment. Surely, children do attempt to communicate, regardless of whether they are thereby fulfilling the purpose of language (if there be such).

The extent to which there is some well-defined divergence,<sup>23</sup> we can say that  $\Gamma_C$  is a *misrepresentation of the target language*, and, consequently, that C is in *error*.

We can generalize this strategy to the case in which the child is exposed to speech from more than two speakers. Let the number of speakers be  $n$ . Now consider the set of utterances  $\{u(A, B, \dots n)\}$  that are perceived by the language-acquirer from *all* of the speakers with whom he or she has been in contact in the course of the acquisition process.<sup>24</sup> A corresponding generalization takes us from  $\Gamma_{AB}$  to  $\Gamma_{AB\dots n}$ , i.e., the grammar shared by the speakers from whom a person acquires his or her grammar. If we wish, we can assign weights to any member or subset of  $\{u(A, B, \dots n)\}$ , to reflect factors such as the amount of attention that the language-learner paid to the utterances, or the extent to which the learner understood those utterances (if these can be independently measured).

It is of course possible—indeed, likely—that no grammar will adequately capture the set of utterances  $\{u(A, B, \dots n)\}$ . For instance, speakers F and G might have divergent linguistic practices. The linguist has a number of options in dealing with this phenomenon. She can assign weights to particular speakers, and count more heavily the utterances (or metalinguistic judgments) of those people who were most prevalent in providing the child with PLD during the acquisition process. Alternatively, she can appeal to the notion of *degrees* of grammaticality—a notion that was initially proposed in Chomsky (1961/1964), and which has an analogue in the

---

<sup>23</sup> There are many conceivable ways of formally characterizing this sort of divergence. No general characterization can be given for all possible types of grammar. A fan of context-free systems might count how many context-free rules exist in  $\Gamma_{AB}$  but not  $\Gamma_C$ , or vice versa, whereas a proponent of old-school transformational grammars might prefer to count the number of transformations instead. Likewise, if the  $\Gamma_i$ s emerge from the P&P approach, then perhaps it will be most natural to compare parameter settings.

<sup>24</sup> For simplicity, I treat only utterances, though, in principle, we can consider metalinguistic judgments as well.

P&P approach.<sup>25</sup> In the limit, the differences between members of the community might be so dramatic that it would be more expedient to treat them as speaking different languages—i.e., drawing on largely disjoint stocks of phones, phonemes, or lexical items. In such a case, we can consider the child as belonging to two distinct language communities.

The linguist may find that explicitly accounting for variation in the community within her theory significantly hampers her research. In such a case, she has the option of idealizing away from such variation, as the linguists in the structuralist tradition found it expedient to do.

Chomsky (1986) describes this strategy in the following passage:

Modern linguistics commonly ... consider[s] an idealized “speech community” that is internally consistent in its linguistic practice. For Leonard Bloomfield, for example, language is “the totality of utterances that can be made in a speech community, regarded as homogenous” (Bloomfield, 1928/1957). In other scientific approaches, the same assumption enters in one or another form, explicitly or tacitly, in the identification of the object of inquiry... Of course, it is understood that speech communities in the Bloomfieldian sense—that is, collections of individuals with the same speech behavior—do not exist in the real world. Each individual has acquired a language in the course of complex social interactions with people who vary in the ways in which they speak and interpret what they hear and the internal representations that underlie their use of language. Structural linguistics abstracted from these facts in its attempts at theory construction...

Comparing the structuralist approach with his own, Chomsky makes clear that analogous idealizations must be made in order to facilitate inquiry into the structure of I-languages.

---

<sup>25</sup> Chomsky (1988: p. 560) argues that the notion of E-language cannot be defined in any principled way because this would require deciding how to treat strings like ‘The child seems sleeping’. For, on the one hand, the string should be included in the E-language on account of its having a clear meaning, but, on the other hand, any attempt to include it in the E-language will inevitably cast the net too wide, potentially bringing into the fold *all* noises and inscriptions. Ironically, Chomsky’s argument overlooks his own work on degrees of grammaticality. The E-language theorist can associate a language, *L*, with a *multi-tiered* set-theoretic structure that takes such degrees into account. Consider, for instance, the fact that ‘The child seems sleeping’ only has a clear meaning in English because of its relation to ‘The child seems to be sleeping’. This asymmetry warrants a distinction between a “core” set of utterances, and a “peripheral” set. The strings in a core set are the fully grammatical sentences. By contrast, ‘The child seems sleeping’ would be in the less-than-grammatical set. Other strings would be in the fully ungrammatical set, and still others wouldn’t be in any of the sets, perhaps because they aren’t made up of elements in the lexicon. (One must, of course, settle what’s in the lexicon—a non-trivial issue, but one that arises for *all* theories of language.) The distinctions between such sets can be made even more fine-grained by reference to a P&P grammar, where the degree of a string’s grammaticality can be seen as a function of how many syntactic principles it fails to satisfy. See Berwick (1991b) for an extended discussion of the latter point.

[Chomskyan linguistics] also abstracts from these facts in posing [the cognitivist] questions [What constitutes knowledge of language, and how is this knowledge acquired and used?], considering only the case of a person presented with uniform experience in an ideal Bloomfieldian speech community with no dialect diversity and no variation among speakers. ... We should also make note of a more subtle theory-internal assumption: The language of the hypothesized speech community, apart from being uniform, is taken to be a “pure” instance of UG in a sense that must be made precise.... We exclude, for example, a speech community of uniform speakers, each of whom speaks a mixture of Russian and French (say, an idealized version of nineteenth-century Russian aristocracy). The language of such a speech community would not represent a single set of choices among the options permitted by UG, but rather would include “contradictory” choices for certain of these options.

Here I want to stress a point that is commonly overlooked in philosophical discussions of generative linguistics. Given that that the study of *both* E-language *and* I-language requires idealizing away from variations—within a community and within a speaker’s idiolect—a theorist who champions the cognitivist conception cannot argue against the neo-structuralist or nominalist approach on the grounds that it requires making such idealizations. Nor can he substantiate the charge that the objects of his opponent’s preferred mode of inquiry are “unreal” or “fictional.” The two approaches are entirely on a par in these respects.<sup>26</sup>

Note also that the conception of E-language that emerges from the approach outlined above is thoroughly *apolitical*. That’s to say that on *this* approach, the concerns of the practicing linguist dictate a delineation of (idealized) language communities that makes no appeal whatsoever to geographical boundaries, ethnic or religious affiliations, power relations among groups, historical conflicts and alignments, and so forth. As Devitt (2006) points out, “a ‘sociopolitical dimension’ does occasionally intrude into such classifications but the intrusion can be resisted by linguists; for example, a linguist may think that, for almost all expressions,

---

<sup>26</sup> Thus, consider the following remark from Collins (2007). “[L]et us grant that linguistic properties *might* arise from a relation between the mind/brain and external stuff, but all the action is at one end of the relation; the other end is noisy, variable and does not submit to independent inquiry” (p. 420). The reply is that *all* natural phenomena are variable and noisy. Idealization is just as necessary in the study of public languages as it is in the study of the mind/brain. The cognitive conception of linguistics derives no advantage from these considerations.

there is no theoretical point to the politically inspired division of Serbo-Croatian into Serbian and Croatian and can simply refuse to go along...” (p. 184).

Nor is the project in any way intended to contribute to the agenda of *prescriptive* linguistics.<sup>27</sup> The construction of a descriptive grammar is a worthy and noble enterprise, and linguists in the generative tradition are—I think *rightly*—unwilling to be seen as engaging in the construction of a prescriptive grammar. No good can come from confusing their theoretical goals with the goals of prescriptive grammarians. I want to urge, however, that it is *equally* erroneous to conflate prescriptive grammar with the study of public languages, or E-languages. It is important to see that taking idiolects or I-languages as the objects of inquiry is *not* the only way to avoid doing prescriptive grammar. As I argued above, one can construct a descriptive grammar for a set of utterances that have been produced by multiple people. And choosing which group of people to study need not involve taking into account messy socio-economic or political issues. Theory-internal considerations can just as easily guide the choice. These considerations include, *inter alia*, the degree to which the members of a group are disposed (i) to produce strings that can plausibly be seen as having the same syntactic structures, (ii) to comprehend linguistic input in similar ways, (iii) to make the same metalinguistic judgments, and (iv) to communicate in a way that both they and the linguist judge to be maximally fluid.

It must be admitted that the approach that I have recommended in the pages above is “rough around the edges,” to put it mildly. Subtle issues arise about how much agreement is required in competent speakers’ intuitions, to what extent—and in what respects—their linguistic output must be similar, to what degree their comprehension of the same utterances must

---

<sup>27</sup> To fix the terminology, a *descriptive* grammar tells us about what *actual* usage is like, whereas the prescriptivist seeks to impose normative “standards of usage” that invariably deviate from the actual linguistic conventions in communities and the systematic behaviors of actual speakers.

coincide, how to gauge the success or fluidity of their linguistic communication, and so forth. But while constructing an apolitical notion of public languages is difficult, we must not confuse difficulty with impossibility. Many linguists are confident that we can gain a great deal of insight into comparatively daunting matters—the evolution of language, for instance. It strikes me as simply irrational to have confidence in *that* sort of enterprise but no hope at all for an apolitical construction of a grammar for a set of utterances produced by, say, two or more native English speakers. Accordingly, I am confident that increasingly precise characterizations of E-languages can be filled in as theoretical and experimental advances are made.

In §2.3, I will argue that the project of individuating I-languages is replete with similar promissory notes. It follows that the lack of precision in our notion of E-language is just as much a feature of our conception of I-languages. My discussion in §2.3 will highlight some of the difficulties pertaining to the competence/performance distinction, the mimicry of “other” dialects, change over time, multilingualism, code switching, the infusion of “foreign-language” lexical items, brain damage, cognitive disorders, and related phenomena. I will argue that meeting these challenges requires making fine-grained decisions that we do not, at present, know how to make on principled grounds.<sup>28</sup> In all likelihood, making the notion of an I-language precise will ultimately require adopting procrustean, legalistic, and sometimes arbitrary policies. Characterizing a dialect or a public language is no different. Accordingly, one would do well to ask what degree of precision it’s really necessary to achieve in order to get on with fruitful theorizing. I suspect that neither in the case of E-languages nor I-languages will any rational theorist insist on *maximal* precision at the outset of inquiry.

---

<sup>28</sup> Some friends of E-language will want to argue that the most natural way of resolving such difficulties—at least in some cases—will involve drawing *normative* distinctions by appealing to one or another socio-cultural construct. This is *not* the line that I will pursue below.



## §2.2 *Do linguists really study I-languages?*

Barber (2004) gives voice to some of the most popular arguments against viewing linguistics as a study of E-language, and for avoiding that sort of project in practice. His first argument opens with the innocent-sounding premise that linguistics is a science—a claim that I take to be quite obviously true, occasional protests notwithstanding (ch. I). Barber goes on, however, to claim that genuine sciences do not make normative pronouncements and that individuating public languages involves making normative (“prescriptive”) claims. It follows, then, that individuating public languages cannot part of a genuine scientific practice and, *a fortiori*, cannot be part of linguistics. As a separate argument—though plainly continuous with the first—Barber adds that individuating public languages involves reference to social and political factors. He goes on to take for granted the claim that “any object whose individuation conditions are determined by socio-political factors should not figure in the fundamental ontology of science.”

There are many things that strike me as mistaken in this line of reasoning. To begin with, Barber is relying heavily on a sharp distinction between the normative and the descriptive. Putnam (2004) has argued—quite compellingly, to my mind—that this distinction is both mythical and pernicious; whatever currency it had in the heyday of positivist philosophy, we should cast a skeptical eye on it today.<sup>29</sup> Of course, Barber might protest that, by ‘normative claim’, all he meant was ‘a claim that makes ineliminable reference to socio-cultural constructs’. But, then, to say that no science can make normative claims in *this* sense is to preclude any

---

<sup>29</sup> See Devitt (2002) for a defense of a position that has come to be called “Cornell Realism” in moral philosophy, which, in its own way, deflates the distinction between descriptive claims and what has traditionally been thought of as non-descriptive, normative discourse.

scientific study of social-level phenomena. It would follow that anthropology, sociology, economics, political science, history, and many other fields are all forever barred from enjoying the status of genuine sciences, on the *a priori* ground that their ontologies are, in some sense, second-rate. It is hard to take seriously a conception of science that animates such a claim.<sup>30</sup>

Consider now what Barber (2004) says about the objects of linguistic theory.

The operating notion of a language is instead that of an idiolect, often in Chomsky's developed sense of this term (an I-language). So although a label like 'Hungarian' will inevitably be used in practice, it is normally intended to be thought of as convenient shorthand for 'The idiolect of some arbitrary but typical inhabitant of present-day Hungary'.

As formulated here, Barber's claim simply cannot be correct. No linguist seriously thinks that Hungarian (or any other language) is, or need be, confined to a particular geographical region.<sup>31</sup> More importantly, it's not at all clear *why* 'the idiolect of some arbitrary but typical inhabitant of present-day Hungary' would be any better an object of study than simply *Hungarian*. Barber's answer has to do with his claim, criticized above, that individuating public languages involves making normative claims and ineliminable references to socio-cultural constructs. But it seems that Barber has run afoul of his own constraint here, for his phrase 'the idiolect of some arbitrary but typical inhabitant of present-day Hungary' *likewise* makes ineliminable reference to socio-cultural constructs, insofar as *present-day Hungary* is one such. We must conclude, I think, that

---

<sup>30</sup> I hasten to point out that Chomsky does *not* fall prey to this error. He writes: "Internalist biolinguistic inquiry does not, of course, question the legitimacy of other approaches to language, any more than internalist inquiry into bee communication invalidates the study of how the relevant internal organization of bees enters into their social structure. The investigations do not conflict; they are mutually supportive. In the case of humans, though not of other organisms, the issues are subject to controversy, often impassioned, and needless" (Chomsky, 2001: p. 41).

<sup>31</sup> Ludlow (2011) makes a similar mistake in taking this strange way of individuating languages to be a feature of folk theory. Of the folk notion, he writes: "Typically, the question of who counts as speaking a particular language like German is determined more by political boundaries than facts about linguistic properties..." (p. 44). It is difficult to believe that the folk would insist that a monolingual German speaker who has, since birth, resided exclusively in France, does not *really* speak German.

Barber fails to adequately characterize the objects of linguistic theorizing, both by his own criteria and by independent standards. How might we do better?

For dramatic effect, suppose that we were to conduct an informal survey of linguists working in the generative tradition, in an effort to determine whether they are studying public languages or idiolects. Given the prevalence of Chomsky's ideology, which is often presented to students of linguistics alongside his truly impressive results in syntactic theory, we would expect to find that many of them will claim that they are studying I-languages. But now suppose that we continue the survey by asking the linguists, "*Whose* I-language are you studying?" Some will have little to say in response to such a bizarre question. Others will give a response that makes reference to socio-cultural constructs—perhaps the very same ones that Barber appealed to in the passage quoted above. Such responses, which can take any of a number of forms, all leave open the possibility that the linguists who make them were, as a matter of fact, *mistaken* when they initially claimed to be studying I-languages. It may be, after all, that they study E-languages (of one stripe or another), but that they are not aware that this is what they are doing, perhaps because they have not reflected on the matter, or because they've uncritically accepted Chomsky's ideological remarks on this issue. There is nothing incoherent about this. It is both common and, within limits, even rational for scientists to pursue empirical questions without reflecting, in any great depth, on the ontological status of their field's subject matter—or by simply taking for granted the received view among their research community (Kuhn, 1962/1996).

A more philosophically inclined linguist might respond as follows: Ultimately, I am studying *my own* I-language. But, while this ingenious response has the virtue of enriching the meaning of the 'I' in 'I-language', it is by no means clear that it is an adequate response, as an

extension of our hypothetical survey would reveal. Consider, for instance, expanding the survey to include the following questions: Do you plan to publish your findings? If so, then why do you suppose that *your* I-language is important for anyone else to learn about? Do you ever ask other people for *their* intuitions about various linguistic constructions, after having articulated your own intuitions about them? If so, *why*? What purpose does it serve to know what data other people's I-languages generate?

There are quite natural ways of answering these questions, which do not bode well for standard Chomskyan line. For instance, it's plausible to say that the purpose of publishing one's results, even those that the linguist takes to be about his or her own idiolect, is to fuel the project of formulating and confirming generalizations that can be projected across a broader population. Indeed, this sort of generalization is a routine practice among linguists.<sup>32</sup> Similarly, asking multiple informants for their intuitions regarding some linguistic construction can easily be seen as a way of gathering information about the *commonalities* between speakers. If so, then the search for generalizations across speakers seems to be the linguist's primary motivation. On the strength of these considerations, one is led to suspect that what the linguist is ultimately interested in is the nature of the public language—perhaps a dialect—not the I-language of an arbitrary individual speaker.<sup>33</sup> Moreover, it should be noted that consensus among multiple

---

<sup>32</sup> Cf. Devitt (2006): “[Public language] classifications seem to be useful in linguistics, for linguistics books and articles are replete with them. The first few pages of Haegeman's GB textbook (1994) have many uses of ‘English’ to classify shared meanings. Then ‘English’ is compared with ‘Italian’; for example, ‘In Italian a subject of a subordinate clause can be moved to the main clause domain across the overt conjunction *che*, corresponding to *that*; in English this is not possible’ (p. 20). And then with Spanish and French (p. 23). And so on throughout the book.” (p. 183, fn 23)

<sup>33</sup> Cf. Devitt (2006): “There is indeed something a little paradoxical about denying the frequency of linguistic conventions. The linguistic method of consulting the intuitions of linguists and other speakers to discover facts about a language ... presupposes masses of conventional regularities among them, even while allowing for some differences in idiolects. Books are written and papers are given about expressions in this or that language, all of which presuppose a great deal of regularity in usage among speakers” (p. 181).

informants—particularly among professional linguists<sup>34</sup>—is a sign (though by no means the only nor the best sign) that their judgments reflect the nature of the language, rather than being a species of performance error. For this reason, even a linguist who takes herself to be studying linguistic competence can benefit from paying attention first to the character of a public language. Doing so can help her to draw a rough-and-ready distinction between phenomena that are indicative of competence and those that reflect only the vicissitudes of performance. Needless to say, the friends of I-language may well have alternative, cogent responses to the hypothetical survey questions posed above—responses that accord with their ideology. But this is by no means obvious and remains to be shown.

In connection with this issue, consider the following scenario, described by Eddington in his 2008 Presidential Address:

It is unfortunate that many linguistic analyses do not submit their hypotheses to experimental test. Instead, they make the critical error of elevating a hypothesis about a phenomenon to the status of an explanation of the phenomenon (Black and Chiat 1981:48, Higginbotham 1991:555, Itkonen 1978:220-221, Ohala 1990:159, Sampson 2001:124, Yngve 1996). Chomsky provides an example of this common fallacy. He claimed that ‘perform’ may only be followed by a count noun, never by a mass noun, which would render ‘perform labor’ incorrect in English (Hill 1962:29). When pressed for evidence that his hypothesis was correct, Chomsky merely responded that he was a native speaker of English. Chomsky’s intuition about the possible predicates of ‘perform’ resulted in a hypothesis. Hypotheses are indeed borne of intuition. However, his intuition at the same time provided him with the supporting evidence for the hypothesis; he saw no need to consult a corpus to confirm it. [fn.2: A quick search on Google yields thousands of cases of ‘perform labor’. Sampson (2001) discusses a similar case in which personal introspection led

---

<sup>34</sup> Devitt (2006) argues that the metalinguistic intuitions of professional linguists are more informed, and hence more reliable, than those elicited from untutored speakers. This claim has been challenged by Culbertson and Gross (2009). The challenge has spawned a fruitful debate; see Devitt (2010a) for a reply, and Culbertson and Gross (forthcoming) for a follow-up. I do not wish to embroil myself in these issues. The claim in the main text is merely that linguists are more likely to have learned about the varieties of performance error and hence more reliable than non-linguists in evaluating whether they are subject to it in some particular case. This is especially true when the judgments are subtle or fine-grained, as they are in the case of, say, degrees of subadjacency violation. Though much of contemporary syntactic theory is based on stable and widely shared judgments (see Sprouse and Almeida [forthcoming] for details), there are isolated cases on which the judgments are quite subtle or fine-grained. Occasionally, such cases turn out to be pivotal for settling major theoretical disputes. As such, they deserve further exploration by all parties to the debate over the relative merits of tutored and untutored linguistic intuitions.

to the idea that central embedding was impossible, when many cases are found in actual language data.]” (p. 8).

For our purposes, the important aspect of this anecdote is that Chomsky replied to the challenge by pointing out that that he is a native speaker of English. What, we might ask, did he mean to convey with his remark? One interpretive possibility that we should reject immediately is that Chomsky meant that he has his own I-language, and no one could refute him about *that*. This interpretation conflicts directly with the fact that Chomsky takes great pains to emphasize that competent speakers of a language can be wrong about their own language and that introspection is not a good (let alone an optimal) tool for the construction of linguistic theory. (See, e.g., *Aspects*, ch. 1.) Charity dictates that we see Chomsky as taking his generalization about the distribution of count nouns and mass nouns to be open to empirical disconfirmation. How might such disconfirmation proceed? Eddington takes the results of a Google search—a very crude sort of corpus analysis—to be evidence against Chomsky’s generalization. There are two ways in which such data might play that role, corresponding to two views of the nature of linguistic theorizing.

On one view, Chomsky was offering a generalization about only one person’s I-language—his own. In this case, the results of Eddington’s crude corpus analysis would disconfirm the Chomsky’s generalization only in the event that Chomsky, himself, looked at them and saw that he perfectly well understands what is meant by ‘perform labor’ and that he finds the phrase either fully acceptable or fully grammatical. The opposing view, by contrast, has it that Chomsky was offering a generalization about a *dialect of English*—possibly, though not necessarily, his own. In this case, the results of Eddington’s crude corpus analysis would refute the generalization outright, regardless of Chomsky’s intuitive judgments—indeed, regardless of whether he ever makes any. Now, we might ask: What’s more plausible? That

Chomsky's theories of syntax have always been about—and *only* about—his own I-language, or that the theories are about the dialect of some reasonably large community of competent speakers? While it seems to me that the latter position is more probably true, I cannot claim to have shown this. Again, all I wish to establish here is that the opposite case must be made.

It may be objected that Chomsky's insistence on the study of I-language does not carry a commitment to the claim that linguists study *some particular speaker's* I-language. Rather, the insistence on studying I-language should be taken as a commitment to either (i) the study of distinctively *psychological* states and processes or (ii) the propriety of a "linguistic theory ... concerned primarily with an ideal speaker-hearer in a completely homogeneous speech-community, who knows its language perfectly" (Chomsky, 1965: p. 3). Let us take these in turn.

Regarding (i), the following point is, I think, crucial: The interest in psychological states and processes is *not in the least bit diminished* by the recognition of the existence of public language, nor by the recognition that the syntacticians can be easily seen as formulating principles of the grammars of dialects rather than I-languages.<sup>35</sup> As we will see in ch. V, the relation between syntactic theory and psycholinguistic models of parsing can be quite complex. A grammar formulated by a syntactician who is either committed to studying E-language or is agnostic about the E-language/I-language distinction and the attendant controversies, can be just as fertile a source of hypotheses for the psycholinguist and the computational modeler as a

---

<sup>35</sup> This is an echo of the point that Devitt (2006) makes concerning the compatibility between adopting his Linguistic Conception and studying knowledge of language. Devitt writes: "[M]y contemplated task must be worthwhile if Chomsky's task (i) is. [Task (i) is answering the question: What constitutes knowledge of language? D.P.] For, although we have distinguished the two tasks we have also related them in a way that makes completing the contemplated task necessary for completing task (i). For, the nature of the speaker's competence studied by task (i) involves the nature of the symbols studied by the contemplated task: those symbols are what the competence produces. Indeed, our earlier discussion (2.3) shows that the contemplated task has a certain epistemic and explanatory priority over task (i). How could we make any significant progress studying the nature of competence in a language unless we already knew a good deal about that language? Just as explaining the bee's dances is a prerequisite for discovering how the bee manages to produce those dances, so also explaining the syntax of sentences is a prerequisite for explaining how speakers manage to produce those sentences" (pp. 28-29).

grammar designed specifically to be “about” I-language. More generally, as noted earlier, the issue regarding whether syntacticians are studying E-language or I-language is *orthogonal* to the issue of whether some grammar or other is psychologically real. It may well be that syntacticians’ grammars are, in the first instance, *about* E-language, but that one or another such grammar can subsequently be shown, on the strength of psycholinguistic evidence to be mentally represented and used for comprehension and production (chs. III-V).

Regarding (ii), there is this to say: We must be open to a theorist’s wish to study a phenomenon by making idealizations—a strategy that is both fruitful and ubiquitous in the natural sciences. A problem does arise, however, if the theorist either mistakes an idealization for reality or if he denies the existence of the very reality that he is idealizing away from. In the case of linguistics, the reality is that there are communities of speakers whose linguistic performance is susceptible to powerful generalizations. For instance, we can say truly of a great many speakers that their uses of the pronoun ‘him’ in the sentence ‘Bill said that John loathes him’ are *not* references to John (assuming, of course, that John  $\neq$  Bill).<sup>36</sup> Such generalizations may, on occasion fail to apply to specific speakers, on account of one or another quirk of individual psychology, education, or other factors. The purpose of idealization, then, might be to exclude such cases as not being indicative of the nature of the public language shared by the community. It need not be seen as a commitment to the study of each community member’s I-language, let alone any one specific I-language—a conception that is neither mandatory nor even a particularly natural way of interpreting the practices of working linguists.<sup>37</sup> Establishing it

---

<sup>36</sup> Needless to say, even more general and theoretically robust claims can be made—generalizations that abstract away from the use of the names ‘Bill’ and ‘John’, and away from the fact that this sentence has only one embedded clause. I have in mind, of course, the generalizations of Binding Theory and its successors.

<sup>37</sup> For further discussion of the issues broached in this section, see Devitt (2006), chapters 2, 7, 8, and 10.



requires an argument that I suspect will be difficult to come by. The difficulties, moreover, are not limited to the ones I have adumbrated thus far. In the next section, I will argue that the project of individuating I-languages faces a number of challenges that are often overlooked by philosophers of linguistics.

### §2.3 *Individuating I-languages*

While I-languages are not themselves collections of text or verbal behavior—the outputs or products of linguistic competence—such things must, of course, serve as *evidence* for the ascription of an I-language to any specific individual, as well as for any claims regarding the individuation of I-languages. A linguist who wishes to write down a grammar for a specific I-language has essentially two options: She can either survey a large corpus of the target individual’s speech, writing, and signing, or she can ask that individual to provide intuitions about the grammaticality of large class of preselected utterances, inscriptions, and signs.<sup>38</sup> For expository convenience, let’s call the target individual ‘Ana’, and let’s give the label ‘utterance’ to any item in the class of Ana’s spoken, written, and signed expressions.

Suppose the linguist decides to go the corpus route. Immediately, certain decisions have to be made about which of Ana’s utterances to keep in the corpus and which to throw out as “bad data”—false starts, utterances made while chewing or yawning, drunken babble, capricious or impulsive mid-sentence switches of topic, interrupted speech, utterances made in the heat of violent passion or overwhelming emotion, careless typos, unforeseen tongue-twisters, spoonerisms, mixed metaphors, and so on. The corpus must also be divided up in a way that reflects the fact that Ana is very likely to switch between a number of dialects, or even be

---

<sup>38</sup> Tellingly, it’s rare to see anyone undertaking the project of individuating an I-language. Linguists typically study individual subjects only when they have some very special property, such as a rare linguistic or cognitive disorder.

multilingual (as the “unreflective folk” would put it). The linguist would, in addition, have to take into account changes in Ana’s speech over long periods of time. For instance, suppose that it’s plain that Ana’s college education effected a radical change in her speech and writing. One might argue, of course, that the education merely boosted Ana’s *performance*, but had no effect on her competence grammar. Though initially tempting, this view raises more questions than it answers. Suppose, for instance that, prior to entering college, Ana consistently rejected sentences of the following form—i.e., she never produced them and, when prompted, judged them strongly unacceptable.

(9) That the man went is a truth.

(10) That we have come to a firm decision is unclear.

(11) That you eat is a fact.

Suppose, further, that Ana gradually became increasingly comfortable with such sentences as she progressed through her college career, balking less frequently or intensely at their appearance in textbooks. We can imagine Ana eventually coming to use such sentences in her writing—at first with trepidation, but later confidently—and, later, in her speech, though only in the rare cases where her conception of the relatively narrow rhetorical register allowed for it. While it’s clear that inebriation and temporary inattentiveness can breed performance errors, it’s not *at all* clear whether Ana’s initial attitudes toward CP-topicalization constitute a performance error.

One might argue that Ana’s I-language prior to enrolling in college consistently allowed other types of topicalization, and that a maximally simple grammar of that I-language would also allow CP-topics. But whatever gains in simplicity this move affords must be carefully weighed against the complexities that must then be introduced into the performance model to support it. Specifically, we would now need to posit a mechanism somewhere in Ana’s linguistic processing routines that plausibly accounts for her initial inability or reluctance to understand, produce, and otherwise accept only *some* types of topicalization, but not others. Perhaps there is an optimal

way of striking a balance between these competing pressures. But it must be admitted that any such judgment regarding relative simplicity of otherwise empirically adequate theories is bound to be extremely subtle.

From this, we can draw two morals. First, the playing field on which the notions of E-language and I-language compete must be a level one. In particular, the theorist who is bent on constructing or refining a notion of public dialects, public languages, or E-languages should be allowed to deploy heavily theoretical resources in the course of her efforts. After all, her counterpart, the I-language theorist, is allowed to appeal to the competence/performance distinction and to employ a (more or less implicit) methodology for drawing that distinction in difficult cases. Second, when the E-language theorist encounters cases in which her theoretical decisions are less than principled—i.e., recognizable as more-or-less arbitrary judgment calls—we should *not* conclude that her project is fundamentally misguided. For, our standards are simply not that high when the topic is I-language. Provisional decisions, even when somewhat arbitrary, are tolerated in isolated difficult cases, if only in the hope that more principled lines can be drawn in the fullness of time, when theoretical progress affords us with conceptual and experimental tools to address more subtle issues.

The friend of crisp I-languages may, at this point, object that we have been unfair to his position. I-languages, he says, are only determinate *at a time*; talk of changes due to education and the like is simply beside the point. But problems loom even for this view. How fine-grained, exactly, is the individuation supposed to be? Does Ana wake up with a new idiolect every day? Does her I-language change from one hour to the next? One minute to the next? One second?

Take a related worry: Suppose Ana is bilingual—a competent speaker of what the folk would unreflectively label ‘Russian’ and ‘English’. Calling Ana bilingual or multi-lingual *hints* at a public conception of language, because without reference to Ana’s similarity to other people—that is, without reference to two separate *languages* or *dialects*—it makes more sense to classify her as *monolingual* (or mono-idiolectal). That is, although Ana would be classified as bilingual by ordinary standards, the position we are now considering entails that her I-language includes *all* of the things that she is disposed to produce, comprehend, and evaluate (again, performance errors aside, whatever that comes to). Thus, given that her I-language includes what we would ordinarily call the English and the Russian language, the crisp-idiolect view has it that Ana’s I-language is best thought of as *Runglish<sub>Ana</sub>*. This view, however, faces a challenge from the opponent who asserts that Ana actually has *two* I-languages, simultaneously—call them *Russian<sub>Ana</sub>* and *English<sub>Ana</sub>*—and that she draws on one of these for some purposes and on the other for others. How to settle this dispute?

Looking at the relevant empirical research, we discover the well-documented fact that polyglots code-switch at the drop of a hat (Gardner-Chloros, 2009). This is especially easy to observe in urban environments. Sitting on a New York subway train, I recently observed a man who, in the course of a brief conversation on his cell phone—the length of about 4 stops on the Q line—switched effortlessly between English, Russian, and Hebrew, producing over a dozen sentences that contained lexical items from all three languages. The literature on multilingualism suggests that this phenomenon, far from being uncommon, is actually the norm in a vast range of human cultures (Auer and Wei, 2007). The prevalence of monolingualism is a relatively recent phenomenon, and even in the present day is confined to social groups that have rather peculiar socio-economic relations to the rest of the human population.

With this in mind, consider asking my fellow Q-train passenger what he thinks of the following sentence.

(12) Я решил definitely принести some wine to the משפחתית tonight.

The trouble starts well before the question is asked, for it is no small matter to decide what the question should even *be*. Asking whether the sentence is acceptable in *his* language presupposes that he has only one. But he may well reject that presupposition, either out of hand or upon careful reflection, on any number of grounds. He may cite, for instance, the fact that *only* Hebrew—not any of his other languages—is used for religious practices. Equally, he may bring up the fact that he remembers phone numbers and performs basic calculations only “in English,” because that’s what he finds easiest. Finally, he might trot out his distinctive preference for swearing in Russian—a language whose agglutinative morphosyntax seems positively designed for the productive generation of ever more intricate forms of profanity. He might add that his preference for Russian swear words is especially pronounced when he is drunk or embroiled in a heated situation. Having brought to bear a wealth of psychological and socio-cultural considerations, the man would surely be puzzled at the linguist’s suggestion (let alone insistence) that, strictly-speaking, he knows only *one* language, “in the technical sense.” And yet the corpus of his utterances contains (12) and a great many such mongrels. Are these performance errors? Do we know how to tell?

One might hope to find the answers to these questions in the field of second-language (L2) acquisition—an area of study whose very title betrays a folksy commitment to either public languages or multiple I-languages residing in one speaker. Unfortunately, there is widespread disagreement amongst theorists who study L2 acquisition about the boundaries between internalized grammars. In some cases, it seems that there is a “core” grammar that is being made

to do double duty, as the speaker deploys it in the course of using “more than one language”—a phenomenon that falls under the heading of *L1 transfer*. In other cases it’s just not clear *what* to say. The questions are, as always, empirical, which warrants the hope that they will one day be resolved on principled grounds. But, in practice, the data that are brought to bear on such issues are often subject to conflicting interpretations, all of which appear antecedently plausible—a situation that requires suspending judgment pending further clarification.

Again, I urge that conceptions of E-language be allowed the same leeway. At present, they are often summarily dismissed by Chomsky and his followers as being too vague and ill defined to be fit for serious empirical inquiry, as though the individuation of I-languages is, by comparison, a straightforward and uncomplicated matter. This double standard is, by my lights, both unfounded and obstructive to progress. Making precise the individuation conditions on I-languages is an ongoing research program. We should view the project of individuating E-languages in the same way, rather than expecting that the friend of E-language will be able to provide crisp and clear definitions in advance of sustained inquiry.

Returning to the thought experiment, suppose that our linguist, undaunted by the difficulties adumbrated above, persists in his quest to divide up Ana’s utterances into exactly two crisp I-languages. He might begin by sorting her lexical items into those that belong to the I-language Russian<sup>Ana on May 19th, 2010, 5:44:06 PM</sup> and those that belong to the I-language English<sup>Ana on May 19th, 2010, 5:44:06 PM</sup>. How might this be accomplished?

One natural suggestion would be to take stock of the similarities and differences in the *phonological* features of the lexical items. But if Ana has made it a point to pronounce “correctly” certain “foreign” words (including names)—in the way that many people recently did with the name of Icelandic volcano, Eyjafjallajökull—then this procedure would not yield a

binary partition. Suppose Ana, like many of us, regularly uses the following names and expressions: *Amen, je ne sais quoi, c'est la vie, que sera sera, gesundheit, chutzpah, gracias, enfant terrible, mea culpa, verboten, machismo, vis-à-vis, a priori, prima facie, Gödel, and Wagner*. It would be rash to insist that these are not part of her lexicon. At the same time, it would be pointless to increase the number of I-languages that we attribute to Ana, simply on the basis of the relevant phonological considerations. The issue becomes even more acute if we learn that, in the course of pursuing her acting career, Ana has fully mastered over a dozen English accents.<sup>39</sup>

Another proposal involves individuating lexical items by reference to the syntactic frames in which they are licensed. But this, too, runs into problems. Consider (13).

(13) Я решил definitely принести some wine to the party tonight.

Is the word ‘definitely’ *licensed* in (13)? Should ‘принести’ (transl. *to bring*) really be ‘to принести’, or is the sentence grammatical as it stands? Is (13) even a *sentence*—i.e., a string generated by a single, consistent grammar? If so, which grammar is that?

The difficulties that we have catalogued thus far can be multiplied *ad nauseam*. Of course, this is not to say that there is absolutely no way of resolving them; perhaps some of them can be resolved on principled grounds, while others are handled legalistically or—in suitably isolated cases—wholly arbitrarily. My aim, rather, has been to point out that, for many theoretical pursuits, the precision may not be worth the effort. One might well simply idealize away from the messy variation.<sup>40</sup> And, as always, the nature of the idealization will depend on

---

<sup>39</sup> For a concrete example, see <http://www.youtube.com/watch?v=3UgpfSp2t6k>

<sup>40</sup> Ludlow (2011: pp. 44-46) makes passing reference to many of the difficulties that I have discussed here. However, he comes to the conclusion that they constitute problems *only* for the friends of E-languages—particularly E-*idiolects*. Having argued that these problems are devastating to the enterprise of E-linguistics, he claims that the very same problems constitute a “research program” for I-linguistics (p. 46). The suggestion seems to be that an

one's theoretical purposes—in this case, the purposes to which a notion of I-language will be put.<sup>41</sup>

For some purposes, the identity conditions on I-languages can be made strict enough that a suitable proportion of Ana's lexical items and utterances are treated as though all they fall neatly into exactly two classes and obey exactly two sets of grammatical principles. Ana can thus be said to have exactly *two* I-languages (on May 19<sup>th</sup>, 2010, 5:44:06 PM). Questions would remain, of course, about whether the project of individuating those I-languages can really avoid making reference—even *tacit* reference—to larger social groups or cultural constructs (§2.1). Likewise, the exact import of the competence/performance distinction would need to be specified, in more fine-grained terms than one often finds in the literature on I-languages.

On the other hand, the identity conditions on I-languages can be left so loose that Ana would count as having only *one* I-language. Perhaps we say this simply because she has only one body, or one brain, or—in the limit—only one *self*. (Of course, individuating *these* entities is no small feat either, as generations of philosophers have come to realize.) Another possibility is to adopt the formula: One I-language per language faculty. But this invites still other kinds of

---

analogous research program in the case of E-language would be unprincipled or theoretically fruitless. But this double standard begs the question against a position like Devitt's, according to which contemporary linguistic theory—a fruitful research program—is actually *about* E-languages. Moreover, Ludlow seems to ignore the fact, which I stressed in a previous section, that a theorist who wishes to study E-languages is entitled to the same sorts of idealizations that animate inquiry into I-languages. Such idealizations serve precisely to abstract away from the problematic types of inter- and intra-speaker variation that Ludlow has in mind.

<sup>41</sup> This point has an analogue in the study of E-language. Devitt (2006) points out that linguistic theory—conceived as an inquiry into the nature of E-language—can proceed at various grains of analysis. In identifying the group of speakers that she wishes to study, the linguist can, for the most part, make do with a term like 'English', "but sometimes [she] need[s] a less precise term like 'Romance' and sometimes a more precise one like 'Australian-English' or 'the Somerset dialect of English' (p. 184).



trouble. Individuating language faculties is itself something of a dark art, as cases of brain damage, bilingual aphasia, agraphia, and alexia serve to highlight.<sup>42</sup>

### §3 Residual Issues and Concluding Remarks

In the pages above, we considered some of the main arguments that constitute the debate between cognitivists like Chomsky and their most recent nominalist opponent, Michael Devitt. Having laid out the central features of both positions, I went on to argue for two claims. In §1, I argued that adopting the very reasonable aim of explanatory adequacy does *not* require a prior commitment to the cognitive conception of linguistics. Explanatory adequacy need not be characterized by appeal to the psychological process of language acquisition. It can be seen, instead, as a reflection of the premium that scientists place on discovering powerful generalizations and offering unifying explanations. Thus, I concluded, one standard line of reasoning in favor of the cognitive conception misses its mark. For all the argument shows, the cognitive conception is strictly *optional*.

In §2, I addressed the cluster of issues surrounding Chomsky's distinction between E-language and I-language. I argued that the notion of E-language might very well be playing a role in the scientific study of language—specifically, in the project of specifying the grammars of the public languages to which a child is exposed in the course of acquisition. A close look at the literature on language acquisition illustrates that such grammars appear to serve as “targets” against which acquisition theorists measure a child's successes and failures throughout development. Specifically, the psychologist's quantitative descriptions of children's patterns of error seem to make ineliminable reference to such grammars. It remains an open question

---

<sup>42</sup> See also the discussion in chapter 10 of *Ignorance of Language*, where Devitt summarizes the results of various neurocognitive studies and draws on them in casting doubt on the unity of the language faculty.

whether such apparent references to public languages can be paraphrased away, e.g., as oblique references to some particular I-language, whether real or idealized. If it cannot, then these considerations will be seen to dovetail with the arguments in Devitt (2006), who concludes that public communicative conventions and belief-ascription practices serve, in the first place, to identify the object of the linguist's inquiry.

There remain a number of issues that divide the friends of E-language from those who follow Chomsky in rejecting the relevance of that notion to linguistic theorizing. One of these is what we might call "the primacy issue." Chomsky (1988) claims that any theory of E-language will "surely have to presuppose grammars of I-languages." In his view, whatever an E-language might be, it is "more abstract" than I-language, on account of its being "more remote from mechanisms" (p. 561). But one might reasonably challenge Chomsky's presumption that the psychological mechanisms of underlying any individual's language use are describable without reference to (apolitical) social facts—without, e.g., a prior statement of the grammar that best captures the conventional language of a particular linguistic community.<sup>43</sup> Several of my arguments in §2 contribute to making the case that references to E-languages play a role in individuating I-languages, and in explaining how they are acquired.

It may be, then, that the notion of E-language has primacy over that of I-language—the exact reverse of the situation that Chomsky envisages. Alternatively, it might be that the two notions are, as it were, coeval and hence on a par with one another. The arguments that I have offered here do not settle which of these two positions is more promising. To do this, one would have to, at a minimum, distinguish between a variety of primacy theses, including those having

---

<sup>43</sup> Needless to say, an anatomist might be able to discuss the underlying neurophysiology of these mechanisms. But that is not what Chomsky has in mind. It's the *psychological* level of description that concerns him, also the level to which the E-linguist's argument is addressed.

to do with the temporal, ontological, phylogenic, ontogenic, explanatory, evidential, and conceptual relations between E-language and I-language.<sup>44</sup> The primacy issue, then, deserves much more discussion. The same can be said of issues pertaining to the contrast between knowledge-how and knowledge-that, and between intensional and extensional conceptions of grammar. I have not dealt with these latter two issues in any detail thus far, but my conclusions in the subsequent chapters will bear strongly on them. By way of foreshadowing, let me introduce them here.

In ch. 6 of *Ignorance of Language*, Devitt argues that Chomsky's cognitive conception of language derives no support from the folk view that we *know* a language, for this is compatible with the folk conception of knowledge-how—the key ingredient in having *skills* or *abilities*. He argues, furthermore, that linguistic competence is indeed a skill, which may or may not be underpinned by a store of declaratively represented knowledge-that. In ch. 11 of *Ignorance*, on the basis of an extensive survey of the literature on the psychology of skills, Devitt goes on to argue that language production and comprehension does *not* require psychological mechanisms to draw on mental representations of syntactic principles. He concludes, then, that grammars are not psychologically real. This, in turn, has a bearing on the issues pertaining to the contrast between intensional and extensional conceptions of grammar.

The types of evidence that would distinguish extensionally equivalent grammars often have to do with psychological processing in an individual—i.e., the performance phenomena involved in acquisition and use.<sup>45</sup> Hence, an intensional conception of grammar is only

---

<sup>44</sup> Devitt (2006) draws these distinctions in the course of his discussion of the relations between language and thought—a related but distinct issue. In ch. 4 (§3.1), I discuss challenges to the claim that thought is explanatory prior to language.

<sup>45</sup> There is an exception to this rule—viz., evidence from diachronic linguistics, where historical studies have yielded generalizations about the systemic character of language change. These generalizations can serve as

appropriate if one has antecedently established that grammars are psychologically real and that they play a computational role in performance phenomena. Devitt puts the point as follows.

[Consider] a popular criticism of the claim that a grammar's rules are psychologically real. The criticism is that we lack evidence as to which grammar's rules are psychologically real: "If we can come up with one grammar for a language, we can come up with many which, though they posit different syntactic rules, are equivalent in their explanation of meaning: they are equally able to capture all the syntactically determined facts about meaning. We need psycholinguistic evidence to show which grammar's rules are in fact playing the role in linguistic processing, evidence we do not have." This criticism is not quite right. We need evidence that the syntactic rules of any grammar are processing rules. These rules may simply be the wrong sort of rules to be processing rules.... Suppose, as seems quite likely, that the human language capacity is an adaptation. Then nature faced the problem of designing this capacity out of pre-existing structures in our ancestors. We should not think, in advance of empirical discovery, that nature solved this problem by making humans go through processes governed by linguistic rules. We should not suppose a priori that a correct account of the linguistic reality will describe the psychological reality. A grammar may have nothing more to do with psychological reality than comes from its meeting the Respect Constraint. (p. 37)

And, as mentioned above, Devitt argues in ch. 11 that grammars are *not*, in fact, psychologically real. Hence, if it turns out that his Linguistic Conception of public-language grammars does not allow us to distinguish between extensionally equivalent formalisms, it is not at all clear that he would be bothered by this finding, nor why he should be.

Suppose, though, that one *were* bothered. Consolation would not be far to seek.

First, let us get clear on about exactly what extensional equivalence amounts to in this context. Chomsky and his followers often presuppose that two grammars are extensionally equivalent when they are both observationally adequate. But the proponent of the extensionalist conception of grammars need not be restricted to observational adequacy. The standard considerations of theoretical simplicity, generality, and explanatory unity can all be invoked by

---

constraints on grammar choice. One of two extensionally equivalent grammars might provide a more natural explanation of the systematic changes than its competitor.

the extensionalist to motivate the adoption of one observationally adequate grammar over another, *without* reference to any data concerning reaction times, error rates, gaze duration, neuronal activity, and the like. For a lucid example of this strategy, see Fodor, Bever, and Garrett (1974: ch. 3) and, indeed, Chomsky (1957). Both works argue for the superiority of transformational grammars to pure phrases structure grammars—what Fodor, Bever, and Garrett call Immediate Constituency (IC) grammars—on the grounds that transformational grammars are intuitively simpler, easier to work with, and capable of stating more powerful generalizations. Moreover, as I argued in §1, these very same considerations can be invoked *once more* to motivate the adoption of one *descriptively* adequate grammar over another, again without appeal to any results from experimental psychology. This, I take it, is what happened in the transition from the Extended Standard Theory to the Principles and Parameters framework. Many would argue that this transition was motivated by concerns having to do with the psychology of language acquisition. By contrast, I would argue that the simplicity, generality, and explanatory power of the P&P framework was *sufficient* to demonstrate its superiority over EST.

It follows that, even if public E-languages constitute the subject matter of formal linguistics, there is no reason to believe that intensional distinctions could not affect our decisions about which grammar to adopt. The study of public E-languages is, of course, bedeviled by the underdetermination of theory by evidence—the problem(s) of induction, essentially—but that is no less true of any other empirical enterprise, including the study of I-languages. Putting aside that set of worries, the study of public E-languages is no less constrained by purely extensional criteria than the study of I-languages.

Whatever the case about that, I will attempt to show in chs. III-V that Devitt's arguments against the psychological reality of language do not succeed. My argument will rely heavily on

empirical findings from psycholinguistics—specifically, studies of on-line comprehension. In chs. III and IV, I will argue that the human sentence processing mechanism (HSPM) constructs metalinguistic representations of the syntactic structures of incoming linguistic stimuli. Any psychologically plausible model of the HSPM must, then, describe the mechanisms that compute such representations. In ch. V, I will demonstrate that such models typically posit either declarative *representations* of one or another grammar, or treat grammars as procedural dispositions to compute specific metalinguistic representations in specific circumstances. I will argue that both approaches can be seen as vindications of the claim that grammars are psychologically real. Moreover, I believe that the kinds of behavioral and neurocognitive data that I will survey in the course of my argument can be used to individuate mental grammars *intensionally*.

While all this is very much in line with the position that I have attributed to Chomsky in the course of this discussion, I do not believe that my conclusions vindicate the cognitive conception of language. As mentioned above, the debate about how to conceive of the objects of linguistic theorizing is, in a certain sense, orthogonal to the psychological reality debate (though it plainly bears on it). As I see it, the logical relationship between the two debates is this: If the cognitive conception is correct, then grammars are trivially psychologically real, for the reasons that Chomsky gives (ch. I, §3). However, the reverse entailment does *not* hold; grammars can be psychologically real even if Devitt’s “linguistic realism” is correct and the objects of the formal syntactician’s concern are public, conventional E-languages. As noted above, it may well be that syntacticians’ grammars are, in the first instance, *about* public E-languages, but that one or another such grammar can subsequently be shown, on the strength of psycholinguistic evidence to be mentally represented and used for comprehension and production (chs. III-V).

Let me conclude, then, by addressing one final question. What should we make of the fact that the so many linguists in the generative tradition would balk at the mere suggestion that their object of study is something other than the mind/brain? To answer this question, I propose to revive an old distinction, familiar from the philosophy of science, between the *context of discovery* and the *context of justification*.<sup>46</sup> My hunch is that syntacticians working in the Chomskyan paradigm are often inspired to formulate their hypotheses about the structure of language in the course of self-consciously tackling questions about language acquisition. Such hypothesis formulation takes place within the context of discovery. Nevertheless, if we focus instead on the context of justification, then we find that what we currently know about language acquisition is neither necessary nor sufficient to establish the truth of those hypotheses. I submit that the core “modules” in many contemporary P&P grammars—the projection principle, X-bar theory,  $\theta$ -theory, Case theory—would have been just as illuminating to the formal syntactician even if (contrary to fact) we *knew* that language acquisition is a matter of purely statistical pattern recognition in a connectionist network.

Focusing on a specific example, consider the motivations for preferring binary-branching phrase structures, like those we find in X-bar theory (ch. V, §5). Haegeman (1994) surveys a large number of purely syntactic arguments for the X-bar formalism—i.e., arguments that appeal to the results of the familiar battery of constituency tests, such as pro-form replacement, topicalization, and coordination. And she notes that binary-branching grammars rank high in simplicity and generality. But she goes on to argue that binary-branching phrase structures are *also* preferable from the perspective of the acquisition theorist.

---

<sup>46</sup> This distinction has fallen on hard times—a case of guilt-by-association with the discredited epistemology of the logical positivists. But I am hopeful that there is a way to hive it off from positivist philosophy. Doing so would allow us to capture an intuitive difference between the psychological processes that lead theorists to advance hypotheses and the arguments that are later used to support those hypotheses in textbooks and the like.

In the course of this chapter, the change from [flat] structure to [binary-branching X-bar] structure was motivated on empirical and theoretical grounds, but there are further advantages to adopting a grammar which allows only the second type. The reader may notice that such a grammar is more aesthetically satisfying, though aesthetics may be a minor preoccupation for linguists. A grammar which allows only binary branching nodes is more constrained than a grammar which freely allows any type of branching node: in the former type of grammar lots of imaginable representations are ruled out in principle. A more constrained grammar is preferred for reasons of economy and elegance and it will also be preferred if we think of the ultimate goal of linguistic theories in the generative tradition. Remember that linguists wish to account for the fact that children acquire language very fast and at an early age. In order to explain their fast acquisition we posit that children are genetically prepared for the task, that they have an innate set of principles which enable them to construct the core grammar of their language on the basis of the data they are exposed to. One component of the child's internalized knowledge of the language, the internal grammar, will concern phrase structure. Theories of phrase structure such as X'-theory attempt to represent the native speaker's internal knowledge of phrase structure. ... A child equipped with a UG that implements only binary branching will have fewer decisions to make when assigning syntactic structure to the data he is exposed to than a child equipped with a less constrained UG which allows ternary or four-way branching. ... If the ultimate goal of our grammar is to account for language acquisition, then it will be natural to aim for the more restricted type of grammar in which fewer decisions have to be made by the child. Fewer choices will automatically mean more speed in the construction of the core grammar of the language acquired. Nowadays most linguists working in the generative tradition tend to adopt some version of the binary branching framework. (Haegeman, 1994: pp. 138-143)

Haegeman's argument in this passage assumes that we know quite a bit about the actual processes underlying language acquisition. Unfortunately, the fact is that relatively little is known about this. In particular, we simply do not know whether, given the child's *actual* learning strategy, a binary-branching grammar presents a substantially easier or more tractable computational problem for the learner. Our present-day understanding of the computational architecture of the brain is far too limited for us to be justified in making firm pronouncements on the matter—much less invoking such pronouncements as premises in arguments for substantive conclusions about the structure of the language. This is not, of course, to suggest that substantial work has not been done in acquisition theory, or that important results have not



been garnered (Guasti, 2002). It is merely to point out that acquisition models do not, at present, dictate a choice of grammar; the situation is quite the reverse.<sup>47</sup>

Given all this, it would be a bit naïve to claim that *the* reason GB Principles replaced EST transformational rules in the 1970s was that former provided a much more compact and parsimonious way of capturing known generalizations and novel data. Talk of “*the* reason” elides the distinction between the contexts of discovery and justification (and, for that matter, the further factor of persuasion). But it is correct, I think, to say that no more is needed to justify the shift from EST transformations to GB Principles than this: The proliferation of transformational rules was extremely ugly, while the appeal to a handful of principles is elegant and satisfying. In particular, no reference need be made to controversial issues in acquisition theory.

While many linguists in the generative tradition are either unaware of these considerations or simply unimpressed by them, there *are* a handful of theorists who take them quite seriously, and pitch their inquiry and their results accordingly. The most prominent example, both in terms of impassioned rhetoric and substantial theoretical results, is the group of linguists who first proposed the feature-based unification formalisms of Generalized Phrase-Structure Grammar (GPSG) and Head-Drive Phrase-Structure Grammar (HPSG). In laying out

---

<sup>47</sup> There is a strange double standard in Berwick and Weinberg’s treatment of this issue. In the opening pages of the 2<sup>nd</sup> chapter of their classic text, *The Grammatical Basis of Linguistic Performance* (1984), they make the following reasonable point about how parsing theory typically fails to constrain grammar construction: “We ought to be able to recruit sentence processing results to tell us something about what the grammar should look like. If we had some independently justified parsing model, we could reject grammars that were incompatible with it. In practice, though, because very little is known about the details of the syntactic parser, confidence in constraining the choice of grammatical theory via this route must be correspondingly weak. If the parsing theory has no independent motivation, we can always change it to suit the grammatical format” (p. 36). However, they go on to suggest that our understanding of the acquisition process is more detailed and hence provides more firm constraints on grammar constructions. But surely, their point the lack of fixed points in parsing theory holds just as well with regard to acquisition. Indeed, the only characteristics of acquisition that B&W bring to bear in their introductory chapter are these: (i) acquisition is accomplished in about 5 years, and (ii) the input to it is variable, partially ill formed, and rarely (if ever) provides negative evidence. This is not more than what we know about parsing. Yet these considerations were marshaled, in their first chapter, as the justification for profound shifts in syntactic theory. If my claims in the main text are correct, the purely formal and superempirical arguments that B&W offer in favor of a GB-style approach are by themselves sufficient to motivate it. Their appeals to acquisition were superfluous.

their metatheoretic commitments, these linguists make absolutely clear that, while the results of their inquiry may well inspire or suggest specific hypotheses for psycholinguistic research, this would be but a serendipitous bonus. The suggestion seems to be that the viability or superiority of GPSG (and later HPSG) is established on grounds that have only a tenuous connection with psychology. I quote their impassioned rhetoric at length.

In view of the fact that the packaging and public relations of much recent linguistic theory involves constant reference to questions of psychology, particularly in association with language acquisition, it is appropriate for us to make a few remarks here about the connection between the claims we make and issues in the psychology of language. We make no claims, naturally enough, that our grammatical theory is *eo ipso* a psychological theory. Our grammar of English is not a theory of how speakers think up things to say and put them into words. Our general linguistic theory is not a theory of how a child abstracts from the surrounding hubbub of linguistic and nonlinguistic noises enough evidence to gain a mental grasp of the structure of a natural language. Nor is it a biological theory of the structure of an as-yet-unidentified mental organ. It is irresponsible to claim otherwise for theories of this general sort. It may even be incoherent... Thus we feel it possible, and arguably proper, for a linguist (*qua* linguist) to ignore matters of psychology. But it is hardly possible for a psycholinguist to ignore language. And since a given linguistic theory will make specific claims about the nature of languages, it will in turn suggest specific kinds of psycholinguistic hypotheses. Stephen Crain and Janet Fodor have argued that [Generalized Phrase-Structure Grammar] does have implications for psycholinguistic concerns. Nonetheless, it seems to us that virtually all the work needed to redeem the promissory notes linguistics has issued to psychology over the past 25 years remains to be done. If linguistics is truly a branch of psychology (or even biology) as is often unilaterally asserted by linguists, it is so far the branch with the greatest pretensions and the fewest reliable results. The most useful course of action in this circumstance is probably not to engage in further programmatic posturing and self-congratulatory rhetoric of the sort that has characterized much linguistic work in recent years, but rather to attempt to fulfill some of the commitments made by generative grammar in respect of the provision of fully specified and precise theories of the nature of the languages that humans employ. Even when that is done, the psychology of language will doubtless have a vast amount of work to do before we have a scientific understanding of how the human species acquires and uses language. After all, geometrical optics long ago provided us with a fairly clear and stable means of characterizing the objects of visual perception, but the psychology of visual perception still has many problems to solve. So far, linguistics has not fulfilled its own side of the interdisciplinary bargain (Gazdar, Klein, Pullum, and Sag, 1985: p. 5).

These ideas are strongly echoed in the remarks of Black and Chiat (1981).

Nothing would be changed if instead of distinguishing between different types of knowledge, we simply distinguished between different systems of rules and representations without making any assumptions about their psychological status. If we agree that it is legitimate to talk about abstract structures and properties, though they may not be identifiable with any particular physical property of strings of symbols or sounds, then the addition of terms like ‘knowledge of’ to grammar and ‘mental’ to rules and representations is simply a matter of theoretical goals and interests. There is no evidence that the practice of theoretical linguistics would be changed if a ‘realistic’ interpretation of the grammar was assumed, rather than a ‘psychologically realistic’ one (cf. Milner 1979). (p. 40)

Other such statements are scattered throughout the literature, suggesting that there is a contingent of linguists who have reflected on the aims and claims of linguistic theory and have come to conclusions that diverge sharply from the main tenets of Chomsky’s cognitive conception. This, too, bolsters my contention that the cognitive conception is not the only game in town.

Chomsky’s view entails that psycholinguists should seek a relatively transparent relation between the syntacticians’ grammar and the “knowledge-base” or “data-structure” that is used by the human sentence processor. While Devitt and many others are deeply skeptical of this claim, there are, I believe, *methodological* grounds for pursuing it ahead of any competing hypothesis. As Berwick and Weinberg (1984: pp. 38-40) point out, a parsing theory that makes direct use<sup>48</sup> of the syntactic principles shored up by the formal syntactician is the simplest and hence the strongest theory that one could adopt. Hence, while we should not expect, let alone presume, that we will find it to be true, we should do our best to subject it to empirical scrutiny *first*, so as to disconfirm it before moving on to weaker hypotheses. As I will report in subsequent chapters, this effort has been underway for over three decades and the results that are now available have not, I think, *contravened* this strong hypothesis, though they have necessitated a much more

---

<sup>48</sup> The term ‘direct use’ is deliberately vague, eliding the distinction between the declarative representation and the procedural embodiment of a grammar. This distinction will take center stage in ch. V.

careful statement of it. All in all, though, the hypothesis that the syntactician's grammar is psychologically real has fared rather well.<sup>49</sup>

The methodological hope of those who are currently working on sentence processing is to find a natural grammar-parser combination. Theorists are certainly aspiring toward that goal and have achieved some measure of success (chs. III-V), but the progress has been rather haphazard. In part, this is because syntacticians are frequently not as concerned with fine-grained psychological data as one would expect them to be if their goal was to construct psychologically real grammars. A survey of the mainstream literature in generative syntax reveals, I think, that psychological reality is a distant, dimly understood, and rarely invoked desideratum. Indeed, it is a notorious gripe in psycholinguistics that syntacticians do not construct grammars with the intent of making life simpler for parsing theorists.<sup>50</sup> For one thing, formal grammars come in and out of fashion every decade or so. Each minor revolution forces the psycholinguist to revise her assessment of the success and relevance of prior experimental results, as well as to devise novel experimental techniques that address issues raised by the dominant formalism of the day.<sup>51</sup> Moreover, syntacticians rely heavily on formal simplicity as a criterion for selecting a grammar. But, as Stabler (1984) argues, it is rather unlikely that this criterion will converge with the psycholinguist's aim of psychological plausibility.

---

<sup>49</sup> Talk of "*the* syntactician's grammar" is actually quite misleading. There are many substantively distinct grammars on offer, e.g., Principles and Parameters grammars, Lexical Functional Grammar, and Generalized Phrase Structure Grammar, to name just a few. And there is even more variety if we consider the theoretically irrelevant but computationally significant notational variants of each of these grammars. See ch. V.

<sup>50</sup> See Frazier (1988). One wonders if increased attention to the grammar/parser distinction would remedy the situation.

<sup>51</sup> The rolling revolution in syntactic theory also makes trouble for *syntacticians*. Ludlow (2011: p. 29) discusses how fans of the Minimalist Program were forced to sacrifice the "sweet results" that issued from the PRO theorem, when that theorem was abandoned by Chomsky (1995)—a case of what Ludlow quite aptly labels "Kuhn loss."

Nevertheless, there is reason to believe that developments in psycholinguistics will reveal the role that grammatical rules and principles play in the psychology of a competent language user. In particular, as I shall show in the chapters to come, detailed models of parsing and comprehension routinely appeal to one or another grammar in the course of explaining a broad range of behavioral and neurocognitive data. The remainder of the present work can be seen as an exploration of the relevant findings in this area and a sustained argument for the psychological reality of grammar. We begin in the following chapter by reviewing the major lines of evidence for the psychological reality of phrase structure representations.

### III

## MENTAL PHRASE MARKERS IN SENTENCE PROCESSING

---

### §1. Introduction

The primary goal of the present work is to assess the bearing of syntactic theory on the psychology of language. In the previous chapters, I argued that there are substantive questions about whether the theoretical constructs of formal linguistics play any role in the psychological processes underlying language use. In the remainder of this work, I will argue that the rules or principles of a grammar are indeed psychologically real, though perhaps not in the way that some philosophers have assumed. Defending this claim will be the burden of ch. V. But the rules and principles of grammar are not the only theoretical construct of interest. In this chapter and the next, I aim to show that the human sentence processing mechanism (henceforth, the HSPM) constructs what I will call *mental phrase markers*. As I use the term, a mental phrase marker is a representation of the syntactic structure of a linguistic stimulus. The claim that the HSPM constructs mental phrase markers in the course of comprehension underlies a great deal of work in psycholinguistics. Here are three typical statements of it by leading figures in the field:

[L]et us suppose (as we surely should, until or unless the facts dictate against it) that the human sentence processing routines compute for a sentence the very structure that is assigned to it by the mental “competence” grammar. (Fodor, J. D., 1989: p. 157).

[W]e assume in the first place [the] [w]eak competence hypothesis: Human syntactic analysis typically involves the explicit recognition of all grammatical relations. (Stabler, 1994: p. 303)

Most models of human language comprehension assume that the processor incorporates words into a grammatical analysis as soon as they are encountered. ... We assume that sentence processing involves the computation of dependencies between the words and phrases that are encountered. For example, in the sentence *The troops found the enemy spy*, the relations include information that *the troops* is the subject of *found*. Often, words are incorporated directly into the representation without breaking existing dependencies. For example, when the main verb *found* is

encountered, the processor forms the dependency between *the troops* and *found*, and does not need to break any other dependencies (e.g., that between *the* and *troops*). (Sturt, Pickering, and Crocker, 2001: p. 283)

While certainly popular, the assumption that the HSPM constructs mental phrase markers is not without its detractors. Consider the following passages, drawn from work in philosophy, psycholinguistics, and artificial intelligence.

[L]anguage use [is] a fairly brute-causal associationist process ... rather than a process involving metalinguistic representations of the syntactic and semantic properties of linguistic expressions. (Devitt, 2006: p. 220)

[N]o independent level of syntactic representation is constructed, operated on, or output by the language analysis process. (Schank and Birnbaum, 1984: p. 220)

[I]t isn't even clear that we do parse, in the sense of constructing an explicit and complete representation of the structure of a sentence in the course of comprehending it. The only language tasks that we really know humans are able to perform are those we can directly observe, including comprehension, production, repetition, and so forth. That mapping from an acoustic signal to a message occurs via deep structure, or some other hierarchical syntactic representation, is merely a convenient assumption—a pervasive and often useful one, but an assumption none the less. In advancing our understanding of the human language processing system, we may do well to question this assumption. (Rohde, 2002: p. 6-7) ... There seems to be little direct evidence that we construct a representation of a syntactic parse tree while comprehending. (p. 18)

In what follows, I present a number of arguments for the psychological reality of mental phrase markers (henceforth, MPMs). From the outset, I wish to stress the paramount significance of behavioral and neurocognitive data, garnered from psycholinguistic experiments. As Fodor, Bever, and Garrett (1974) put it, “detailed empirical constraints can and should be placed upon [parsing] models. Any theory of sentence recognition must be committed to an array of testable predictions” (p. 319). In the standard case, such predictions will be concerned with the speed and accuracy of verbal reports, button presses, and eye movements, as well as the character of the neuronal activity that underpins—and ultimately constitutes—language processing.

Many ingenious experimental designs have been used to study one or another aspect of written or spoken language is processed.<sup>1</sup> A subject in a typical psycholinguistic experiment is given a task designed to test whether some controlled variable affects her ability to recognize and categorize an expression, to disambiguate it, to recall it, to comprehend its meaning, to make inferences from its meaning to other information, and so on (Fernandez and Cairns, 2011). Such tasks can be performed in non-optimal conditions, with degraded or highly ambiguous stimuli, and in conjunction with ancillary task demands. The completion of a task is signaled either by a behavioral response or by a neurophysiological signal. The results of such experiments constrain models of comprehension; a model that stays faithful to the available behavioral and neurocognitive data has, ipso facto, a stronger claim to psychological plausibility.<sup>2</sup>

Since the inception of the cognitivist study of language, literally *thousands* of experiments have been performed, many of them geared toward establishing one or another claim regarding the psychological reality of grammars, and of the mental phrase markers that are couched in the proprietary vocabularies of those grammars. Plainly, a comprehensive review of the relevant findings will not be possible in the space of a chapter. In what follows, then, I

---

<sup>1</sup> Studies of sign language processing are quite rare, but they do exist. See, e.g., Bavelier, D., Corina, D. P., and Neville, H. J. (1998). "Brain and Language: A Perspective from Sign Language Properties of ASL," *Neuron*, Vol. 21, pp. 275–278, and references therein.

<sup>2</sup> It perhaps goes without saying that my insistence on the primacy of behavioral and neural data is not an expression of behaviorist or crypto-eliminativist principles. Indeed, I plan to argue, on the basis of such data, that certain processes and representations are *psychologically real*. If insisting on the primacy of third-person data were a sign of an underlying philosophical agenda, the agenda in question would best be described as *anti-Cartesianism*—a methodology that de-emphasizes wanton intuition mongering and the deliverances of consciousness and introspection. The fruitlessness of Cartesian theorizing is apparent not only from the historical failure of 19<sup>th</sup>-century introspectionist psychology (Lyons, 1986; Kukla and Walmsley, 2006), but also from the endless, seemingly intractable debates to which such theorizing has recently given rise in the field of consciousness studies. For a discussion of the pitfalls of relying on introspective data in psycholinguistics, see Karttunen and Zwicky (1985: p. 22) and Tanenhaus, Carlson, and Seidenberg (1985).



examine a small but suggestive sample of the results, focusing in large part on studies of comprehension.

### *§1.1 Outline of the chapter*

I begin in §2 by taking a brief look at what is currently known about the neural processes underlying language comprehension. Though neurolinguistics is currently thriving, the field is in its very early stages. In the 1990s, results emerging from a number of labs suggested that a phase of distinctly syntactic processing can be identified via neurophysiological measures. Subsequent work showed, however, that the matter is more complicated than these initial studies suggested. The present state of the field is characterized by heated debates about the proper interpretation of various neural markers, particularly the P600 and the N400.<sup>3</sup> Although it is too early to tell how such debates will pan out, we shall see that the psychological reality of mental phrase markers is supported by all of the available findings.

Turning to behavioral studies in §3, I discuss the results of a number of experiments designed to examine a phenomenon known as structural priming. These studies provide strong support for claim that the HSPM constructs representations of distinctly syntactic properties of the input. I also review some evidence that casts doubt on the claim that the HSPM constructs the D-structure representations posited by the Government and Binding theory. The evidence suggests, rather, that the HSPM computes S-structures directly.<sup>4</sup>

---

<sup>3</sup> I discuss the significance of these labels in §2.

<sup>4</sup> Many theorists were persuaded by the arguments in Fodor, Bever, and Garrett (1974) that the transformations posited by a variety of syntactic theories are not psychologically real operations. Fodor, Bever, and Garrett argued against what they called the derivational theory of complexity (DTC), on the basis of a number of experimental findings. Following Berwick and Weinberg (1984: ch. 2), Phillips (1994, 1996) and Phillips and Lewis (forthcoming), I doubt that the findings do in fact disconfirm the DTC. (Phillips and Lewis write: “We do not mean to claim that the DTC was substantiated, but the reports of its defeat strike us as somewhat stylized history.”) Nevertheless, I am aware of no empirical grounds for positing D-structure representations in the course of sentence

In §4, I turn to the psycholinguistic experiments that reveal what are known as “garden-path” phenomena in sentence processing. These are cases in which the HSPM encounters a locally ambiguous input and resolves the ambiguity in a way that turns out to be incorrect relative to the completion of the sentence or phrase. I discuss two principles of ambiguity resolution, Minimal Attachment and Late Closure, which form the foundation of most psychologically plausible parsing models. I then examine the experimental results that have led a number of theorists to posit a stage in processing where only syntactic information plays a role in determining the HSPM’s ambiguity resolution preferences. I argue that the evidence currently available supports positing such a stage, hence rejecting the models on which the HSPM employs semantic information from the very outset of the comprehension process.

§5 is devoted to a survey of the fascinating attempts to demonstrate experimentally the presence of empty categories—specifically, *wh*-traces and NP-traces—in the mental phrase markers that the HSPM constructs. The evidence currently available strongly suggests that *wh*-traces are psychologically real and that the HSPM employs rather sophisticated strategies in searching for them, making use of the cues provided by their antecedents as well as considerable knowledge of grammatical constraints. I discuss the implications of this fact for the processing model proposed by Fodor, Bever, and Garrett (1974), and for the psychological reality issue more generally. Finally, I review recent experimental results that provide support for the claim that NP-traces are likewise present in mental phrase markers.

The evidence provided here will allow us to critically assess, in ch. IV, two models on which MPMs play no role in language comprehension: the “integrated processing” models developed by Roger Schank and his colleagues at Yale (Schank and Birnbaum, 1984) and the

---

processing. See ch. V (§4) for further discussion.

“brute-causal” model tentatively advanced by Devitt (2006). Moreover, several of the findings discussed in this chapter—particularly the explanatory success of principles like Minimal Attachment and Late Closure—will play a crucial role when we turn to the psychological reality of syntactic rules and principles in ch. V.

## §2. The Argument from Neurocognitive Data

Let us begin by surveying some of the results emerging from the field of neurolinguistics that bear on the psychological reality of mental phrase markers. As noted above, neurolinguistics is currently in its infancy. So we must be rather tentative when drawing conclusions from the available findings. Nevertheless, it would be unwise to ignore a range of data that strongly suggests that the HSPM constructs mental phrase markers. In arguing for an even stronger claim—viz., that “real-time processes assemble syntactic representations that are *the same as those motivated by grammatical analysis*” (emphasis mine)—Phillips and Lewis (forthcoming) say the following:

[S]tudies that use highly time-sensitive measures such as event-related brain potentials (ERPs) have made it possible to track how quickly comprehenders are able to detect different types of anomaly in the linguistic input. This work has shown that speakers detect just about any linguistic anomaly within a few hundred milliseconds of the anomaly appearing in the input. Different types of grammatical anomalies elicit one or more from among a family of different ERP components, including an (early) left anterior negativity ((e)LAN’; Neville et al., 1991; Friederici, Pfeifer, & Hahne, 1993) or a P600 (Osterhout & Holcomb, 1992; Hagoort, Brown, & Groothusen, 1993). Many questions remain about what the different components reflect and what determines which components are evoked in any individual situation (Hagoort, 2003; Friederici & Weissenborn, 2007; Lau, Phillips, & Poeppel, 2008; Gouvea et al., 2009), but for current purposes the most relevant outcome from this research is that more or less any grammatical anomaly elicits an ERP response within a few hundred milliseconds. If the on-line analyzer is able to immediately detect any grammatical anomaly that it encounters, then it is reasonable to assume that it is constructing representations that include sufficient grammatical detail to detect those anomalies.

In fact, it is no exaggeration to say that virtually *all* of the work presently being conducted in

neurolinguistics presupposes not only that the HSPM constructs mental phrase markers, but that it does so in consultation with a mentally represented grammar. In a text that provides a comprehensive review of the existing work on the neural underpinnings of language processing, Bornkessel-Schlesewsky and Schlewsky (2008) begin by outlining the conditions that must be satisfied by any psychologically plausible model of comprehension.

Properties of the processing system required by incremental comprehension: The processing system must impose a structure on the input as quickly as possible. ... [It must then use] the structure/meaning assigned to generate predictions about the incoming input. ... Structure building must clearly be constrained by the grammar assumed for the language being processed. Thus, the grammar constrains the types of structures that can be constructed from the input. (Bornkessel-Schlesewsky and Schlewsky, 2008: p. 90)

In the next several sections of this chapter, we will review the evidence for the claim that comprehension is “incremental,” in the sense these authors intend. For the time being, let us turn our attention to the nature of the experiments that Phillips and Lewis mention in the passage quoted above.

A number of techniques are currently used in obtaining recordings of neural activity. These include electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), positron emission tomography (PET), and transcranial magnetic stimulation (TMS). Bornkessel-Schlesewsky and Schlewsky (2009) provide an in-depth analysis of the strengths and weaknesses of these techniques. The most salient contrasts between them concern the accuracy that they achieve in measuring the speed of processing and in discerning the precise location at which processing occurs.

EEG and MEG recordings are time-locked to the presentation of a stimulus and are capable of yielding impressively precise measurements of the time at which a neurophysiological response occurs, with an accuracy of up to 4 milliseconds. Unfortunately, the cost is a marked

*imprecision* with respect to the location of the source of any such signal. By contrast, recordings from fMRI devices yield quite precise data concerning the location of a neurophysiological response, but leave open a window of up to several seconds during which the response might have occurred. Our primary concern here is to establish the existence of a processing phase at which the HSPM is sensitive only to syntactic properties. As such, we do well to focus on the EEG data, which has provided the lion's share of the evidence bearing on this issue.

EEG devices measure what are known as *event-related potentials* (ERPs), also sometimes called *evoked potentials*. These are small voltage differences between electrical activities in the brain, recorded by electrodes placed on the scalp. ERP data is multi-dimensional, providing information about the degree, polarity, and location of neuronal activations. In addition, as noted above, ERPs are time-locked to a stimulus event. Hence, they admit of functional interpretation, by reference to the properties of the stimuli that evoke them. (Illustrations below.)

In studies of linguistic comprehension, a typical piece of EEG data will specify the latency, degree, polarity, location, and distal cause of an ERP, in a graph like the one displayed in Fig. 1 below. There, we see a comparison between the electrical activity evoked by two distinct stimuli—in this case two German sentences, one grammatical, the other not. At the top left, the gross location of the activity is specified ('F7'). The values on the y-axis represent the degree of the signal and its polarity (positive or negative). The values on the x-axis represent the signal's latency—i.e., time at which it occurs, relative to the onset of the stimulus.

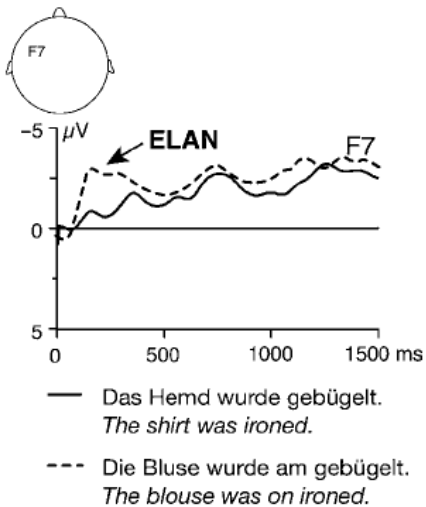


**Figure 1:** A typical display of ERP data, showing the latency, degree, polarity, location, and distal cause of the neuronal signal. In this case, the signal is an ELAN—an early negativity in the left anterior region of the brain. As the graph shows, the ELAN occurs roughly 125 milliseconds after the onset of the critical stimulus. Like the studies discussed in the main text, the experiment from which this data was derived used two German sentences as stimuli. The first, indicated by the unbroken line, is a grammatical sentence. The second, indicated by the broken line is *ungrammatical*—i.e., it exhibits a basic phrase structure violation. The critical stimulus is the word ‘ironed’. The ERP associated with the grammatical sentence are significantly different from the one associated with the ungrammatical sentence. The graph shows a distinct negativity (conventionally plotted upward on y-axis). *Source: Bornkessel-Schlesewsky and Schlewsky (2009: p. 110)*

The most widely used experimental paradigm in ERP studies is known as *the violation paradigm*. Participants are shown a variety of sentences, some of which contain violations with respect to one or another linguistic property. For instance, in an early ERP study, Neville et al. (1991) used the following materials:

- |                                                              |                                           |
|--------------------------------------------------------------|-------------------------------------------|
| (1) *The man admired Don’s <u>of</u> sketch the landscape.   | syntactic violation                       |
| (2) *The man admired Don’s <u>headache</u> of the landscape. | semantic/pragmatic violation <sup>5</sup> |
| (3) The man admired Don’s sketch of the landscape.           | control (no violation)                    |

As subjects read such sentences, an EEG device monitors their brain activity, particularly at the

<sup>5</sup> In view of the notoriously shaky status of the semantics/pragmatics distinction, I simply slur over it in what follows, labeling various properties of sentences ‘semantic’ regardless of whether they would be classified as semantic or pragmatic by a theorist who (mistakenly, in my view) insists on drawing the distinction. Though I cannot address the issue here in any detail, I also don’t see that anything essential to the argument hangs on it.

crucial regions, underlined in (1) and (2). The logic of the paradigm is straightforward. The ERPs associated with the anomalous stimuli differ significantly from those associated with stimuli that are free of linguistic violations. This in turn provides information about where and when in the brain specific kinds of violation are represented. For instance, Neville et al. found that sentence (1) evoked a negative-polarity response in the left anterior region of the brain approximately 125 milliseconds after the onset of the word ‘of’, followed by a positive-polarity response several hundred milliseconds later. This fast negative-polarity response has come to be known as the ELAN—early left anterior negativity—and is depicted in Fig. 1. By contrast, the non-syntactic violation in sentence (2) evoked a negative-polarity response approximately 400 milliseconds after the onset of the word ‘headache’. This has been dubbed the N400.

In another early study, Friederici, Pfeifer, and Hahne (1993) found the same pattern—a replication that is especially striking given the fact that, unlike Neville et al., Friederici et al. used German rather than English sentences and presented them auditorily rather than visually. The materials from the Friederici et al. study appear in (4)-(6).

- |                                                                    |                                 |
|--------------------------------------------------------------------|---------------------------------|
| (4) *Der Freund wurde im besucht.<br>the friend was in-the visited | syntactic violation             |
| (5) *Die Wolke wurde begraben.<br>the cloud was buried             | semantic/pragmatic violation    |
| (6) Der Finder wurde belohnt.<br>the finder was rewarded           | control sentence (no violation) |

Indeed, virtually the same pattern has been observed in dozens of subsequent studies. The natural interpretation is that incoming words are incrementally incorporated into a mental phrase marker, with syntactic information being accessed quite early—125 milliseconds after stimulus onset—while other properties of the stimulus are recovered significantly later.

Bornkessel-Schlesewsky and Schlewsky (2009) discuss a series of further studies

designed to test whether the integration of syntactic information—i.e., the construction of a mental phrase marker—is, strictly speaking, a *prerequisite* for the integration of semantic information. This can be determined by constructing sentences in which the semantic violation is *combined* with the syntactic violation. Consider, for instance, sentence (7).

- (7) \*Das Gewitter wurde im gebügelt.                      combined syntactic *and* semantic violation  
the thunderstorm was in-the ironed

What Hahne and Friederici (2002) found was that sentence (7) evokes an ELAN, which is characteristic of syntactic violations, but not the N400, which seems to be correlated with semantic violations. It appears, then, that the (presumably syntactic) ELAN is capable of *blocking* the (presumably semantic) N400. Crucially, it has also been discovered that the reverse is *not* true. That is, a “semantic” N400 evoked *prior* to a “syntactic” ELAN cannot “block” the ELAN.<sup>6</sup> Researchers have thus concluded that “existing ERP findings provide strong converging support for the assumption that constituent structure information *hierarchically dominates* other information types such as semantics/plausibility” (Bornkessel-Schlesewsky and Schlewsky, 2009: p. 113; emphasis in the original).

The kind of work reviewed above also bears on a key issue that will resurface in later discussion (section 4). There, we will turn to the question of how the HSPM goes about assigning structure in *locally ambiguous* circumstances—as in the famous “garden-path” cases, exemplified in (8)-(12).

- (8) Jake tells students he intrigues to stay.  
(9) Benny knows the boy hurried out the door slipped.

---

<sup>6</sup> The sudden rash of scare-quotes in this sentence indicates the extreme caution with which we must proceed here. Further studies may well show that this categorization of the phenomena is in need of revision. For instance, Stroud and Phillips (2010) report an array of findings that cast doubt on the initial picture of the N400. None of the findings they report, however, seem to affect the argument here. For, they do not challenge the claim that the ELAN is an unambiguous reflection of distinctly syntactic violations. Furthermore, the refinements they suggest to our interpretation of the ERP signal known as the P600 seem to do no violence to the analysis presented above.



- (10) The soldier persuaded the radical student that he was fighting in the war for to enlist.
- (11) Aron gave the man who was eating the fudge.
- (12) The old train the children.

It has been found such sentences cause various kinds of processing difficulty for ordinary readers. The work reviewed below will show that this happens on account of isolated failures of the HSPM's heuristics for resolving local ambiguities. Evidently, when the HSPM's preferred structural assignment in the middle of the sentence turns out to be incorrect by the time the sentence is complete, additional computational load is incurred. A number of theorists have proposed processing models on which the HSPM recovers syntactic information at the outset, and then performs an operation of *reanalysis* in light of more sophisticated semantic information. There is ongoing debate about whether such a model can be sustained. Bornkessel-Schlesewsky and Schlewsky (2009) put the matter as follows:

While the need for constituent (phrase) structure is undisputed in both theoretical linguistics and psycholinguistics, its precise role in real time comprehension is somewhat more controversial. A number of positions on this issue have been advocated, the two extremes of which can be described as follows: (a) constituent structuring is the prerequisite for all further processing steps, both syntactic and interpretive in nature, and (b) constituent structure is not accorded any special role but rather interacts with all other available information types (e.g. animacy, frequency, thematic roles) to determine sentence structure and interpretation. (p. 107)

The experiments discussed above plainly bear on this issue. Indeed, early ERP studies were designed for the explicit purpose of testing such hypotheses, thus adding to the pool of behavioral data that we discuss in §4 below. The findings support those models on which purely syntactic information plays a privileged role in determining the HSPM's initial parsing preferences. For instance, Friederici (1995) conducted a series of studies in which the ERP data was found to conform rather tidily to the predictions of so-called *syntax-first* models. She concludes her report of the findings in the following passage:

The combined findings discussed here suggest that on-line structuring processes are subserved by brain systems located in the anterior part of the left hemisphere [DP: the site of the ELAN], whereas processes of structural reanalysis seem to involve different brain systems. ... [W]e may take the available data to suggest a parser with two subcomponents, a first subcomponent responsible for the early structuring of the input seemingly working in a highly time-dependent procedural manner and a second subcomponent responsible for syntactic integration and reanalysis *consulting grammatical knowledge* which may be represented in a less time dependent form. With such an architecture, the syntactic processing system would (a) be fast in assigning structure to the incoming information and would (b) be most flexible in selecting the valid structure for adequate thematic role assignment and the ultimate interpretation. (Friederici, 1995: p. 278; emphasis added)

In a more recent publication, Friederici and Brauer (2009) tout the continuing success of this interpretation in light of new evidence. Having reviewed the currently available homological comparisons with animal neuroanatomy as well as a number of lesion studies and child language processing studies, they conclude that

[s]o far, empirical evidence supports the assumption of two separate functional brain networks underlying syntactic processes. Local phrase structure building involves the frontal operculum and the anterior part of the STG [i.e., the superior temporal gyrus], while the processing of complex hierarchical structures requires the participation of Broca's area and very likely the posterior part of the STG. These functional networks correspond to structural connections in the brain between associated areas within each of these circuits. Relevant data from the non-human brain and from children support the view that the network responsible for complex syntax might have evolved late during phylogeny and only develops late during ontogeny. (p. 502)

Importantly, some of the studies reported in Friederici and Brauer (2009) seem to show unequivocally that the relevant brain circuits are processing specifically *syntactic* information. This can be seen from the fact that the regions are activated in the course of processing sentences of *artificial* languages, such as the ones depicted in fig. 2 below. Plainly, there is no possibility of semantic processing in such cases; the "words" are all nonsense syllables.

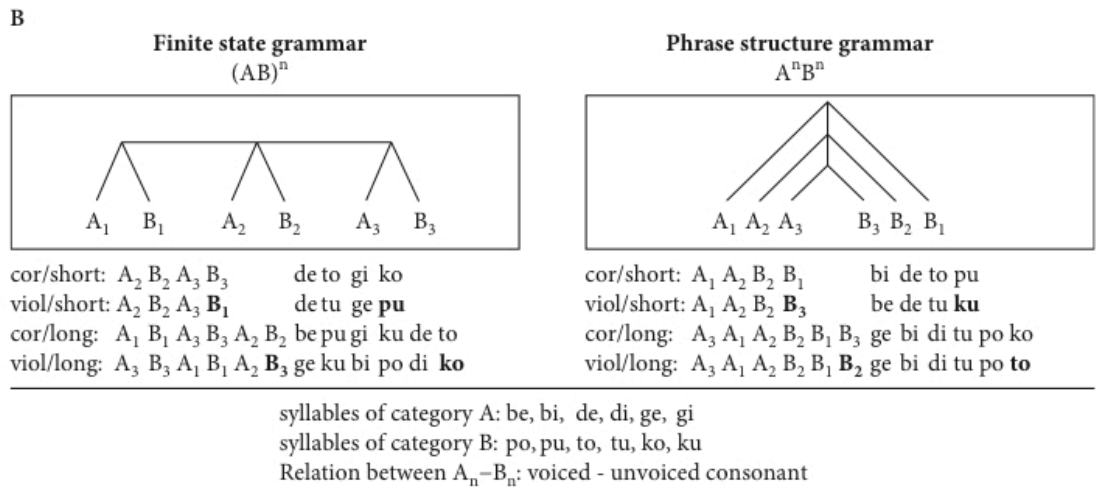


**Figure 2:** The artificial grammars used for some of the ERP research reported in Friederici and Brauer (2009). Note that the examples marked ‘cor’ (which is short for ‘correct’) illustrate what each grammar can generate, while those marked ‘viol’ (short for ‘violated’) illustrate what they cannot generate.

### Conclusion

Although the discussion above paints a tidy picture of the available data, it would be wrong to assume that there are no competing interpretations. For instance Bornkessel-Schlesewsky and Schlewsky (2009: ch. 12) discuss an array of data that suggests the need for a revision in our view of various ERP results. Ongoing debate among neurolinguists about the status of the ERP known as the P600 (a late positivity signal) serves as a sobering reminder of the need to be tentative in one’s conclusions. Nevertheless, we have seen reasons for thinking that the HSPM constructs mental phrase markers in the course of comprehension. In the next section, we will review a range of behavioral studies in support of the same conclusion. §§4-5 expand on these findings to provide a clearer picture of the internal operations of the HSPM. As we proceed through the discussion, I will occasionally note the empirical support for the claim, expressed clearly in the above quotation from Friederici (1995), that the HSPM performs its syntactic analysis in consultation with a mentally represented grammar.

### §3. The Argument from Structural Priming

Early behavioral studies of the psychological reality of syntactic constituency were rather primitive. Some amounted to little more than simply asking subjects which strings of words in a sentence seem to them to constitute natural groupings.<sup>7</sup> In their classic text, Fodor, Bever, and Garrett (1974) reported a wide range of studies that they interpreted as demonstrating the psychological reality of mental phrase markers—i.e., structural descriptions of stimulus sentences. Among these were the famous click experiments, in which subjects monitoring a speech stream misheard short clicks as if they occurred at constituent boundaries. However, as Jurafsky and Martin (2007: pp. 424-5) note, much of this evidence was rather weak.<sup>8</sup>

One of the issues was that many of these studies failed to control for semantic biases that correlate with syntactic structure. After all, many effects that can be explained by the hypothesis that subjects group words into a *syntactic* perceptual unit can equally well be explained by the hypothesis that the grouping is a *semantic* one. Convincing arguments for the psychological reality of syntactic constituency must, therefore, be based on data that is independent of semantic confounds. Jurafsky and Martin point out, furthermore, that “since there are many non-constituent-based theories of grammar based on lexical dependencies, it is important to find evidence that cannot be interpreted as a *lexical* fact; that is, evidence for constituency that is not based on particular words.” Recent evidence from priming studies fits the bill.

The logic of priming experiments is this: A mental representation activated at time  $t$ —say, in response to a stimulus—continues to be active for some time after  $t$ , and raises the probability of that representation’s influencing cognitive processing as long as the activation

---

<sup>7</sup> See, for instance, Levelt, W. J. M. (1970). “A Scaling Approach to the Study of Syntactic Relations,” in d’Arcais, G. B. F. and Levelt, W. J. M. (eds.), *Advances in Psycholinguistics*, pp. 109–121. North-Holland, Amsterdam.

<sup>8</sup> Clark and Clark (1977) discuss some of the methodological problems of the early experiments.

persists.<sup>9</sup> Pickering and Ferreira (2008) discuss the importance of this phenomenon to recent research in sentence processing.

In the past couple of decades, research in the language sciences has revealed a new and striking form of repetition that we here call *structural priming*. When people talk or write, they tend to repeat the underlying basic structures that they recently produced or experienced others produce. This phenomenon has been the subject of heavy empirical scrutiny. Some of this scrutiny has been because, as in other domains in cognitive psychology (e.g., priming in the word-recognition literature; e.g., McNamara, 2005), the tendency to be affected by the repetition of aspects of knowledge can be used to diagnose the nature of that knowledge. [T]he tendency to repeat aspects of sentence structure helps researchers identify some of the representations that people construct when producing or comprehending language. As we shall see, much structural priming is unusually abstract, evidently reflecting the repetition of representations that are independent of meaning and sound. This is therefore informative about how people represent and use abstract structure that is not directly grounded in perceptual or conceptual knowledge. One possibility is that the representations that it identifies can be equated with the representations assumed in formal linguistics.

An early and influential study in this vein is reported in Bock and Loebell (1990). The researchers were careful in eliminating the semantic and lexical confounds mentioned above, by constructing their stimulus materials in such a way as to vary syntactic structure independently of lexical and semantic structure, and vice versa. This was made possible by the fact that some verbs in English are ditransitive—capable of being used in semantically identical but syntactically distinct expressions. Examples of ditransitive verbs include ‘give’, ‘sell’, and ‘send’. The examples in (13)–(15) illustrate their alternations between a double-object dative construction, i.e., a V-NP-NP structure, as in the first in each of the pairs, and a prepositional dative construction, i.e., a V-NP-PP structure, as in the second in each of the pairs.

- (13) a. Quentin [<sub>V</sub> gave [<sub>NP</sub> Oliver [<sub>NP</sub> a car]]]. / b. Quentin [<sub>V</sub> gave [<sub>NP</sub> a car [<sub>PP</sub> to Oliver]]].  
(14) a. Quentin [<sub>V</sub> sold [<sub>NP</sub> Oliver [<sub>NP</sub> a car]]]. / b. Quentin [<sub>V</sub> sold [<sub>NP</sub> a car [<sub>PP</sub> to Oliver]]].  
(15) a. Quentin [<sub>V</sub> sent [<sub>NP</sub> Oliver [<sub>NP</sub> a car]]]. / b. Quentin [<sub>V</sub> sent [<sub>NP</sub> a car [<sub>PP</sub> to Oliver]]].

---

<sup>9</sup> This formulation is somewhat loose. For a fine-grained distinction between priming and persistence, see Pickering and Ferreira (2008: p. 428). The distinction does not, however, affect the present argument.

Bock and Loebell's made use a picture description paradigm. Participants were first asked to read some sentences out loud. Unbeknownst to them, these sentences served as primes and were selected by the experimenters for having a preposition after the verb (i.e., a V-NP-PP structure), but differing from (13b)-(15b) in their semantics and lexical constituency. For instance, although a sentence like (16) has the same *syntactic* structure as those in (13b)-(15b), it has none of the same words as the sentences in (13b)-(15b)<sup>10</sup> and, crucially, has a different semantic interpretation—e.g., the preposition carries a locative meaning, as against the dative meaning of the prepositions in (13b)-(15b).

(16) IBM [<sub>V</sub> moved [<sub>NP</sub> a bigger computer] [<sub>PP</sub> to the Sears store].

Having read out loud sentences like (16), participants were shown pictures and asked to describe them. The pictures depicted events that involve an agent giving something to someone. Such events can be described equally well by sentences that employ the double-object dative construction (e.g., *The boy gave the girl flowers*) and ones that employ the prepositional dative construction (e.g., *The boy gave flowers to the girl*).

What Bock and Loebell found was that participants exhibited a strong priming effect. Those who initially read out loud sentences that employ the double-object dative construction were more likely to employ the same construction in describing the events depicted in the pictures. Similarly, those who had recently read out loud sentences that employ the prepositional dative construction were more likely to employ *that* construction in describing the events depicted in the very same pictures. This strongly suggests that the subjects in the experiment constructed a mental representation of the syntactic properties of the sentences that they were initially asked to read, and then used that representation in repeating those sentences out loud.

---

<sup>10</sup> Below I discuss the one word they do have in common—viz., the preposition 'to'. To foreshadow, this commonality has been shown to be irrelevant to the argument at hand.

The representation then remained active in their language processing system, hence more likely to be reused later in production.

The experiment just described was one of the first in what has become a very long line. Indeed, the area of structural priming research is currently thriving, in large part because the results are so robust and the data so telling. This leads Pickering and Ferreira (2008) to speculate about the “intriguing possibility that *all* levels of processing that occur during production show priming and therefore that the absence of priming suggests the absence of a corresponding level of representation” (p. 429, emphasis added).

Bock and Loebell’s initial conclusions have been refined and extended in a number of ways. Pickering and Ferreira (2008) offer a comprehensive review of the relevant literature. In it, they discuss studies which show that structural priming is not restricted to the constructions mentioned above—e.g., it occurs with active-passive pairs and other constructions. Moreover it is *not* due to the presence of common closed-class words in the stimulus materials—e.g., the preposition ‘to’ in sentences (13)-(16).<sup>11</sup> Nor is structural priming restricted to a single language; the phenomenon has been observed in German, and bilingual English-German speakers exhibit *cross-linguistic* priming effects—i.e., primed production across their two languages, with respect to suitable constructions. Children and aphasics also exhibit structural priming effects, ruling out the possibility that the phenomenon is restricted to some special set of language users. Further studies rule out the possibility that subjects produce forms similar to the prime because they want to stay in the same rhetorical register, e.g., formal speech. Other communication-related effects have also been controlled for. Similarly, structural priming is

---

<sup>11</sup> Bock (1989) finds priming across sentences with different prepositions—e.g., ‘to’ and ‘for’.

independent of both prosody and argument structure (i.e.,  $\theta$ -assignment), and can be elicited cross-modally from spoken to written language and vice versa.<sup>12</sup> Finally, the same findings have been replicated using experimental paradigms other than the picture-description paradigm. These include sentence recall, written sentence completion, and spoken sentence completion.<sup>13</sup> Having surveyed and ruled out a range of possible confounds, Pickering and Ferreira write:

In conclusion, taken together, these results provide compelling evidence for autonomous syntax: The production of a sentence critically depends upon an abstract syntactic form that is defined in terms of part of speech forms (e.g., nouns, verbs, prepositions) and phrasal constituents organized from those (noun phrases, verb phrases, prepositional phrases), and this abstract syntactic form has a large influence upon structural priming.

#### *Evidence against the psychological reality of D-structure representations*

There remains a question, of course, about precisely *which* structural descriptions the HSPM constructs. Although experiments like those reviewed above show that the HSPM constructs representations of syntactic structure—in addition to representations of semantic, prosodic, and argument structure—this leaves open the possibility that the HSPM constructs only certain kinds of syntactic descriptions and not others. For instance, we might ask whether the HSPM computes not only the surface structures of sentences, but also the D(eep)-structure representations posited by the Government and Binding theory (Chomsky, 1986; Haegeman, 1994) and earlier versions of transformational grammar (e.g., Chomsky, 1965).

The experiments in Bock, et al. (1992) provide evidence against the hypothesis that, in the course of producing passive sentences, the HSPM constructs D-structure representations. In

---

<sup>12</sup> This contradicts a contention of Devitt (2006) to the effect that there may well be no modality-neutral language faculty. Pickering and Ferreira discuss what they take to be “strong evidence that at least those aspects of structural knowledge that underlie structural priming are modality independent—they are used in the same way both when speaking and when writing” (p. 439).

<sup>13</sup> See Pickering and Ferreira (2008: p. 428) for a detailed list of references.



these experiments, the materials employed were active and passive sentence pairs, in which the animacy values of the nouns were varied, as in the following examples:

- (17) Five people carried the boat.  
*active, animate surface subject, inanimate deep object*
- (18) The boat was carried by five people.  
*passive, inanimate surface subject., inanimate deep object*
- (19) The boat carried five people.  
*active, inanimate surface subject, animate deep object*
- (20) Five people were carried by the boat.  
*passive, animate surface subject, animate deep object*

Bock, et al. (1992) asked participants to read sentences like (17)-(20) out loud. These served as the primes. The participants were then asked to describe pictures in which, e.g., a boy wakes up to the sound of an alarm clock ringing. This event can be described using either an active or a passive form, as in the examples below:

- (21) The alarm clock awakened the boy.  
*active, inanimate surface subject, animate deep object*
- (22) The boy was awakened by the alarm clock.  
*passive, animate surface subject, animate deep object*

In interpreting the results, it is important to understand that the Government and Binding framework treats passive sentences like (18), repeated below, as having a D-structure representation in which ‘the boat’ is located in the object position, as depicted in (23).

- (18) The boat was carried by five people.  
*passive, inanimate surface subject, inanimate deep object*
- (23) was [<sub>VP</sub> carried [<sub>NP</sub> the boat] by five people.  
*D-structure representation of sentence (18)*

This “deep object” is then moved by a transformational step into what appears on the surface to be a subject position, resulting in the surface form shown in (18).

With this in mind, we can appreciate the logic of the experiment. The materials are designed to tease out whether the participants are primed by the surface structures of (17)-(20) or by their deep structures. In accordance with the structural priming results described above, it was expected that participants who were primed with active sentences (17) and (19) would be more likely to describe the event using the active sentence in (21), and that those who were primed with passive sentences (18) and (20) would produce passive descriptions like (22). The data bear this out, thus providing additional confirmation of the structural priming hypothesis. However, what was *not* known in advance was whether participants who heard sentences with, say, animate deep objects—e.g., (19) and (20)—would produce a description with an animate deep object—i.e. the sentence in (21). If they do, then this would provide evidence for the hypothesis that participants are constructing D-structure representations. This, however, was *not* what Bock, et al. observed.

Instead, the data show that participants were more likely to produce (21) after being primed with (18) and (19) than if they were primed with (17) and (20). As Pickering and Ferreira point out, the crucial thing to look at is the difference between the priming effects of sentences (18) and (20). Both are passive, so both are less likely to prime for the active sentence (21). However, whereas (18) shares with (21) the property of having an inanimate surface subject, (20) does not. Thus, the fact that (18) primes the production of (21) more than (20) does “reveals that semantic representations—representations with particular animacy values—are mapped onto surface structure positions, not deep structure positions, contradicting an approach to production that is directly translated from transformational accounts of syntactic representation in formal linguistics” (Pickering and Ferreira, 2008: p. 433). For this and other reasons and others, today’s most promising comprehension models are typically *not* designed to

compute D-structure representations. In general, they directly compute S-structure representations, of one stripe or another.<sup>14</sup>

### *Conclusion*

The phenomenon of priming provides an insight into the nature and effects of the representations we construct in the course of many cognitive processes. Structural priming affords us a glimpse into the character of the representations constructed by the HSPM in the course of language comprehension and use. A number of structural priming experiments shore up strong evidence for the psychological reality of mental phrase markers. They also allow us to rule out substantive proposals that were entertained in the early stages of the cognitivist study of language—particularly, the idea that the HSPM constructs D-structure representations.

In the sections that follow, I survey several other arguments for the psychological reality of mental phrase markers. The studies reviewed below shed further light on the character of mental phrase markers, as well as the information that the HSPM employs in constructing them. By examining fine-grained behaviors, such as eye movements, the studies also provide information about the character of the HSPM's processing routines, especially with respect to ambiguity resolution.<sup>15</sup>

---

<sup>14</sup> Below, I discuss the much-studied question of whether the HSPM constructs representations that include empty categories, such as *wh*-traces, PRO, *pro*, and NP-traces. These are posited by the Government and Binding theory, but not by (some versions of) competing grammatical frameworks, such as Head-Driven Phrase-Structure Grammar (Pollard and Sag, 1994) and Lexical-Functional Grammar (Bresnan, 1978, 2001). The Government and Binding framework posits *wh*-traces and NP-traces at S-structure and at the distinct level of representation, LF. As has been widely noted, there is no need for a parser or a generator to construct D-structure representations, given that all of the information in those representations is arguably already encoded by antecedent-trace relations at S-structure. See, for instance, Fodor (1989: p. 178). The most recent incarnation of transformational grammar in the Principles and Parameters tradition—viz., the Minimalist Program—does away with the level of D-structure altogether. I discuss Minimalist parsing models in ch. V (§6).

<sup>15</sup> Phillips and Lewis (forthcoming) discuss an issue raised by the studies reviewed above. Structural priming has been observed between sentences that, from the point of view of a sophisticated syntactic theory (e.g., Government and Binding theory), exhibit fine-grained syntactic differences. That is, for the purpose of structural priming,

## §4. The Argument from Garden-Path Effects

### §4.1 Garden Path Phenomena

A classic form of argument for the psychological reality of mental phrase markers begins with the observation that competent language users have problems reading and understanding sentences like (8)-(12), repeated below.

- (8) Jake tells students he intrigues to stay.
- (9) Benny knows the boy hurried out the door slipped.
- (10) The soldier persuaded the radical student that he was fighting in the war for to enlist.
- (11) Aron gave the man who was eating the fudge.
- (12) The old train the children.

From the point of view of formal syntax, all of these sentences contain a local ambiguity that is resolved at or before the end of the sentence. What could explain the fact that even proficient readers encounter measurable processing difficulties with regard to such sentences? A standard explanation appeals to the on-line construction of mental phrase markers.

A parsing routine that computes phrase markers incrementally will update its representation of a sentence in accordance with the words or phrases that it encounters, up to the point at which the sentence becomes syntactically ambiguous. At that point, the parser has to make a choice among the possible ways of continuing the phrase marker that it has thus far constructed.<sup>16</sup> Any ambiguity resolution strategy will sometimes lead a parser to make *incorrect*

---

syntactic structure A might be “the same as” syntactic structure B, even though A and B are, in theory, quite different. Does this show that the mental phrase markers constructed by the HSPM in processing a sentence are not of the same type as the phrase markers assigned to that same sentence by a syntactic theory? The answer is that it doesn’t. Phillips and Lewis point out that this finding can be taken as simply showing that “the structural priming paradigm is a relatively blunt tool for investigating structure, because relatively coarse-grained similarity between structures is sufficient to cause structural priming. This would leave open the possibility that on-line processes build fine-grained structures” (p. 13).

<sup>16</sup> Here, I assume a serial architecture. Parallel models will build multiple grammatically licensed structures and perhaps ranking them. Serial and parallel models yield identical observable results in most (though perhaps not all) psycholinguistic experiments. In those cases where a serial model has made a mistake, it will incur extra computational load by being forced to backtrack. Parallel models will likewise incur extra computational load, on account of their having to re-rank the parses that they have constructed and stored in working memory. For further

choices—i.e., choices that give rise to expectations that the remainder of the sentence will serve to disconfirm. Herein lies the explanation of the aforementioned processing difficulties. In sentences like (8)-(12), the human sentence processor’s preferred structural assignment turns out to be incorrect by the time the sentence is complete. Additional computational load is then incurred in revising the mental phrase marker—a process known as *reanalysis*. This additional processing burden shows up in behavioral and neurophysiological indicators of processing difficulty, such as error rates and delayed reaction times. The success of this explanation of the observed processing difficulties constitutes evidence in favor of the claim that human sentence processing routines construct mental phrase markers.

Countless instances of this argument can be found in the sentence processing literature. To take a classic study, Rayner, Carlson, and Frazier (1983) examined the eye movements involved in reading sentences (24a) and (24b).<sup>17</sup>

- (24) a. The spy saw the cop with a revolver, but the cop didn’t see him.
- b. The spy saw the cop with the binoculars, but the cop didn’t see him.

The data show that readers fixate immediately after the word ‘revolver’ in (24a) for a significantly longer time than they do after the word ‘binoculars’ in (24b). (See fig. 3.) Rayner *et al.* argue that the increased fixation is a result of the fact that the HSPM initially constructs a representation of sentence (24a) in which the prepositional phrase ‘with a revolver’ attaches to

---

discussion of this issue, see Crocker, Pickering and Clifton (2000). Note also that I am *not* equating parallel models with constraint-satisfaction models (to be discussed later in this section). As a matter of sociology, these tend to be associated. But, as Bornkessel-Schlesewsky and Schlewsky (2009: p. 106, fn. 5) point out, the question of whether the HSPM constructs mental phrase markers in parallel is logically distinct from the question of what information it uses to resolve local ambiguities or to rank the parses it constructs. (For instance, Gibson, 1991 endorses a parallel parsing model, but not of the constraint-satisfaction variety.) Nor should we simply assume that parallel parsing models will have a connectionist architecture. The best-known parsing models for classical computational architectures employ the CYK algorithm or the Earley algorithm, both of which are parallel, in the sense that they store all possible parses consistent with the (explicitly represented) grammar in a data structure known as a “table” or a “chart.” (See Jurafsky and Martin, 2007: ch. 12.) I discuss these algorithms in ch. V.

<sup>17</sup> For an extensive discussion of the eye-tracking paradigm, see Rayner (1998).

the verb ‘saw’, not the noun ‘the cop’.<sup>18</sup> Fractions of a second later, this initial attachment preference comes into conflict with a person’s encyclopedic information—specifically, a belief to the effect that one is less likely to use a revolver to see something than to see someone who is in possession of a revolver. This gives rise to a quick *reanalysis* of the sentence, in the course of which the prepositional phrase is attached to the noun ‘the cop’. By contrast, the initial attachment preference in the case of sentence (24b) is consistent with the semantic interpretation of the sentence, so processing is not delayed and fixation times stay low.

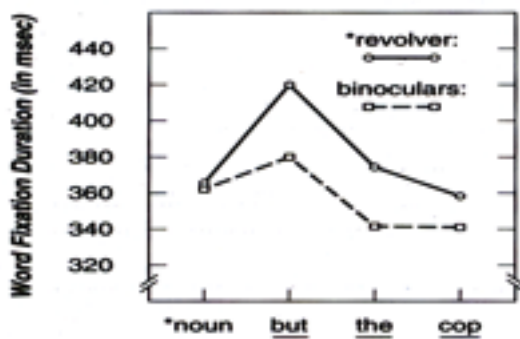


**Figure 3:** Eye-tracking data from Rayner, et al. (1983). The graph depicts a significant spike in fixation times at the critical region of sentence (24), represented by the unbroken line. Note the clear contrast between the fixation times at the critical region in sentence (25), depicted by the broken line.

Traxler and Pickering (1996) report the results of a similar experiment in which an eye tracker was used to measure the difference between fixation times associated with the critical regions in sentences like (25)-(28).

(25) I recognized you and your family would be unhappy here.

(26) I recognized you and your family right away.

(27) I recognized she and her family would be unhappy here.

(28) I recognized that you and your family would be unhappy here.

Sentence (25) exhibits a classic local ambiguity. The verb ‘recognize’ can take as a complement both a noun phrase, as in (26), and a full clause, as in (27). When a sentence contains a

---

<sup>18</sup> This attachment preference is an instance of a more general processing principle, known as Minimal Attachment. I discuss Minimal Attachment below.

complementizer (e.g., ‘that’) after the verb, as in (28), only a sentential complement can follow. However, the complementizer is optional in constructions like (25), making for a local ambiguity that is resolved relatively late in the sentence. In the case of (25), the HSPM prefers the noun phrase continuation.<sup>19</sup> That is, it initially analyses the string ‘you and your family’ as a noun phrase—the object of the verb ‘recognize’. When the auxiliary verb ‘would’ is encountered, the analysis has to be revised in favor of the sentential complement reading. The revision process exerts a computational load on the system, giving rise to elevated fixation times.

The experiment reported in Traxler and Pickering (1996) was designed to test whether, and to what extent, the HSPM makes use of information regarding the syntactic property known as Case. In a common dialect of American English, only a handful of nouns are overtly marked with Case. The pronouns ‘I’, ‘she’, ‘he’, ‘we’, and ‘they’ take different forms depending on their Case, as shown in the following examples:

(29) I kissed her. / She kissed me. / \*Me kissed her. / \*I kissed she. / \*Her kissed me.

(30) They kissed us. / We kissed them. / \*Them kissed us. / \*They kissed we. / \*We kissed they.

Unlike those above, however, the pronoun ‘you’ does not exhibit Case overtly. Its morphology is the same regardless of whether it has Nominative, Accusative, or Dative Case.

(31) I kissed you. / You kissed me. / I gave a gift to you. / I gave you a gift.

It is in this respect that sentences (25) and (27), repeated here, differ.

(25) I recognized you and your family would be unhappy here.

(27) I recognized she and her family would be unhappy here.

The fact that the pronoun ‘she’ in (27) overtly carries Nominative Case can provide an important clue to the HSPM, for it entails that ‘she’ is the subject of a subordinate clause—a sentential

---

<sup>19</sup> Again, this preference is in accordance with Minimal Attachment, discussed below.

complement—rather than the object of the verb ‘recognize’. Because the HSPM is sensitive to this information, participants do not fixate on the word ‘would’ in the sentence (27) for an extra period of time, as they do in the case of sentences like (25). With this datum in hand—as well as some ancillary data that control for various confounds—Traxler and Pickering conclude that the HSPM makes rapid use of Case information in the early stages of processing.

Similar questions can be raised with regard to information about pronoun gender. Sentences like (32) demonstrate that, lacking a complementizer, the verb ‘know’ has a *very* strong preference for an NP complement, as against a sentential complement.

(32) I know the man who believes the Queen shaves himself.

The sentence in (32) presents enormous difficulties for most readers. Indeed, the effect is so pronounced that readers will not only fail to correctly answer simple questions on the basis of the information it contains (e.g., “What fact is the speaker claiming knowledge of?”) but they will routinely insist, sometimes doggedly, that the sentence is outright ungrammatical. That the sentence is, in fact, perfectly grammatical can be seen by attending to sentence (33), which differs from (32) only in the addition of the optional complementizer ‘that’.

(33) I know that the man who believes the Queen shaves himself.

Once it’s made plain that the complement of ‘know’ is sentential, there is no longer anything impeding the binding of the reflexive pronoun ‘himself’ to the NP ‘the man who believes the Queen’. Indeed, given that both the pronoun and the NP are marked with a masculine feature—to which the HSPM is demonstrably sensitive—this is the natural binding relation.

The studies reviewed above, and many others like them, serve to demonstrate three points. First, in the course of processing, phrase structure is imposed on the linguistic input. That is, a mental phrase marker is constructed, in a manner that is sensitive to the syntactic



properties of the stimulus. Second, garden-path effects are not limited to cases in which the processing difficulty is *consciously* detected, as it typically is in the course of processing the following sentences:

(34) The ship floated down the river sank.

(35) Fat people eat accumulates.

(36) The cotton clothing is made of grows in Arkansas.

(37) The daughter of the king's son admires himself.

Sentences like (34)-(37) have been dubbed ‘conscious garden paths’. Though they are exceedingly rare, one finds them mentioned with some frequency in the literature outside of psycholinguistics—particularly in neighboring disciplines like philosophy. The reason is not far to seek; such cases illustrate vividly the distinction between grammaticality and acceptability.

Unfortunately, the press these sentences have gotten has engendered some confusion on the part of nonspecialists; it gives rise to the assumption that *all* garden-path effects are conscious—i.e., that if an anomaly is not *consciously* detected, then it has not occurred. This is, however, a mistake, due either to overgeneralization or to a conflation of mentality with consciousness. There is simply no compelling reason to think that garden-path effects always give rise to a *conscious* awareness of anomaly.<sup>20</sup> Nonconscious awareness of anomaly can be observed from the third-person perspective by close (theory-laden) attention to fine-grained

---

<sup>20</sup> “The fact that that hearers are not always conscious of having made a mistake in the analysis of such sentences (as they are for notorious garden-paths such as *The horse raced past the barn fell*) is not, we submit, a good argument against this kind of perceptual complexity” (Frazier and Fodor, 1978: p. 296). See Pritchett (1992: p. 33) for illuminating reflections on this point. Pritchett argues that although conscious reflection cannot tell us anything about sentence processing, the fact that some processing difficulties are consciously registered while others are not nevertheless constitutes evidence that the former are more serious than the latter.

behavioral data—e.g., the eye tracking data obtained by Rayner, *et al.*—and, more recently, neurophysiological markers.<sup>21</sup>

The third point illustrated by the studies reviewed above is that the HSPM seems to construct mental phrase markers in accordance with a quite general ambiguity resolution strategy, known in the literature as *Minimal Attachment* (Frazier, 1979). The strategy, simply put, is to attach incoming material into the existing mental phrase marker in such a way as to minimize the number of nonterminal nodes in the resulting structure. Minimal Attachment is a least-effort principle. Operating in accordance with it, the parser avoids postulating more structure than it absolutely must at any given point. An application of Minimal Attachment to sentences (24a) and (24b) is illustrated in fig. 4 below.

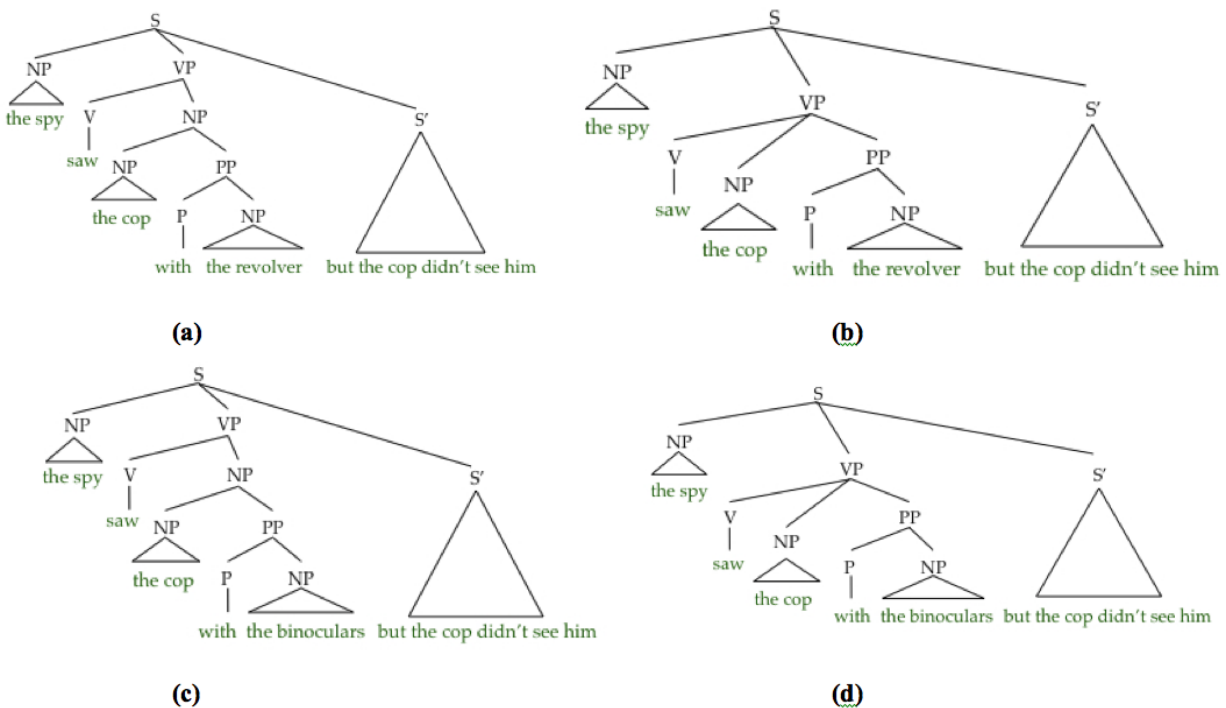


**Figure 4:** In accordance with the principle of Minimal Attachment, when sentence (24a) is the input, the HSPM will initially fail to build the plausible and grammatical structure in (a), preferring instead the semantically implausible structure in (b). The reason is that, in (a), the verb phrase (VP) dominates *six* nonterminal nodes—one more than it does in the structure

<sup>21</sup> For a detailed and compelling treatment of the varieties of nonconscious awareness, see Rosenthal (2005).

depicted in (b). The additional nonterminal node in (a) is an NP, which is needed to accommodate the attachment of the prepositional phrase ‘with the binoculars’ to the NP ‘the cop’. Minimal Attachment rules against constructing a mental phrase marker with this additional internal structure. When sentence (24b) is the input, Minimal Attachment likewise dictates avoiding the structure depicted in (c) and building instead the structure shown in (d). In this case, the attachment preference turns out to be semantically plausible. These diagrams are adapted (with modifications) from a PowerPoint presentation by Alan Garnham. The original slides are available at: [www.lifesci.sussex.ac.uk/home/Alan\\_Garnham/Teaching/LTM/sentproc.ppt](http://www.lifesci.sussex.ac.uk/home/Alan_Garnham/Teaching/LTM/sentproc.ppt) The tree structures in these diagrams are far simpler than would be posited in contemporary syntactic theories, but they are sufficient to illustrate the structural distinctions relevant to the present discussion.

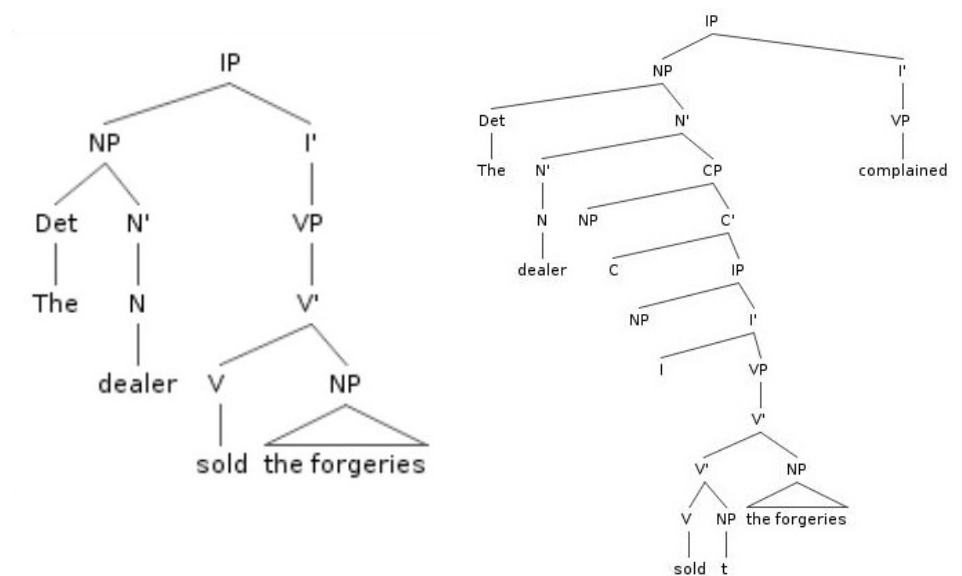


**Figure 5:** In accordance with the principle of Minimal Attachment, the HSPM assumes that the noun ‘the dealer’ is the subject of that verb ‘sold’. It thus attaches the verb to the existing structure in the manner depicted in the left panel. Subsequent input reveals that the HSPM’s assumption was incorrect, thus necessitating reanalysis. The difficulty of reanalysis in this case is a function of the sheer amount of additional structure needed to accommodate a passive participle reading.

Minimal Attachment also accounts for the conscious garden path effect associated with sentences that have a local main-clause/relative-clause ambiguity, such as the famous sentence (38) and its cohorts, (39)-(41).

(38) The horse raced past the barn fell.

(39) The ship floated down the river sank.

(40) The dealer sold the forgeries complained.

(41) The man sent the letter cried.

The verb-forms ‘floated’, ‘raced’, ‘sold’, and ‘sent’ are ambiguous. They can serve either as past-tense verbs that are part of a main clause or as past participles that introduce a reduced relative clause. (In these cases, the optional complementizer ‘that’ has been omitted.) The locally ambiguous structures associated with these sentences are illustrated in fig. 5.

Traxler and Pickering’s results can be used to argue that Case information comes into play *before* Minimal Attachment applies. Consider again sentences (25) and (27), repeated here.

(25) I recognized you and your family would be unhappy here.

(27) I recognized she and her family would be unhappy here.

The syntactic structures of (25) and (27) are represented in fig. 6.

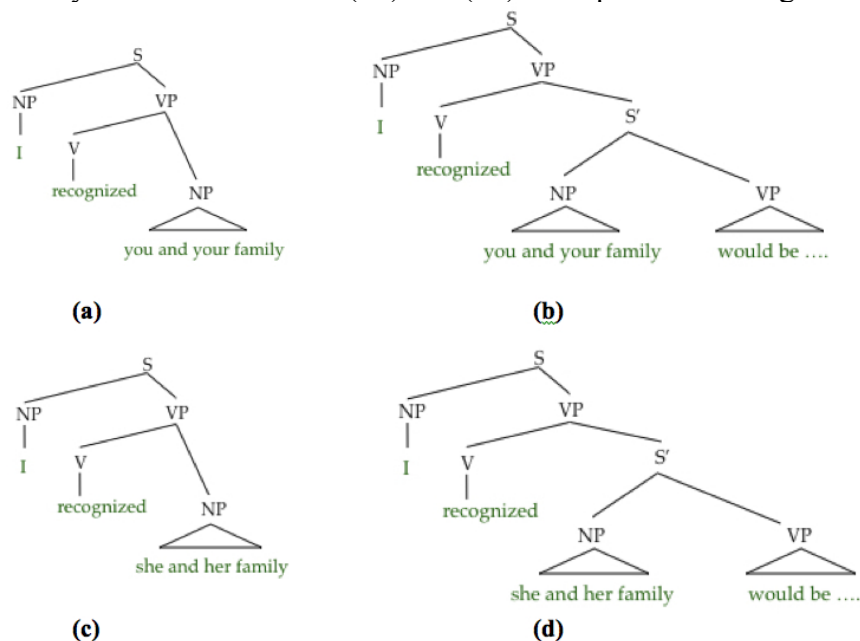


**Figure 6:** In accordance with the principle of Minimal Attachment, the HSPM assumes that the noun phrase ‘you and your family’ is the object of the verb ‘recognize’, thus building the structure depicted in (a). This assumption is disconfirmed by subsequent input, giving rise to a process of reanalysis, which results in structure (b). However, when Case information is available, as in sentence (27), structure (c) is never built. Evidently, the HSPM uses the information about Case *prior* to applying the principle of Minimal Attachment, thus building the structure in (d).

Minimal Attachment is one of a handful of principles that, taken together, explain a wide range of the HSPM's ambiguity resolution preferences. Another such principle is Late Closure, which dictates that the parser will incorporate newly encountered material into the most recent phrase or clause of the mental phrase marker that it has already constructed. Late Closure is invoked to explain the HSPM's preference in cases like the one illustrated in fig. 7.<sup>22</sup>

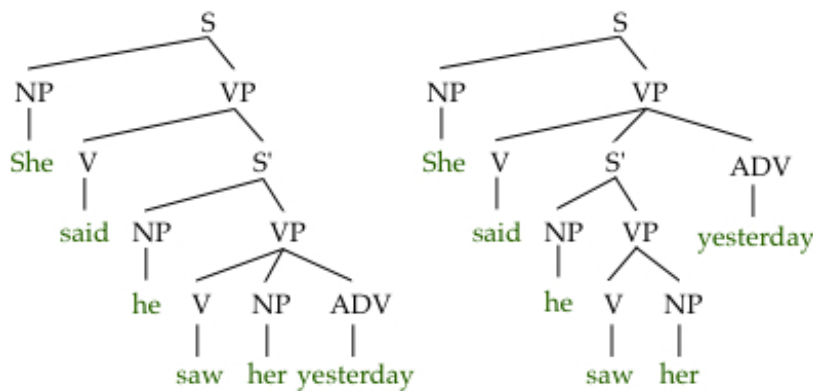


**Figure 7:** Having built a structure for the input ‘She said he saw her’, the HSPM receives the adverb ‘yesterday’. The grammar licenses two possible attachments, represented by the left and right panels. In accordance with the principle of Late Closure, the HSPM resolves this ambiguity in favor of the structure depicted in the left panel. That is, it attaches the adverb to the most recent phrase of the structure it had already built—in this case, the verb phrase ‘saw her’.

Both Minimal Attachment and Late Closure are close cousins of a number of other sentence processing principles that have been discussed over the past several decades. In a remarkably prescient work, Kimball (1973) proposed seven principles of sentence processing designed to account for the way in which the HSPM assigns a phrase-structure analysis to ambiguous input. These principles differed in a crucial respect from the grammatically

<sup>22</sup> The Late Closure principle encountered a bit of trouble when it was discovered that in Spanish and Dutch (and perhaps other languages), adjuncts sometimes prefer to attach high in the tree structure, rather than low as Late Closure would predict. (See Cuetos and Mitchell, 1988.) However, this issue was effectively resolved by Fodor (1998a), who proposed that prosody—whether explicit or implicit—determines whether or not Late Closure applies. Roughly, the idea is that prosodically “heavy” material is attached high, while prosodically light material is attached low. The details of the Late Closure crisis and its resolution are beyond the scope of the present discussion.

unmotivated heuristics proposed by Fodor, Bever, and Garrett (1974). Unlike those authors, Kimball assumed that the HSPM would assign syntactic structure—i.e., build its mental phrase markers—in a way that draws on the rules of a surface-structure grammar.<sup>23</sup> More precisely, his view entailed that the processing principles that are true of the HSPM would make ineliminable reference to the proprietary notions of an independently motivated syntactic theory—e.g., the notion *number of nonterminal nodes*.

Thus, in addition to explaining a broad range of data, Kimball's work made it possible to see, if only in dim outline, how one might incorporate an independently motivated syntactic theory into a model of sentence processing—an idea that Fodor, Bever, and Garrett (1974) rejected, for reasons that have come to be recognized as spurious.<sup>24</sup> Kimball's approach thus offered psycholinguists an alternative to the proposals in Fodor, Bever, and Garrett (1974), allowing them to seek more concise and predictive principles that accord with the theories emerging from the field of descriptive syntax.<sup>25</sup>

In a series of influential follow-ups to Kimball's pioneering work, Frazier and Fodor (1978) and Frazier (1979) refined Kimball's principles into what we now know as Minimal Attachment and Late Closure. Other theorists extended Kimball's ideas in similar directions. For instance, Wanner and Maratsos (1978) describe a nondeterministic Augmented Transition Network (ATN) parser that mimics, in its operations, a recursive phrase structure grammar with

---

<sup>23</sup> Kimball discussed only the parsing of surface structures without traces. He did not offer a comparably detailed treatment of transformational dependencies (§5). In this respect, he did not hew to the prevailing syntactic theory. Fodor (1978) discusses an extension of Kimball's approach that deals with transformational phenomena.

<sup>24</sup> See Berwick and Weinberg (1984: ch. 2), Phillips (1994, 1996) and Phillips and Lewis (forthcoming). I discuss the issue in detail in ch. V (§4).

<sup>25</sup> Frazier (1979) examines in detail the merits and drawbacks of Kimball's approach. A more recent treatment appears in Pritchett (1992: pp. 26-30), where Kimball's work is situated in the context of subsequent theories of sentence processing. Pritchett's own theory of sentence processing is directly inspired by Kimball's work, though it draws on a more recent grammar, to which Kimball did not have access.

built-in attachment preferences. Similarly, Marcus (1980) proposed a deterministic parsing model that implemented some of the same principles. Ford, Bresnan, and Kaplan (1982) proposed principles that they called “Lexical Preference” and “Final Arguments,” which do much of the same work, though in the context of a parser modeled on Lexical Functional Grammar. Pereira (1985) discusses a “shift-reduce” parser where Kimball’s principle of Right Association is implemented as a preference for shifts over reductions, and Fodor and Frazier’s principle of Minimal Attachment is implemented as a preference for maximally long reductions. Subsequently, Pritchett (1992) and Frazier and Clifton (1996) proposed refinements of Kimball’s principles that appeal to the  $\theta$ -roles of Government and Binding theory and the related distinction between arguments and adjuncts.<sup>26</sup>

It is important to be clear about the status of the processing principles discussed above. To my knowledge, it has never been suggested that Minimal Attachment or Late Closure are *represented in* the HSPM, or that the HSPM has *knowledge of* them, however tacit. Rather, these are intended to be *descriptive* principles; they are *true of* the HSPM in much the same way that the principles of celestial mechanics are *true of* our solar system. Admitting the descriptive accuracy of such principles does not, then, automatically commit one to any view regarding the psychological reality of syntactic rules or principles—at least not without additional argument. Still, given that Minimal Attachment, Late Closure, and related principles make ineliminable appeal to the proprietary notions of syntactic theory, it follows that the HSPM works *in accordance* with a particular theory of syntax. In ch. V, we shall tackle the question of how to properly cash out the notion of “working in accordance with.” I shall argue there that the correct way of cashing it out commits us to a position stronger than what Devitt (2006) calls “the

---

<sup>26</sup> I discuss the ATN model, the shift-reduce parser, and  $\theta$ -roles in some detail in ch. V.

minimal position (M)” which he formulates as follows:

(M) A competence in a language, and the processing rules that govern its exercise, respect the structure rules of the language: the processing rules of language comprehension take sentences of the language as inputs; the processing rules of language production yield sentences of the language as outputs. (p. 57)

If we take Minimal Attachment, Late Closure and the like to be true of the HSPM—as the data suggest—then we should not rest content with the minimal claim embodied in (M). The relation between a descriptive grammar and the operations of the HSPM is not *simply* that the latter is able to process sentences described by the former. The fact that the HSPM works “in accordance with” a descriptive grammar entails that the rules or principles of that grammar are either represented or embodied in HSPM. We return to this issue in ch. V.

#### *§4.2 Initial Parsing Preferences: On the Use of Non-syntactic Information*

There is ongoing debate in psycholinguistics about what properties of the stimulus are encoded in mental phrase markers, as well as about the character of the parsing routines that construct them. For instance, Ni, Crain, and Shankweiler (1996) write:

The existence of garden path effects and semantic anomalies is evidence that the human sentence processing mechanism (the parser) makes rapid decisions about which alternative, grammatically well-formed structural representation to adopt when the input is ambiguous. The factors that influence the decision making of the parser and the manner in which the parser is influenced are matters of controversy, however. (p. 285)

In a similar vein, Traxler and Pickering (1996) describe the state of the field as follows:

Researchers in the field of sentence processing have identified many potentially relevant sources of information [that the HSPM might make use of]. They do not, however, necessarily agree about which information is used during initial processing. These information sources include lexical preferences of individual verbs (Ford, Bresnan, & Kaplan, 1982), plausibility (Rayner, Carlson, & Frazier, 1983), discourse context (Crain & Steedman, 1985), punctuation (Mitchell & Holmes, 1985), prosody (Marslen-Wilson, Tyler, Warren, Grenier, & Lee, 1992), gender agreement (Brysbaert & Mitchell, 1996), and case-marking (Trueswell et al., 1993).



The most widely known restricted-parsing account is the “garden-path” model (Frazier, 1979, 1987). On this account, initial parsing ignores plausibility and discourse context. These factors can therefore only play a role in reanalysis (e.g. through the “thematic processor” discussed by Rayner et al., 1983). The parser can make use of prosodic information if it is inconsistent with particular syntactic analyses (though it may be very difficult to specify when a prosodic contour is inconsistent with a particular analysis and when it is merely preferentially associated with a particular analysis). The parser probably also employs punctuation such as commas if it is inconsistent with particular analyses (e.g. Mitchell, 1987). (p. 993)

Traxler and Pickering’s remarks highlight an important issue concerning what information the HSPM makes use of in the *initial* stages of processing. In particular, controversy surrounds what they call the “restricted-parsing” account. On this account, the assignment of syntactic structure to the input is accomplished in stages, where in the first stage the HSPM takes into account only the *syntactic* properties of the input and uses only *that* information in resolving structural ambiguities, in accordance with Minimal Attachment and Late Closure. In the later stages, the HSPM may revise its initial decisions on the basis of *nonsyntactic* information.

The question of early parsing preferences has generated a heated debate. Bornkessel-Schlesewsky and Schlesewsky (2009) cast the issue as follows:

[A]s numerous empirical studies have demonstrated effects of non-syntactic information types (such as discourse context or pragmatic plausibility) on sentence processing, the crucial question underlying the debate on modularity vs. interactivity is not whether these information sources influence processing (which appears indisputable) but *when* this influence occurs. (p. 98)

The disagreement, then, is between theorists who favor a “modular” parsing model, according to which only syntactic information influences early parsing decisions, and those who favor an “interactionist” model, on which *all* types of information exert an influence from the very outset of the parsing process. This debate has a bearing on issues that will arise in ch. IV in connection with Schank et al.’s “integrated processing hypothesis.” For that reason, I will devote the

remainder of this section to a discussion the evidence favors syntax-first models over their interactionist rivals.

In an influential early discussion, Forster (1979) presented the schematic model of the HSPM depicted in fig. 8. As Frazier (1988) notes, this model is highly modular. It predicts, among other things, that there will be no syntactic effects on the initial stages of phonological processing, and no semantic or pragmatic effects on the initial stages of syntactic processing.

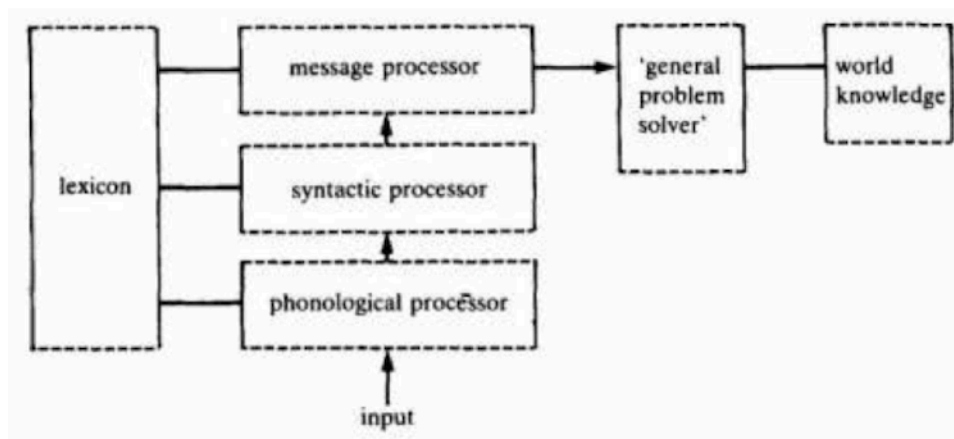


**Figure 8:** A modular model of the HSPM, due to Forster (1979). The lines with arrowheads represent the flow of information processing. The lines without arrowheads represent the accessibility and exchange of information.

Frazier discusses a range of experimental confirmations of the predictions mentioned above. First, when either of a pair of homonyms (e.g., ‘right’ and ‘write’) are heard, both are activated in the lexicon, regardless of syntactic plausibility. This demonstrates that syntactic information—even a rather strong syntactic bias—does not constrain the initial steps of lexical retrieval on the basis of phonological input. Next, passive constructions take longer to process than active ones, regardless of whether the passive sentence is semantically “reversible”—i.e., equally plausible when the subject and object are swapped. Similarly, sentences take longer to read when their subjects are extraposed than when they appear in their canonical position, again

regardless of context and semantic disambiguation. For instance, a sentence like (42) takes more time to read than a sentence like (43).

(42) Inside the house were two detectives.

(43) Two detectives were inside the house.

More generally, even when prior semantic context would clearly disambiguate between two grammatically licensed attachments, the HSPM does not make use of this information in its initial parsing decisions. A celebrated example from Ferreira and Clifton (1986) illustrates this point. The experimenters used materials like sentence (44) and (45) in an eye tracking study aimed at discerning the HSPM's attachment preferences in the face of a syntactic main-clause/relative clause ambiguity that contains a built-in semantic disambiguation.

(44) The witness examined by the lawyer turned out to be unreliable.

(45) The evidence examined by the lawyer turned out to be unreliable.

Before the word 'by' appears, the verb 'examined' is ambiguous. It might be a tensed verb that introduces a main clause, as in sentence (46).

(46) The witness examined the specimen.

But, as sentences (44) and (45) demonstrate, it might also be a passive participle that introduces a reduced relative clause. As we saw above, the HSPM consistently prefers the main clause reading, in accordance with Minimal Attachment. This preference gives rise to a strong, conscious garden path effect in famous cases like (34)-(37), repeated here.

(34) The ship floated down the river sank.

(35) The horse raced past the barn fell.

(36) The dealer sold the forgeries complained.

(37) The man sent the letter cried.

In sentences (44) and (45), ‘examined’ turns out to be a passive participle, so we would expect—on the basis of syntactic considerations alone—to find a garden path effect in these cases as well. However, as noted above, these sentences were constructed in such a way that semantic information could, in principle, be used to avoid the incorrect reading. If HSPM has access to semantic information from the outset, then it should take into account the *semantic* fact that witnesses are typically capable of examining things while evidence is typically not. In this case, we would expect the HSPM to predict the relative clause reading of (45) and, hence, to encounter no difficulty when the word ‘by’ disambiguates the structure in favor of that reading.

This is not, however, what Ferreira and Clifton observed. They report that the fixation times at the word ‘by’ were quite comparable in the case of both (44) and (45). This suggests that initial parsing decisions are made without the influence of semantic information. Ferreira and Clifton’s result was subsequently replicated by Rayner, Garrod, and Perfetti (1992), who performed a similar experiment using materials designed to test, in addition, for effects of discourse focus. They report their findings in the following passage:

Our primary finding was that subjects were garden-pathed even when there was biasing context. However, when the target sentence was in the discourse focus, subjects were able to recover more readily from their initial erroneous parse of the sentence. The data thus support models of sentence parsing that postulate that the parsing of a sentence is based upon structurally based principles and the influence of semantic or pragmatic information makes itself felt only after the initial parsing decision has been made. (p. 109)

The view supported by these and several other convergent studies briefly came under fire when Trueswell, Tanenhaus, and Garnsey (1994) cast doubt on the adequacy of the methods and materials used in the original Ferreira and Clifton (1986) study. In an attempt to replicate Ferreira and Clifton’s results using improved methods and materials, Trueswell *et al.* claimed to have found that semantic information *does* eliminate (or reduce) the garden path effect. For

some time, this was seen as a demonstration of the failure of the syntax-first approach.<sup>27</sup>

However, in a more recent study, Clifton, et al. (2003) reexamined Trueswell et al.'s claims and found no evidence supporting their main conclusions. In their concluding discussion, they write:

Our findings ... cast substantial doubt on Trueswell et al.'s (1994) frequently cited conclusion that semantic factors can override syntactic processing biases. A favorite argument against serial models is that the syntactic preferences they propose can in fact be completely eliminated by semantic and pragmatic factors, and Trueswell et al. is frequently cited in support of this argument. While it is certainly possible that syntactic preferences can be eliminated in some domains, it appears that the reduced relative clause domain studied by Trueswell et al. is not one of them. (Clifton, et al., 2003: p. 331).

In discussing these issues, Pickering and van Gompel (2006: pp. 474-5) point out that, until recently, it has been difficult to obtain experimental evidence that tells *clearly* against constraint-satisfaction models.<sup>28</sup> The reason is that any time processing difficulties are observed in the course of parsing a locally ambiguous input, the constraint-based models can account for this by claiming that the effect is due to a computationally taxing *competition* within the HSPM. That is, the HSPM may, at those points, be having trouble deciding between two syntactically distinct structures that have received comparable support from the evidence it has thus far encountered. For instance, in processing a sentence like (47), the HSPM might rank *equally highly* the syntactic preference for a main-clause reading and the semantic preference for burglars that get arrested, rather than ones who do the arresting themselves:

(47) The burglar arrested by the cop was in court.

---

<sup>27</sup> Clifton et al. (2003) speculate that the sudden shift of the tide, as exemplified by introductory textbooks in cognitive psychology, was not entirely a function of the experimental results. They attribute it in part to (i) the fact that psychologists became wary of the ever-changing syntactic theories, hence eager to conduct future experiments on the basis of something other than syntax, and (ii) the fact that constraint-satisfaction models came to be associated with connectionism, which was rapidly gaining popularity in the mid-1990s.

<sup>28</sup> For this reason, Clifton et al. (2003) do not claim to have resolved the debate. They are clear that their findings “do not provide a basis for choosing between serial, depth-first parsing models and parallel, constraint satisfaction models” (p. 331).

The data are thus consistent with both the syntax-first view, which explains the processing difficulty by appeal to a computationally taxing reanalysis procedure, and the constraint-satisfaction view, which explains it by appeal to a local competition between syntactic and semantic constraints.

Recent studies have sought to overcome this difficulty by turning away from *locally* ambiguous inputs and focusing instead on the predictions that the rival models make about how the HSPM will deal with *globally* ambiguous inputs. Specifically, unlike syntax-first models, constraint-satisfaction models predict that globally ambiguous sentences like (48) should be more difficult to process than related unambiguous sentences like (49) and (50).

(48) I read that the bodyguard of the governor retiring after the troubles is very rich.

(49) I read that the governor of the province retiring after the troubles is very rich.

(50) I read that the province of the governor retiring after the troubles is very rich.

The reasoning here is straightforward: Constraint-satisfaction models posit a computationally taxing competition between syntactic and semantic information. It follows that globally ambiguous sentences like (48) will engender such a competition, whereas their unambiguous counterparts, e.g., (49) and (50) will not. The data, however, do not confirm this prediction. Van Gompel et al. (2005) present results that show globally ambiguous sentences to be *easier* to process than their unambiguous counterparts.

### §4.3 Conclusion

In this section we have surveyed a sample of the studies that shed light on the internal workings of the HSPM by examining fine-grained behaviors such as eye movements. The studies uniformly support the claim that the HSPM constructs mental phrase markers. They also provide information about the character of its processing routines, especially with respect to

ambiguity resolution. In the initial stages, the HSPM's parsing preferences are captured by two principles—Minimal Attachment and Late Closure—whose formulation requires ineliminable reference to abstract grammatical notions. Later stages of processing take into account nonsyntactic information, occasionally giving rise to a computationally costly process of revision known as *reanalysis*. In the course of reanalysis, the syntactic relations encoded by a mental phrase marker are adjusted in accordance with *both* grammatical constraints *and* various types of nonsyntactic information.

In the next section, we examine another well-studied aspect of the comprehension process—the assignment of filler-gap relations. Studies of this phenomenon shed further light on the syntactic information encoded in mental phrase markers, as well as on the processing routines of the HSPM. Indeed, a number of theorists have expressed the hope that these studies will allow us to discern which descriptive grammar most closely resembles the grammar employed by HSPM. We will also encounter two new experimental paradigms. Like those discussed in §3, the experiments used in studying filler-gap processing also make use of the phenomenon of priming, though not *structural* priming per se.

## **§5. The Argument from Filler-Gap Processing**

Consider the questions and relative clause constructions, exemplified in (51)-(54).

(51) What instrument do you play?

(52) I play the guitar that my parents bought.

(53) Whom did they consult before buying it?

(54) I don't remember which of their friends they consulted.

The verbs ‘play’, ‘bought’, and ‘consulted’ are transitive—that is, they take noun phrase as a complement, as in (55)-(57).

(55) I play the guitar.

(56) My parents bought the guitar.

(57) They consulted their friends.

In declarative sentences like (55)-(57), the position of the two nouns around the verb makes it clear who is performing the action denoted by the verb and what the action is being performed on. For instance, the position of the nouns in (56) makes clear who is doing the buying and what is being bought. However, in (51)-(54) the noun phrases are not in their canonical positions.

Rather, they appear at the front of a clause. Yet the verbs are understood as having those nouns as their objects. The nouns must therefore maintain some sort of long-distance relationship with the object positions of the verbs. Different linguistic theories employ different formal devices to represent this relationship. Government and Binding (GB) theory (Chomsky, 1986; Haegeman, 1994) uses so-called “empty categories.” An empty category is a theoretical posit. While not overtly present in writing and speech, it is assumed to occupy a position in the underlying syntactic structure of a sentence and to behave in many ways as though it *were* overtly present.<sup>29</sup>

Sentences (62) and (63) are relatively uncontroversial examples of constructions in which, according to GB, empty categories play this role.

(62) Alia plays the keyboard and Jeffrey the bass.

(63) Igor loves Alia and she him.

---

<sup>29</sup> Empty categories are sometimes called “phonologically null” categories, because they are not spoken or heard. Of course, empty categories are also *orthographically* null. See ch. I (§4.1) for a discussion of the ontological status of empty categories.



Sentences (62) and (63) are understood as having a verb between the two final nouns. And not just *any* verb; sentence (62) cannot be interpreted as meaning that Alia plays the keyboard and Jeffrey *sells* the bass. The missing verb in (62) must be ‘play’ and in (63), ‘love’. This relationship can be represented as follows:

(64) Alia will play<sub>i</sub> the keyboard and Jeffrey [<sub>VP</sub> EC<sub>i</sub> the bass].

(65) Igor loves<sub>i</sub> Alia and she [<sub>VP</sub> EC<sub>i</sub> him].

The empty category (EC) in (64) and in (65) is said to be *coindexed* with the overt verb in the prior clause. The subscript ‘i’ encodes a relationship of identity between the overt verb and its missing counterpart, which in turn constrains the semantic interpretation of the sentence. The overt verb in (64) and (65) is sometimes called a *filler* and the empty category—or the position it occupies—is a *gap*.

Returning now to sentences (51)-(54), we see that the same device can serve to encode the relationship between verbs and their objects, as in (66)-(69).

(66) [What instrument]<sub>i</sub> do you play EC<sub>i</sub>?

(67) I play [the guitar]<sub>i</sub> that my parents bought EC<sub>i</sub>.

(68) [Whom]<sub>i</sub> did they consult EC<sub>i</sub> before buying the guitar?

(69) I don’t remember [which of their friends]<sub>i</sub> they consulted EC<sub>i</sub>.

In these cases, the overt noun or *wh*-phrase is the filler, while the empty category—or the position it occupies—is the gap.

Passive sentences can receive a similar treatment. Recall from §3 that Government and Binding theory posits a level of syntactic structure, known as D-structure, at which noun phrases are in their canonical positions relative to verbs. Movement operations then apply to these nouns, yielding the surface forms (S-structures) that we read and hear. These S-structures,

however, are also assumed to contain empty categories that serve to mark the relationship between the displaced nouns and their D-structure positions. Consider (70) and (71):

- (70) The boat was carried by five people.  
(71) Five people were carried by the boat.

These sentences have the following D-structures:

- (72) was [<sub>VP</sub> carried [<sub>NP</sub> the boat] by five people.  
(73) were [<sub>VP</sub> carried [<sub>NP</sub> five people] by the boat.

After the movement operations apply, the resulting S-structures are (74) and (75).

- (74) [<sub>NP</sub> The boat]<sub>i</sub> was carried EC<sub>i</sub> by five people.  
(75) [<sub>NP</sub> Five people]<sub>i</sub> were carried EC<sub>i</sub> by the boat.

Government and Binding theory posits empty categories to account for two other syntactic phenomena: raising and control. The first has to do with so-called *raising verbs*, such as ‘seem’ and ‘appear’, in sentences like (76) and (77).

- (76) Jennifer appears to be reading a book about neuroscience.  
(77) Igor seems to enjoy action movies.

In these sentences, the verbs ‘reading’ and ‘enjoy’ are understood as having a subject. For instance, in (76) the subject of ‘reading’ is ‘Jennifer’. However, the subjects in sentences like (76) and (77) are not located in their canonical position; they have been “raised” to a position in the higher clause at the front of the sentence. One way to account for the semantic relationship between the lower verb and its subject is to posit an empty category in the subject position of that verb and coindex that empty category with the noun phrase that is understood as its subject. Here again, the understood subject is taken to be located in the subject position at D-structure and moved to its new position at S-structure. The movement leaves an empty category in its wake—a kind of trace. Thus, sentences (76) and (77) are taken to have the following S-structure representations.

- (78) Jennifer<sub>i</sub> appears EC<sub>i</sub> to be reading a book about neuroscience.  
 (79) Igor<sub>i</sub> seems EC<sub>i</sub> to enjoy action movies.

Similar considerations motivate the posit of an empty category in constructions involving so-called *control verbs*, exemplified in (80) and (81).

- (80) Jeffrey tried to record a song.  
 (81) Konstantine promised to attend the party.

As with the cases discussed above, the verbs ‘record’ and ‘attend’ are understood as having subjects. This relation is captured by the empty categories displayed below.

- (82) Jeffrey<sub>i</sub> tried EC<sub>i</sub> to record a song.  
 (83) Konstantine<sub>i</sub> promised EC<sub>i</sub> to attend the party.

In sum, empty categories allow us to encode various semantically relevant structural relationships between the words in a sentence. Let us distinguish the four types of empty categories introduced thus far. Following convention, I use the labels ‘trace’ and ‘PRO’.

- |                                                                                         |                    |
|-----------------------------------------------------------------------------------------|--------------------|
| (84) Igor loves <sub>i</sub> Alia and she EC <sub>i</sub> him.                          | VP-ellipsis marker |
| (85) [What instrument] <sub>i</sub> do you play EC <sub>i</sub> ?                       | <i>wh</i> -trace   |
| (86) [ <sub>NP</sub> The boat] <sub>i</sub> was carried EC <sub>i</sub> by five people. | NP-trace           |
| (87) Jeffrey <sub>i</sub> tried EC <sub>i</sub> to record a song.                       | PRO                |

Given that empty categories play an explanatory role in GB theory, we can raise a question about whether they “psychologically real.” Does the HSPM construct mental phrase markers that contain, e.g., *wh*-traces? It is possible, after all, that the HSPM encodes filler-gap relations in some other way. Alternative descriptive grammars—rivals of GB—do not posit traces or PRO. Specifically, Lexical Functional Grammar and Head-driven Phrase Structure Grammar

encode filler-gap relations without positing empty categories.<sup>30</sup> Hence a demonstration of the psychological reality of empty categories can be used to argue that Government and Binding theory (or the newer Minimalist grammars) more closely resemble the grammar employed by the HSPM than rival grammatical formalisms.<sup>31</sup>

Fodor (1989, 1991) summarizes a number of early experiments designed to address this issue. The most straightforward of the bunch demonstrates what psycholinguists call “the filled-gap effect.” This is similar to the garden-path effect that we examined in §4; in both cases, the HSPM makes a prediction that is disconfirmed by later input, giving rise to computationally costly revisions that are, in turn, reflected in longer reading times. What’s distinctive about the filled-gap effect is that it has to do with the HSPM’s prediction that some position in the syntactic structure of the input will contain a trace.

Limiting our attention for the moment to *wh*-traces, we can begin by considering sentences (88) and (89), which demonstrate that such a trace can appear in a variety of locations in the input sentence.

(88) Who<sub>i</sub> could the little child have forced *wh*-trace<sub>i</sub> to sing those stupid French songs for Cheryl last year?

(89) Who<sub>i</sub> could the little child have forced us to sing those stupid French songs for *wh*-trace<sub>i</sub> last year?

---

<sup>30</sup> This is an oversimplification. For details regarding the way in which LFG and HPSG treat *wh*-constructions and relative clauses see Bresnan (1978, 2001) and Pollard and Sag (1994), respectively. For a briefer treatment of the issue, with a focus on its relevance for psycholinguistics, see Fodor (1989: p. 177-186) and Featherston (2001)

<sup>31</sup> As noted at the end of ch. II, there are methodological grounds for initially favoring the hypothesis that the HSPM employs whatever grammar happens to satisfy the desiderata of formal linguistic theory. For further discussion of this issue, see Steedman, 1985: p. 361; 2000: pp. 227-8; Fodor, 1989: pp. 177-8; Featherston, 2001: pp. 2-3. Still, hypothesis may well be false; the psychologically real grammar might not meet the compactness/non-redundancy/elegance constraints imposed by syntacticians (Soames, 1984; Stabler, 1984). On the other hand, Jackendoff (2002) argues that syntacticians *ought not* impose such constraints, precisely because biological and psychological phenomena are rarely neat or “perfect,” in the sense assumed by Chomsky’s Minimalist Program.

The HSPM is demonstrably sensitive to the fact that the word ‘Who’ is an antecedent, to which a *wh*-trace will have to be bound—i.e., a filler awaiting a gap. Thus, it faces the task of actively searching for legitimate positions at which to posit the *wh*-trace. The filled-gap effect consists in the fact that the HSPM predicts that the gap will occur after ‘forced’, in both (88) and (89). In the case of (88), the prediction is correct. But in the case of (89), it leads the parser astray, giving rise to measurable processing difficulties at just the point in the sentence where the word ‘us’ occupies the predicted position of the *wh*-trace.

Fodor (1989) notes that there is no *obvious* reason why the HSPM should predict a gap where there is none, instead of simply waiting to see whether some overt material occupies that position. She writes:

Gap anticipation of this kind falsified Fodor’s (1978) proposal that the processor, when in need of a certain constituent (say, a noun phrase), will always try the next constituent in the input to see if it is what it needed, before concluding that the constituent is missing and that an empty category must therefore be assumed.  
(1989: p. 163, fn. 8)

The HSPM’s prediction also cannot be accounted for by the ambiguity resolution principles discussed above—Minimal Attachment and Late Closure. The reason is that there is no structural ambiguity to resolve in these cases. Whether the position after ‘forced’ in sentences (88) and (89) is filled by an overt lexical item (e.g., ‘us’) or by a *wh*-trace, the structural relations in the sentence remain the same.

To account for the filled-gap effect, deVincenzi (1991) proposed a third processing principle, known as the Minimal Chain Principle (MCP). According to the MCP, the parser will avoid postulating unnecessary gaps/traces and, having identified an item in the input as an

antecedent, will posit a gap/trace in the very first position at which it is licensed by the grammar.<sup>32</sup>

For present purposes, what is important about the MCP is that it makes ineliminable reference to abstract grammatical notions—e.g., ‘position at which a trace is licensed’. This reference to grammatical licensing is on a par with the kind we saw earlier, in connection with Minimal Attachment and Late Closure. Hence, we can draw the same conclusions: Even if the processing principles are not themselves represented by the HSPM, but are merely *true of* its operations, the fact that such principles must be formulated in terms of the proprietary notions of a grammar entails that the HSPM employs a grammar, by either embodying or representing it.<sup>33</sup>

To substantiate this line of reasoning, we must show the HSPM is, indeed, linguistically informed. Fodor (1989) considers just this question and discusses a number of experiments that were designed to settle it. She begins by summarizing the studies that have yielded hints of evidence for the claim that the HSPM draws on information about the subcategorization properties of verbs—i.e., what complements a verb allows or requires. Although the experiments she discusses provide relatively weak evidence, more recent work has demonstrated that the HSPM does indeed make use of this information. For instance, Gorrell (1991) used a

---

<sup>32</sup> The second conjunct is equivalent to what Frazier and Clifton (1989) refer to as the “Active Filler Hypothesis.” I use the locution ‘gap/trace’ to avoid commitment to grammatical formalisms that have traces and movement operations in their theoretical toolkit.

<sup>33</sup> Stabler (1983: pp. 396-7) considers an argument due to Fodor, Bever, and Garrett (1974: pp. 369-370) that looks, at first blush, to be quite similar to the argument presented here. He then gives what are, to my mind, quite decisive reasons to reject that argument. If the appearance of similarity between my argument and theirs were veridical, then Stabler’s rejoinders to their argument would apply to mine as well. But the appearance is wholly illusory; Fodor, Bever, and Garrett were concerned with a quite different matter than the one at issue here. Their argument was intended to show that formulating linguistic universals—i.e., formal constraints on the rules and principles of a grammar—commits one to the claim that these rules or principles are not only represented by the HSPM, but that they are represented *in a particular formalism*. My argument, by contrast, has nothing at all to do with linguistic universals and carries no implications for the formalism in which a psychologically real grammar might be represented. Moreover, my argument appeals directly to empirically verified *processing* principles, rather than to abstract considerations about the adequacy conditions on descriptive grammars.

reaction time measure on a lexical decision task to show that the HSPM makes a distinction between transitive and intransitive verbs within milliseconds of lexical retrieval—i.e., without any appreciable delay. French-Mestre and Pynte (1995) obtained evidence for the same conclusion by conducting an eye tracking study. Their findings are even more convincing than Gorrell's, given the inherently more fine-grained nature of eye tracking data.

Stowe (1986) conducted an early experiment that bears on the question of whether the HSPM is sensitive to so-called *island constraints*—i.e., restrictions on the structural configurations that allow for fillers to be bind their corresponding gaps. The question of how best to characterize these configurations in a descriptive grammar is a difficult one; the principles appear to be quite abstract. Demonstrating that the HSPM employs such principles in its computations would lend strong support to the claim that they are psychologically real. To address this issue, Stowe (1986) examined sentences like (90).

(90) The teacher asked what<sub>i</sub> [the silly story about Greg's older brother] was supposed to mean *wh*-trace<sub>i</sub>.

In the vocabulary of the syntactician, the bracketed noun phrase in (90) is an island. What this means is that the *wh*-trace associated with the word 'what' cannot appear within that phrase. Nevertheless, there is a position within the island at which a processor ignorant of the island constraints would be tempted to posit a *wh*-trace. That position is right after the word 'about', which requires a noun phrase complement.

The logic behind Stowe's experiment is simple: If the HSPM knows about island constraints, then it will not predict a *wh*-trace after 'about'; if it doesn't know about them, then it might. Having posited the *wh*-trace after 'about', a linguistically ignorant processor would exhibit the filled-gap effect discussed above. Specifically, reading times would be elevated at the word 'Greg'. Stowe observed no significant rise in reading times at that position, as

compared to the analogous position in the control sentence (91).

(91) The teacher asked if [the silly story about Greg's older brother] was supposed to mean anything.

This demonstrates the *absence* of a filled-gap effect. Stowe took this as confirming the hypothesis that the HSPM has access to a grammar that covers island constraints.

Fodor (1989: p. 169) notes that the data in Stowe's experiment are rather weak.

Moreover, at that time, what appeared to be conflicting evidence was emerging. In light of this, Fodor reasonably suspended judgment about whether the HSPM is sensitive to island constraints. However, further experimental work has largely settled the matter. The following passage from Phillips and Lewis (forthcoming) summarizes the current state of play and provides a number of references to key studies that have been conducted in this area:

[One] body of on-line studies has examined whether on-line structure building respects various grammatical constraints, i.e., whether the parser ever creates grammatically illicit structures or interpretations. Many studies have found evidence of immediate on-line effects of grammatical constraints, such as locality constraints on *wh*-movement (Stowe, 1986; Traxler & Pickering, 1996; Wagers & Phillips, 2009), and structural constraints on forwards and backwards anaphora (Kazanina et al., 2007; Nicol & Swinney, 1989; Sturt, 2003; Xiang, Dillon, & Phillips, 2009). These findings extend to complex cases that present apparent challenges for incremental application of grammatical constraints, such as constraints on backwards anaphora in Japanese, where the constraints must apply before any verb has appeared in the input (Aoshima, Yoshida, & Phillips, 2009), and constraints on parasitic gaps inside complex subjects in English, where the parasitic gap precedes its licenser (Phillips, 2006). *Findings such as these imply that the structures created on-line include sufficient structural detail to allow the constraints to be applied during parsing.* (Phillips and Lewis, forthcoming; emphasis added)

It is worth examining one of the studies that Phillips and Lewis mention in this connection—viz., Nicol and Swinney (1989). This study initiated a line of research that makes use of an experimental paradigm known as *Cross-Modal Priming*. To see how this works, consider an example from Nicol and Swinney's experiment. Sentence (92) contains a relative clause with a *wh*-trace in the object position of the verb 'accused'.



(92) The policeman saw the boy<sub>i</sub> [that<sub>i</sub> the crowd at the party accused *wh*-trace<sub>i</sub> of the crime].

In (92), there is only one position at which the *wh*-trace is grammatically licensed and only one noun phrase that can legitimately serve as the antecedent of that *wh*-trace—viz., ‘the boy’.

However, the sentence contains a number of other noun phrases—‘the policeman’, ‘the crowd’, and ‘the party’—any of which a linguistically ignorant processor might take to be the antecedent.

Let’s call these *distracters*.

To test whether the HSPM is temporarily fooled into taking any of the distracters as the antecedent of the *wh*-trace, Nicol and Swinney had participants listen to sentence (92) while fixating on a computer screen. When the word ‘accused’ was spoken, participants saw a word appear on the screen. In some trials the word was semantically related to ‘boy’, which is the antecedent of the *wh*-trace that appears after ‘accused’. For example, participants saw the word ‘girl’. In other trials, participants saw words that were comparable in length to ‘girl’, but semantically related to one of the distracters. They were asked to read the word out loud and their reaction times were measured.

Nicol and Swinney made the following assumption: If the HSPM posits a *wh*-trace after ‘accused’, then it will activate the meaning of the antecedent ‘boy’ at that point, which would in turn prime the recognition of semantically related words, like ‘girl’.<sup>34</sup> By contrast, the recognition of words that are semantically related to one of the distracters would *not* be primed. As expected, this is what they found. The results were quite robust and have been replicated a

---

<sup>34</sup> Priming studies had, by this time, demonstrated quite clearly that the recognition of a word primes the recognition of semantically related words.

number of times in subsequent studies.<sup>35</sup> On the basis of these studies, *wh*-traces had, for some time, been taken to be psychologically real.

As noted above, transformational grammars posit a number of empty categories in addition to *wh*-trace. The question arises, then, whether these, too, have psychological reality—i.e., whether they are explicitly represented in the mental phrase markers constructed by the HSPM. Evidence for a positive answer to this question emerged from a number of studies that employed an experimental paradigm similar to Cross-Modal Priming, known as *Visual Probe Recognition*. In this paradigm, participants are presented with sentences, one of which is hypothesized to have an empty category (such as PRO or NP-trace) in its underlying structure. In the studies discussed below, the relevant empty category was almost always an NP-trace.<sup>36</sup>

The Visual Probe Recognition experiments rest on assumptions similar to those inherent in the Cross-Modal Priming paradigm. The presence of an NP-trace is hypothesized to reactivate the meaning of the antecedent, thus priming for faster or more accurate responses. The differences between the two paradigms consist in their dependent variables, as well as in the nature of the tasks they rely on. Whereas Nicol and Swinney (1989) measured response times on reading tasks that were conducted mid-sentence, MacDonald (1989) and McElree and Bever (1989), using the Visual Probe Recognition paradigm, asked participants to wait until the sentences had been presented and then to judge whether a particular word was present in the

---

<sup>35</sup> In connection with the earlier discussion of the modularity-vs.-interaction debate, it is worth pointing out that other aspects of the data from Nicol and Swinney's experiment provide further evidence for the "modular" or "syntax-first" view. In reporting their findings, they wrote: "When structural information cannot serve to constrain antecedent selection, then pragmatic information may play a role, *but only at a later point in processing*" (p. 5, emphasis added).

<sup>36</sup> Recall that NP-traces are posited by the Government and Binding theory to account for the relation between a verb and its object in passive sentences like (74) and raising constructions like (79), both repeated here:

- (74) [NP The boat]<sub>i</sub> was carried NP-trace<sub>i</sub> by five people.  
(79) Igor<sub>i</sub> seems NP-trace<sub>i</sub> to enjoy action movies.

input. On many trials, the word was not in fact present, so participants would answer ‘No’. But in cases like those shown in (93) and (94), the answer was ‘Yes’. The dependent variable was the speed of participants’ responses.

(93) The terrorists wanted to disrupt the ceremonies. [The new mayor at the center podium]<sub>i</sub> was shot NP-trace<sub>i</sub>.

(94) The terrorists wanted to disrupt the ceremonies. [The new mayor at the center podium] was furious.

Both MacDonald (1989) and McElree and Bever (1989) found that the presence of an NP-trace primed the recall of the word ‘mayor’. For instance, participants’ responses were faster when primed with (93) than with (94). On the strength of these results, NP-traces were tentatively declared to be psychologically real, on a par with *wh*-traces.

The emerging picture was, however, complicated by the fact that *wh*-traces elicited different effects in the Cross-Modal Priming experiments than did NP-traces and PRO. For instance, in a study that used an improved version of the paradigm, Osterhout and Swinney (1992) found that NP-traces primed for the recognition of their antecedents, but only when they appeared at increasingly longer distances from their antecedents. In the limit, it took a whopping 1000ms for the effect to show up. This contrasts sharply with *wh*-traces, which have priming effects almost immediately. Osterhout and Swinney proposed two alternative explanations for this difference.

According to the first account, the difference has to do with the fact that the antecedents of *wh*-traces always appear in nonargument positions and provide the HSPM with an obvious cue that a gap will appear later in the sentence. The antecedents of NP-traces, by contrast, appear in argument positions, so the HSPM has no warning that the input will contain a gap. According to the second account, the difference is that *wh*-traces are inserted into the mental phrase marker during syntactic processing, whereas the effects associated with passive, raising, and control

verbs are due to subsequent *semantic* processing. On this account, NP-traces, unlike *wh*-traces, are not psychologically real—i.e., not present in the HSPM's mental phrase markers.

Although Osterhout and Swinney did not offer any grounds for privileging one of these accounts over the other, they made clear the implications of adopting the second explanation. As noted earlier, a variety of syntactic theories—most notably Lexical Functional Grammar and Head-driven Phrase Structure Grammar—dispense with the theoretical construct of NP-traces and PRO, preferring to deal with filler-gap binding in passive, raising, and control constructions by alternative means. Hence, if it could be shown that NP-traces and PRO are not psychologically real, this would constitute evidence that the grammar employed by the HSPM is more like these nontransformational grammars than like Government and Binding formalism. For theorists who hold out hope that a single grammar will be both descriptively adequate and psychologically real, this conclusion would carry great significance.

In subsequent discussions, a number of theorists went farther than Osterhout and Swinney, questioning the validity of the studies that were supposed to establish the psychological reality of *wh*-traces. In particular, proponents of Head-driven Phrase Structure Grammar were beginning to claim that their formalism requires *neither* NP-traces *nor* *wh*-traces. In addition, it became apparent that *all* of the experimental data discussed above can be accounted for by the assumption that the effects are due to semantic rather than syntactic processing. A particularly forceful statement of this view can be found in Sag and Fodor (1994). Questioning the psychological reality of *wh*-traces, they write:

What would it take to show, on the basis of sentence processing, either that there is, or that there is not, a syntactic entity in an extraction gap? Fodor (1993) noted that no psycholinguistic finding could even in principle qualify as evidence for empty constituents unless it were established that the data pertained to the syntactic processing or representation of a sentence. But for all we know at present, the experimental techniques that have been used to study gap processing to date might

be providing information about semantic processing only. If so, they would be completely uninformative about traces, because they wouldn't tell us whether the sentence meaning is computed from a trace or from a traceless gap in the syntax. (Sag and Fodor, 1994: p. 8)

We see here yet another instance of an argument pattern that has emerged several times in the history of the cognitivist study of language. Recall that the earliest experiments purporting to demonstrate the existence of mental phrase markers were criticized primarily on account of their failure to control for semantic factors. The structural priming results examined in §3 teased apart syntax and semantics, in the face of deep skepticism concerning a distinctively syntactic level of representation. And while careful and extensive experimentation sufficed, in large part, to quell those concerns in connection with structural priming studies, the drive to deemphasize syntactic explanations in favor of semantic ones cropped up again when we turned our attention to the literature on ambiguity resolution. Our tentative conclusion there was that models which explain initial parsing preferences by appeal to distinctively syntactic properties of the input fare better with respect to a range of experimental findings than rival accounts which stress the influence of semantics. Still, the controversy surrounding these matters has not yet been resolved. Accordingly, our conclusions about the primacy of “syntax-first” models must be tentative. Moreover, we see now that the broader debate regarding the adequacy of explanations pitched at a distinctively syntactic level of processing recurs in discussions of the psychological reality of empty categories. Theorists who seek to eliminate empty categories from both descriptive grammar and from psychological models of sentence processing manifest, once more, the tendency to explain experimental findings by appeal to a level of semantic representation.

Although a decisive resolution of this issue is currently out of reach, recent experimental research does shore up a range of findings that bear on it. Before closing this section, then, let us

review the results of a series of studies reported in Featherston (2001). As we shall see, Featherston's conclusions favor a view that locates filler-gap effects at a *syntactic* level of representation.

Featherston (2001) begins by noting that there are currently *four* separate accounts of the data from the Cross-Modal Priming and Visual Probe Recognition studies discussed above.<sup>37</sup> Two of these we have already examined. The first is the Trace Reactivation account, which takes at face value the claim that the HSPM posits traces at various positions in the input. On this account, the priming data show that traces are syntactic entities whose presence in the mental phrase marker facilitates the reactivation of the meanings of their antecedents—i.e., that traces are, in this respect, on a par with overt anaphoric elements like pronouns. The second is the Semantic Processing account, which, as the name suggests, construes effects at gap position as being indicative of activity at a phase of processing distinct from the phase at which syntactic structure is assigned—viz., the semantic phase.

The third is what Featherston calls the Depth of Processing account, discussed in Fodor (1991). In reviewing the Visual Probe Recognition (VPR) data presented above, Fodor writes:

[T]here is known to be an inverse relation between VPR performance on a word, and its expectability and ease of processing in the sentence. Earlier research (not all of it on language) had shown that the greater the effort that goes into processing a stimulus, the more distinctive its subsequent memory representation is. ... Let us call this the depth-of-processing effect. It links three things: the less predictable an item is, the greater the effort of processing it, and the easier it is to recall. This provides a good explanation for why passive sentences fare better than adjectival sentences in VPR experiments. Because passive sentences are less common than adjectival sentences, they receive more attention during processing, so they are easier to probe in short-term memory. (pp. 238-9)

---

<sup>37</sup> Featherston discusses many experiments that were performed between the late 1980s and his own 2001 studies. For the sake of brevity, I have not mentioned these here. Their implications were taken into account in the design of Featherston's experiments.

To illustrate, let's apply this to examples in (93) and (94), repeated here. The data showed that participants were faster at recalling the presence of the word 'mayor' in (93) than in (94).

(93) The terrorists wanted to disrupt the ceremonies. [The new mayor at the center podium]<sub>i</sub> was shot NP-trace<sub>i</sub>.

(94) The terrorists wanted to disrupt the ceremonies. [The new mayor at the center podium] was furious.

The Depth of Processing account entails that the facilitation of participants' recall had nothing to do with the presence of an NP-trace in the mental phrase marker for the second sentence in (93), but instead has to do with the fact that this sentence is a passive.

The fourth account Featherston considers, which he calls the Direct Association hypothesis, is due to Pickering and Barry (1991). According to it, "the effects found are in fact due to the association of the argument with the verb, and that the confusion of this with trace reactivation is due to the position, adjacent to the verb, at which direct object traces are posited" (Featherston, 2001: p. 221). It follows from this view that if it were possible to dissociate the location of a trace from the position adjacent to a verb, then priming effects would be observed only at the latter position, not at the location of the trace. As Featherston notes, this dissociation is all but impossible to achieve in English. However, materials from a language like German, which differs in the crucial respects from English, can be used to tease apart the predictions of the Trace Reactivation account and the Direct Association account.

Featherston (2001) reports the results of six separate experiments. I will mention here only the findings from experiments 2-5.<sup>38</sup> Experiments 2 and 3 made use of the Cross-Modal Priming paradigm, applied to German sentences. With these materials, antecedent reactivation effects are predicted *only* by the account that takes traces to be syntactic entities on a par with

---

<sup>38</sup> Experiment 1 was argued to be flawed. Experiment 6 made use of an experimental paradigm known as *sentence matching*, which we have not introduced here and whose validity has not yet been subjected to sufficient scrutiny.

overt pronouns, whose presence in a mental phrase marker facilitates the reactivation of the meanings of their antecedents. In discussing the results of the experiments, Featherston writes:

The data clearly suggest that the mechanism of trace reactivation does indeed account for some of the experimental data, since no other account predicts the range of effects observed. This result could lead us to pay greater attention to grammatical and processing models which make use of trace, since their empirical adequacy receives support. (p. 223) ... Experiments 2 and 3 are strong results supporting the psychological reality of movement trace since these results are not predicted by any traceless account... (p. 229)

Experiment 4 made use of Visual Probe Recognition paradigm discussed above. The results of the experiment provided no confirmation of any of the four competing accounts. No depth-of-processing effects were observed and no evidence of *wh*-traces in so-called tough-movement constructions was found.<sup>39</sup> Featherston reasons that the Visual Probe Recognition paradigm is simply “blind” to *wh*-traces and concludes, with Fodor (1991), that it may well not be a suitable paradigm for the study of empty categories in sentence processing.

Experiment 5 differed from the rest in that the data consisted of recordings of event-related potentials (ERPs), discussed in §2 above. In addition, it tested specifically for the psychological reality of NP-trace and PRO. The materials were designed to contrast raising, control, and ordinary transitive constructions in German—the latter serving as a baseline measure. Participants were asked to read sentences that, according to Government and Binding theory, contain NP-traces, PRO, or neither.

The results showed significant differences in the ERP signatures of all three constructions. This provides evidence for the claim that the mental phrase marker associated

---

<sup>39</sup> For reasons of space, we have not discussed tough-movement constructions in this chapter. An example of such a construction is ‘Igor is tough to convince’, where the adjective ‘tough’ seems to raise the object of the verb ‘convince’ to the subject position in the higher clause. There are numerous syntactic analyses of this construction and there is, as yet, no convergence amongst syntacticians as to what its underlying structure is. Here, we simply note that Government and Binding theory treats it as containing multiple empty categories, including *wh*-trace.



with each construction differs from either of the other two. As the sentences were matched for length and other confounding factors, the natural interpretation of the data is that the ERP signatures track the presence of NP-trace and PRO. Featherston draws this conclusion in the following passages:

The results showed significant differences between all three, but particularly so between raising and equi,<sup>40</sup> where the difference was plain in three successive 200ms time windows. This result is not predicted by the [Head-driven Phrase Structure Grammar] analysis of the difference between raising and equi, since it locates the contrast in the semantic role assignment of the matrix verb. In this way the principled distinction between raising and equi implemented in [the Government and Binding theory and Minimalist grammars] by the use of different ECs is supported. (p. 228)

The pool of evidence for local ECs has lost the visual probe recognition data but gained a new recruit in the form of our ERP result. Since the former is the methodologically least sophisticated and the latter the most sophisticated, we must conclude that the case for local ECs has been strengthened by our studies. (p. 229)

Our ERP Experiment 5 ... offers robust support for the existence of local ECs, since the alternative possible causes of the P600 we found are excluded by our two follow-up experiments. In sum we have a significant body of data suggesting that ECs play a role in on-line processing as well as in the grammar. Since this hypothesis is a strong and testable claim, we feel that any evidence in its favour is a surprising fact and should be treated as a success for linguistics. (pp. 229-230)

### *Conclusion*

It goes without saying that all of the findings reviewed above can be challenged in a number of ways and may very well fail to hold up to further scrutiny. Nevertheless, at present they offer us grounds for the claim that empty categories—including *wh*-trace, NP-trace, and PRO—are psychologically real elements of mental phrase markers. For reasons such as these, Phillips and Lewis (forthcoming) are quite correct to assert that “there is now much experimental

---

<sup>40</sup> To avoid the ambiguity in the term ‘control’, Featherston refers to syntactic-control verbs as equi-verbs, reserving the term ‘control’ for the baseline measure in an experiment.

evidence [supporting] the notion that real-time processes assemble syntactic representations that are the same as those motivated by grammatical analysis.”

Whatever the case about that, one thing is clear: the HSPM is *not* a naïve mechanism. Regardless of the details concerning the structures it builds, the evidence reviewed above establishes beyond reasonable doubt that, in building them, the HSPM has access to sophisticated knowledge of the dependency relations that hold between grammatical antecedents and the syntactic positions to which they are bound. As Fodor (1989) observes,

long-distance binding of traces should provide many pitfalls for a rough-and-ready processor which relies on informal strategies rather than consulting the information provided by the grammar (except perhaps as an emergency back-up). The hypothesis that the [HSPM] is such a device (Fodor, Bever, and Garrett, 1974; Bever et al., in press) becomes quite implausible in the face of the speed and accuracy with which the [HSPM] interprets traces. We can conclude, instead, that the [HSPM] is very closely attuned to the grammar of the language. If that is so, then differences in how the processor responds to different (putative) empty categories can be taken seriously as evidence of how they are treated by the grammar. (Fodor, 1989: p. 205)

The final remark in this passage is particularly significant for the linguist who seeks to formulate a grammar that is not only descriptively adequate but also psychologically real.

In previous chapters, I argued that descriptive adequacy is a worthy goal, in and of itself, for a syntactician to pursue in constructing a grammar. And, as Devitt (2006) argues at length, it seems to be the *only* goal that many syntacticians pursue. Nevertheless, it's plain that finding a descriptive *and* psychologically real grammar is a much more exciting prospect. In this chapter, I have laid the groundwork for a defense of the claim that some grammar is, indeed, psychologically real. Our discussion of the Minimal Attachment, Late Closure, and the Minimal Chain Principle suggests that psychologically plausible parsing models will have to, in one way or another, make use of grammatical rules or principles in constructing mental phrase markers.

In the next chapter, we will examine two models that attempt to get by without drawing without constructing mental phrase markers. I will argue that such models face serious difficulties. In chapter V, I will survey a range of computational models that overcome these difficulties, in large part by constructing mental phrase markers and making use of either an embodied or an internally represented grammar.

## IV

### ON TWO ATTEMPTS TO DO WITHOUT MENTAL PHRASE MARKERS

---

#### §1. Introduction

In the previous chapter, we saw a number of arguments for the psychological reality of mental phrase markers—i.e., explicit representations of the distinctly *syntactic* structure of linguistic input. Neurocognitive data, structural priming experiments, and studies of garden-path and filler-gap processing all point to the same conclusion: Any psychologically plausible model of human language comprehension must, at some stage, construct mental phrase markers. My aim in this chapter is to reinforce this conclusion, by casting doubt on two prominent proposals for doing without mental phrase markers.

The first of these is due to Roger Schank and his colleagues at Yale, who throughout the 1970s and 80s, constructed models that relied almost exclusively on semantic and pragmatic resources in the course of comprehension. The second is the “brute-causal” model tentatively endorsed by Devitt (2006). According to the strongest version of this model, the human sentence processing mechanism (HSPM) maps phonetic representations *directly* into thoughts, which are assumed to have a syntactic structure that closely resembles that of public-language sentences. I will demonstrate the weaknesses of such models and argue that they are unworkable in light of the evidence we currently have regarding human language comprehension.

I begin by summarizing the motivations that Schank and Birnbaum (1984) give for pursuing what they call the “integrated processing hypothesis,” according to which “meaning and world knowledge are often crucial even at the earliest points in the process of understanding language” (212). We saw in the previous chapter (§4) that this hypothesis is at variance with a

range of experimental results, which show that the HSPM makes use of *solely* syntactic resources and represents *exclusively* syntactic properties at the earliest stages of comprehension. I briefly review these results in §2.1, and then draw on them in defusing a number of arguments that Schank and Birnbaum (1984) put forward in favor of the integrated processing hypothesis. I then turn to their stronger claim, “that no independent level of syntactic representation is constructed, operated on, or output by the language analysis process” (220).

The models that Schank and his associates developed were rich in semantic and pragmatic resources—representations that encoded not only the argument structures of various predicates, but also a wealth of encyclopedic knowledge. Relying solely on these resources, Schank *et al.* were able to achieve rather impressive results, which I review in §2.2.

Unfortunately, their achievements were quite limited, primarily due to the fact that the integrated processing models made reference only to the linear order of the words in the input string. No representation of the string’s hierarchical syntactic structure was used in determining which phrases play which thematic roles. Drawing on Marcus (1984), I show in §2.3 that such models cannot handle a wide variety of sentences. On account of their impoverished resources for representing syntactic structure, they fail to assign correct interpretations to sentences with embedded clauses, as well as sentences with “gaps,” like those we encountered in ch. III (§5).<sup>1</sup> Plainly, the inability of these systems to cope with such ubiquitous phenomena disqualifies them as plausible models of human sentence processing.

I turn next to Devitt’s brute-causal model. Devitt (2006) tentatively endorses the infamous Language of Thought Hypothesis (LOTH), due originally to J. A. Fodor (1975).

---

<sup>1</sup> To a first approximation, gaps are syntactic positions at which the semantic structure of a predicate requires there to be an argument, but where no material is overtly present.

LOTH has it that mental representations are couched in a linguistic medium, which Fodor—following Wilfrid Sellars—calls “Mentalese.” The central idea of LOTH is that the neurochemical vehicles of thought are *syntactically structured*. A question arises, then, about the relationship between a person’s mental representations—her Mentalese sentences, as it were—and her public-language representations, speech acts or sentence tokens: Does a person’s Mentalese syntax bear a close resemblance to the syntax of her public language? Devitt argues that if we accept LOTH, then we should also take on board the view that this resemblance is relatively strong. This yields what I call the *Strengthened* Language of Thought Hypothesis.

(SLOTH) A person’s thoughts and other mental representations are couched in a *linguistic medium* whose syntax bears a close resemblance to his or her public language.

Devitt argues that accepting SLOTH would bolster his brute-causal model of language processing, whose central claim is this:

(NO-MPM) The HSPM engages in *brute-causal processing*; it does *not* construct or operate upon mental phrase markers, but instead maps phonetic representations *directly* into thoughts.

He reasons that if public language expressions have syntactic properties that are largely mirrored by the corresponding Mentalese sentences, then language comprehension should be a relatively straightforward process. The HSPM would be charged with the task of directly activating those thoughts that match incoming linguistic expressions in respect of syntax and semantics. Devitt conjectures that all the work can be accomplished by what he calls a “brute-causal process”—a series of associationist transitions. This view, which combines *both* SLOTH *and* brute-causal processing, is the strongest version of Devitt’s position. The resulting model of language comprehension, which I will call the Strong Brute-Causal (SBC) model, will be my primary target in §3.

I §3.1, I discuss the standard arguments for LOTH, which appeal to the alleged systematicity, productivity, and inferential coherence of thought. After raising some concerns about the cogency of these arguments, I turn to Devitt's argument for SLOTH. The argument takes the form of an inference to the best explanation. According to Devitt, a strong resemblance between the syntactic structures of thoughts and public-language expressions explains *why* public languages have the syntax that they do. I offer three objections to this argument, and then examine a recent attack on its conclusion, due to Pinker (1994).

First, I point out that Devitt's argument relies on a contentious thesis to the effect that thought is *explanatorily* prior to language. Other priority claims—e.g., those having to do with ontogeny, phylogeny, and causation—are, of course, uncontroversial. But the *explanatory* priority of thought over speech faces serious challenges (Sellars, 1958; Rosenthal, 1986/2005). Second, I show that there is a satisfying *alternative* explanation of the syntactic properties of public-language expressions—indeed, one that Devitt himself seems to endorse at several points in *Ignorance of Language*. This explanation stems from a perspective that accords no explanatory priority to either thought or language. Third, I point out that Devitt's explanation of the syntactic properties of public language expressions rests on an assumption that is undermined by other commitments of the brute-causal model. I demonstrate this internal tension by reference to a connectionist model of sentence processing due to Rohde (1999). Finally, I examine two arguments against SLOTH, due to Pinker (1994). Although Devitt succeeds in rebutting Pinker's first argument, he does not discuss the second. I offer reasons for thinking that Devitt's best option for responding to Pinker involves an unpalatable commitment to the existence of synonymous-but-nevertheless-distinct thoughts.

Having cast doubt on the viability of SLOTH, I go on to present Devitt's characterization of brute-causal processing, and to assess whether the operations of the HSPM fit that characterization. I argue that they do not. Language comprehension, even at the earliest stages, is neither reflex-like (in the sense that Devitt intends) nor associative. More importantly, we have good reasons to believe that the comprehension process constructs and operates on structured metalinguistic representations—mental phrase markers. I will discuss Devitt's reasons for denying this claim, specifically his reliance on a distinction between *representing* something as having a certain property and *responding* to it as having that property. I conclude that the operations of the HSPM cannot be viewed as mere responses, in his sense.

## **§2. The Integrated Processing Model of Comprehension**

In their landmark paper, “Memory, Meaning, and Syntax” (1984), Roger Schank and Lawrence Birnbaum report the results of a nearly two-decade-long effort on the part of linguists and computer scientists at Yale to implement versatile and realistic natural language processing (NLP) systems.<sup>2</sup> The models that emerged from this effort were certainly impressive and continue to this day to inform work in NLP and in AI more generally. However, these models were also limited in certain key respects. In particular, they were incapable of processing a wide range of constructions, including many types of relative clause, *wh*-question, and anaphora. Marcus (1984) argues persuasively that these limitations can be traced back to a single design decision that Schank and his colleagues adopted throughout their work—viz., to largely eschew explicit representations of the syntactic structure of linguistic input.

---

<sup>2</sup> Schank, R. *et al.* (1970, 1972, 1975a, 1975b, 1977, 1978, 1979a, 1979b, 1980, 1981, 1984)



We saw in ch. III that a wide range of psycholinguistic evidence points to the need for such representations. Needless to say, if the “mostly-semantic” models developed by the Yale school were satisfactory on independent grounds, there would be grounds for reevaluating this evidence. For we would then have a working example of how natural language can be processed *without* mental phrase markers. As it happens, however, the exact opposite is the case. The limitations of Schank *et al.*’s models are both striking and pervasive—so much so that they were abandoned by the NLP community in favor of models that do not share their fatal design flaw.

In §2.1, I discuss the motivations for the integrated processing hypothesis and find them groundless. In §2.2, I lay out the theoretical commitments that Schank and his associates adopted and review the results that they achieved. In §2.3, turn to the criticisms that Marcus (1984) levels against their approach, focusing on a handful of cases that illustrate its inadequacy. I take these criticisms to be decisive, and hence to constitute further evidence for the claim that representing syntactic properties is crucial to the success of language comprehension, whether in artificial NLP systems or in humans.

### *§2.1 Assessing the Motivations for “Integrated Processing” Models*

Schank and his associates developed their comprehension models in accordance with a philosophical and methodological outlook that they take to be in direct opposition to the dominant Chomskyan paradigm. They are skeptical of the import of the competence/performance distinction (ch. I), and they do not see the point of constructing an autonomous syntactic theory. Indeed, they claim that “it is unlikely that a purely linguistic theory could be *in any sense* adequate” (p. 209, emphasis added). On their view, because language is inextricably bound up with thought, we should be suspicious of the claim that a purely syntactic analysis of a stimulus is necessary prior to the assignment of meaning.

Moreover, they claim that “pure linguistic theory” can only be concerned with “questions of form,” and they infer from this that such a theory cannot tell us about linguistic meaning or about how people process language. “Can one form an adequate theory of language-processing behavior without first constructing a purely linguistic theory? Of course!” (p. 210)

On the basis of these commitments, Schank *et al.* make the following proposal:

Our working hypothesis is that meaning and world knowledge are often crucial even at the earliest points in the process of understanding language. We call this position *the integrated processing hypothesis*. ... [T]his hypothesis stands in direct opposition to theories that posit a logically autonomous syntactic analysis procedure, preceding, and providing input for, semantic processing. (pp. 212-13).

Two questions immediately arise: First, what reason do we have for adopting the integrated processing hypothesis (henceforth, IPH) or, indeed, the background theoretical commitments from which it arises? Second, how do Schank *et al.* propose to implement the IPH in a comprehension model? I address the first of these questions in the remainder of this section, and take up second in §2.2.

The IPH is motivated in several ways. First, Schank *et al.* note that the early models of machine translation, which made no use of semantic information at all, failed to deal with the massive ambiguity of natural language. This, however, does not support the IPH, but only the weaker and wholly uncontroversial claim that effective comprehension systems must draw on semantic and pragmatic resources at *some* point. The IPH is a stronger claim, to the effect that such resources must come into play “even at the earliest points in the process” (*ibid.*)

Similar remarks apply to Schank *et al.*’s second argument. They point out that it is typically much easier for people to understand text in a foreign language with which they have some familiarity, than to produce or to paraphrase such text. “It would seem, then, that

understanding language requires far less knowledge of syntax than generation...” (p. 213).<sup>3</sup>

While their observation is interesting and certainly merits further investigation, it does not support the IPH.

First, familiarity with foreign languages is not exclusively a matter of knowing the meanings of foreign lexical items. In picking up fragments of a foreign language, one also acquires rudimentary syntactic knowledge. Of course, such knowledge may very well be influenced by their knowledge of the syntax of their first or primary language—a well-studied phenomenon known as *L1 transfer*. In those cases where the primary language differs significantly from the foreign language, L1 transfer will cause one to have an incorrect conception of the syntax foreign language.<sup>4</sup> But it does not follow that one has *no* syntactic resources at one’s disposal, nor that one fails to bring those resources to bear on the comprehension task. It may well be that one comprehends expression in a familiar foreign language by accessing familiar lexical items and ascertaining (however crudely) their syntactic properties. Once accessed, those properties can be used to assign an exclusively syntactic structure to the input, though possibly a very rudimentary one. Such a syntactic analysis may, in turn, serve as the input to the “semantic analyzer,” contrary to the IPH. Moreover, as we shall see in §2.3, to the extent that comprehension of the foreign text does *not* proceed this way, it will be reliable only for the case of simple, clause-sized expressions.

---

<sup>3</sup> This is contrary to Devitt’s claim that syntactic knowledge is less operative in production than in comprehension. As we shall see in §3.3, his claim is made against the background of a strong version of the language of thought hypothesis. It is likely that Schank *et al.* would not endorse this position. Recall, moreover, the syntactic priming evidence that we reviewed in ch. III (§3), which strongly suggests that mental phrase markers play a role in language production. (See esp. Bock, *et al.* 1989, 1990, 1992.)

<sup>4</sup> Note that my use of the terms ‘knowledge’ and ‘conception’ is informal throughout. I argue in ch. V that the relevant states are *subpersonal* and do *not* require the possession of concepts (in the sense of the term that has common currency in the philosophy of mind).

Schank *et al.*'s third argument for the IPH rests on the following observation: In those cases where our interlocutor stops speaking in mid-sentence, introspection seems to reveal that we have already interpreted what has been said and that we've generated a rich body of expectations about how the utterance will continue. For instance, upon hearing 'Seven Libyan gunmen shot their way into...', we do not expect 'a paper bag' to follow, and we would not notice the ambiguity in 'a bank' if it were the completion of the sentence. Schank *et al.* draw the following conclusion: "Clearly, we do not wait for a complete syntactic analysis before we begin to process for meaning" (210). To evaluate this argument, it pays to be clear about what the opponent of the IPH is committed to. Crucially, she is *not* committed to the claim that the HSPM constructs a global syntactic analysis of an *entire sentence*, or even a whole clause, before initiating semantic processing. While such claims have, indeed, been made,<sup>5</sup> they are far too strong and now known to be at variance with experimental data.

All parties to the debate concerning this issue are agreed that a wide variety of different types of information are brought to bear on the task of comprehension *at some point or other*.<sup>6</sup> What is controversial is whether the HSPM brings all of these to bear *at the very earliest stages of the process*, when the initial decisions regarding attachment and dependency are being made. The disagreement, then, is between theorists who favor a "syntax-first" parsing model, on which only *syntactic* information influences the earliest parsing decisions—with other information

---

<sup>5</sup> Fodor, Bever, and Garrett (1974).

<sup>6</sup> Recall the following remarks from Traxler and Pickering (1996): "Researchers in the field of sentence processing have identified many potentially relevant sources of information [that the HSPM might make use of]. They do not, however, necessarily agree about which information is used during initial processing. These information sources include lexical preferences of individual verbs (Ford, Bresnan, & Kaplan, 1982), plausibility (Rayner, Carlson, & Frazier, 1983), discourse context (Crain & Steedman, 1985), punctuation (Mitchell & Holmes, 1985), prosody (Marslen-Wilson, Tyler, Warren, Grenier, & Lee, 1992), gender agreement (Brysbaert & Mitchell, 1996), and case-marking (Trueswell et al., 1993)" (p. 993).

types exerting an influence only after a short delay—and those who favor an “interactionist” model, on which *all* types of information exert an influence from the very outset of the process.

The evidence reviewed in ch. III (§4) strongly favors the syntax-first hypothesis. We saw there that, even when prior semantic context would clearly disambiguate between two grammatically licensed attachments, the HSPM does *not* make use of this information in its initial parsing decisions. Recall that Ferreira and Clifton (1986) used materials like (1) and (2) in an eye tracking study aimed at discerning the HSPM’s attachment preferences. They were interested in what the HSPM would do when faced a syntactic main-clause/relative-clause ambiguity that contains a built-in semantic disambiguation.

- (1) The witness examined by the lawyer turned out to be unreliable.
- (2) The evidence examined by the lawyer turned out to be unreliable.

Prior to encountering the word ‘by’, the HSPM might interpret the verb ‘examined’ as a tensed verb that introduces a main clause, as in sentence (3).

- (3) The witness examined the specimen.

But, as sentences (1) and (2) demonstrate, the verb ‘examined’ might also be a passive participle that introduces a relative clause in which the optional complementizer ‘that’ has been omitted. As we noted in ch. III, the HSPM consistently prefers the main clause reading, in accordance with the principle of Minimal Attachment (MA). (See ch. V for further discussion of MA.)

In sentences (1) and (2), ‘examined’ turns out to be a passive participle, so we would expect—on the basis of syntactic considerations alone—to find a garden path effect in these cases as well. However, as noted above, sentences (1) and (2) were constructed in such a way that semantic or pragmatic information could, in principle, be used to avoid the incorrect reading in (2). If the HSPM has access to semantic or pragmatic information from the outset, then it should take into account the fact that witnesses are typically capable of examining things, while

evidence is typically not. In this case, we would expect the HSPM to predict the relative clause reading of sentence (2) and, hence, to encounter no difficulty when the word ‘by’ disambiguates the structure in favor of that reading. This is not, however, what Ferreira and Clifton observed. They report that the fixation times at the word ‘by’ were quite comparable for both (1) and (2). This finding strongly suggests that initial parsing decisions are made *without* the influence of semantic information.

Ferreira and Clifton’s result has been replicated several times. Rayner, Garrod, and Perfetti (1992) performed a similar experiment using materials designed to test, in addition, for effects of discourse focus. They write:

Our primary finding was that subjects were garden-pathed even when there was biasing context. However, when the target sentence was in the discourse focus, subjects were able to recover more readily from their initial erroneous parse of the sentence. The data thus support models of sentence parsing that postulate that the parsing of a sentence is based upon structurally based principles and the influence of semantic or pragmatic information makes itself felt only after the initial parsing decision has been made. (p. 109)

Similarly, in reporting the results of a cross-modal priming experiment, Nicol and Swinney (1989) found that “[w]hen structural information cannot serve to constrain antecedent selection, then pragmatic information may play a role, *but only at a later point in processing*” (p. 5, emphasis added).

Finally, as discussed in ch. III (§2), researchers in neurolinguistics have garnered a wealth of “ERP findings [that] provide strong converging support for the assumption that constituent structure information *hierarchically dominates* other information types such as semantics/plausibility” (Bornkessel-Schlesewsky and Schlewsky, 2009: p. 113).

What, then, should we conclude about the introspective observations that Schank, *et al.* report? To begin with, we should note that uncritical reliance on introspection is simply not

appropriate in this context. For instance, Schank, *et al.* claim that, upon hearing the string ‘*Seven Libyan gunmen shot their way into...*’, we do not notice the ambiguity in the sentence completion, ‘a bank’. But, while it’s true that awareness of this ambiguity does not enter into our phenomenology, sophisticated behavioral and neurocognitive experiments reveal that the unintended meaning (i.e., a riverbank) *is*, in fact, weakly activated in the mind of the hearer.<sup>7</sup> We can conclude that Schank *et al.*’s introspective observations do *not* constitute compelling grounds for adopting the IPH.

We now know that syntactic analysis operates very quickly (about 125ms after stimulus onset), and integrates sub-clausal units incrementally—typically morpheme by morpheme.<sup>8</sup> This is sufficient to explain how exposure to only the first few words of a sentence can give rise to a rich semantic analysis, without sacrificing the core claims of the syntax-first approach. Moreover, we have solid experimental evidence against “interactionist” models of initial parsing preferences—the models that, in effect, adopt Schank *et al.*’s “integrated processing” approach.

We can bring these conclusions to bear on Schank *et al.*’s fourth argument for the IPH—the argument on which they seem to rest most of their case. Throughout their paper, they make much of the fact that we process new stimuli in light of our past experiences—our memories. Specifically, they hold that understanding something, whether it be a sentence or a situation, is a process of finding the memories that are most appropriate or relevant to the current context, and using them to create a new memory structure for the present input. It follows, then, that our memories of past situations must color our expectations about the current situation. So, “the

---

<sup>7</sup> Steedman (2002) writes: “experiments like those of Swinney (1979) and Tanenhaus, Leiman, & Seidenberg (1979) show[ed] early transient activation of irrelevant lexical alternatives, including irrelevant syntactic types, which are only subsequently eliminated as an effect of prior biasing extra-sentential context” (p. 98). See also Fernandez and Cairns (2011: ch. 6) and references therein.

<sup>8</sup> See Bornkessel-Schlesewsky and Schlewsky (2009), chs. 7-12.

commonsense motivation for integrated processing is that it doesn't make sense not to make the fullest possible use of this relevant information in processing further input" (215). To illustrate, they ask us to consider the following sentence.

(4) The old man's glasses were filled with sherry.

This sentence typically gives rise to a conscious garden-path effect. It is reasonable to suppose that the effect is due to our interpreting the ambiguous word 'glasses' in light of our prior understanding of 'old man'.<sup>9</sup>

This error can be explained only by assuming that people analyzing this sentence apply their knowledge of the relation between age and eyesight before completing any putative low-level analysis of the sentence. If 'bartender' is substituted for 'old man' ..., the opposite assumption is made in processing, so no backing up is needed." (215)

Given that 'old man's glasses' immediately precedes 'filled with sherry', it seems that semantic information is being brought to bear rather quickly, precisely as the IPH would predict.

The remarks above about incremental syntactic processing serve to undermine this line of reasoning. For, as just noted, syntax-first models are perfectly compatible with semantic information being accessed *quickly*, so long as syntactic information is accessed *first*. In the case of sentence (4), it is not at all implausible to think that the HSPM assigns the category NP to 'The old man's glasses' and passes along this partial phrase marker to the "semantic analyzer" well before 'filled with sherry' is processed. This hypothesis is entirely consistent with what we know about the time course of syntactic and semantic processing. Moreover, it underwrites the very same explanation of the garden path effect associated with (4) that Schank *et al.* propose. Thus, the presence of this effect does *not* force the IPH on us, which is just as well, given that we have excellent independent reasons for doubting it.

---

<sup>9</sup> In a footnote, they speculate that the same is true of the famous garden-path sentence, *The horse raced past the barn fell*, originally due to Thomas Bever. We discussed such sentences at some length in ch. III.



Turning from anecdotal and introspective observations to experimental results, Schank *et al.* examine several studies that they take to show that

- i) people do not use the syntactic rules of transformational grammar in the course of language processing, and
- ii) the very strong claim of a completely autonomous syntactic processor is false.

The first such study is Slobin (1966), which used active and passive sentences, of which half were semantically reversible and half nonreversible.

- |                                                                  |                               |
|------------------------------------------------------------------|-------------------------------|
| (5) John saw Bill. / Bill saw John.                              | <i>active, reversible</i>     |
| (6) John was seen by Bill. / Bill was seen by John.              | <i>passive, reversible</i>    |
| (7) John ate the apple. / #The apple ate John.                   | <i>active, nonreversible</i>  |
| (8) The apple was eaten by John. / #John was eaten by the apple. | <i>passive, nonreversible</i> |

The sentences marked with a ‘#’ exhibit a semantic defect or pragmatic infelicity.<sup>10</sup> They serve to illustrate what it means for *John ate the apple* and *The apple was eaten by John* to be “nonreversible.”

Slobin asked his subjects to match a picture to each such sentence, and measured their reaction times on this task. He reasoned that, in the nonreversible case, the semantic properties of the words, combined with background encyclopedic knowledge, might be sufficient to correctly interpret the sentence. In the reversible case, however, those resources are *not* sufficient; syntactic information *must* be brought to bear on interpretation. Slobin found that, in the reversible case, people took less time to respond to the active form (5) than to the passive form (6). By contrast, in the nonreversible case, processing the active form (7) and the passive form (8) took an equal amount of time. From this, Schank *et al.* conclude that “semantic grounds, if sufficient, can override the necessity for certain kinds of syntactic processing” (221).

---

<sup>10</sup> Following the policy of ch. III, the present discussion does not presuppose any particular way of drawing the semantics/pragmatics distinction.

The problems with this argument have been widely discussed, notably by Forster and Olbrei (1973). Using a paradigm known as rapid serial visual presentation (RSVP), Forster and Olbrei found that nonreversible passive sentences actually took significantly *longer* to process than their active counterparts. This finding directly contradicts Slobin's. And although the RSVP paradigm has since been shown to be inadequate for certain purposes, it is surely more reliable than Slobin's picture-matching paradigm. Forster and Olbrei argue that the task of matching sentences to pictures does not reveal anything about sentence processing *per se*. They point out that subjects were slower at matching passive sentences to pictures even when there was a 3 second delay between the presentation of the sentences and the pictures. This is a problem for Slobin, because the HSPM is known to take significantly less than 3 seconds to process a sentence, regardless of whether it has an active or a passive form. A reasonable conclusion, then, is that the slower response times that Slobin observed in the case of passive sentences are to be explained by appeal to processing effects outside of the HSPM—perhaps in the visual processing system, or at the language-vision interface. We will return to Slobin's experiment in ch. V. In the meantime, let us look at the other results that Schank *et al.* discuss. It turns out that none of these constitutes a compelling case for the IPH.

Schank *et al.* present data from a study by Tyler and Marslen-Wilson (1977), whose purpose was to test the model proposed by Fodor, Bever, and Garrett (1974). On that model, syntactic analysis *within the clause* proceeded by the operation of a completely autonomous syntactic processor; no higher-level knowledge could bear on the process until a clause boundary was reached. However, as we have seen, this is not a theoretical commitment that the opponent of the IPH should take on board. On a reasonable syntax-first model, both syntactic and semantic analysis are strongly incremental, operating in a morpheme-by-morpheme fashion,

even within the clause.<sup>11</sup> Thus, while Fodor, Bever, and Garrett's model has been disconfirmed, the syntax-first approach has not.<sup>12</sup>

Turn now to considerations pertaining to pronominal reference. According to Schank *et al.*, Schwartz (1980) found no evidence to suggest that hearers use syntactic information to resolve binding ambiguities in pronouns. They take this kind of result to be important because, "to the extent that processes that might have been thought to depend on explicit syntactic representations can be found not to, the integrated processing hypothesis is strengthened" (222). This reasoning relies on a premise to the effect that the role of syntactic knowledge is to *resolve* binding ambiguities. But this is false. As we will see in §2.3, the role of syntactic knowledge is merely to *rule out* certain binding possibilities, thereby *constraining* semantic analysis. Besides filtering out grammatically impossible binding relations, syntactic representations *leave open* how a pronoun should be bound.

This concludes our assessment of the arguments that Schank *et al.* provide for the IPH. I hope to have shown that these arguments are, without exception, wholly underwhelming. Moreover, as I have noted here and in ch. III, a range of behavioral and neurocognitive studies provide compelling grounds for rejecting models on which semantic and pragmatic resources are brought to bear on the HSPM's initial parsing decisions. These results make it unlikely that the models developed by Schank and his associates are psychologically plausible. Nevertheless, it

---

<sup>11</sup> The most striking evidence for this comes from some of the neurocognitive studies that we looked at in ch. III (§2). For instance, consider the way in which Hahne and Friederici (2002) combined semantic and syntactic violations in one and the same stimulus, by relying on the HSPM to take note of one morpheme just before taking note of another morpheme in the very same word.

<sup>12</sup> Schank *et al.* also discuss studies that show that processing indirect speech *in context* takes the same amount of time as processing direct speech, and less time than processing indirect speech *out of context*. This shows that pragmatic resources, when available, speed up the comprehension process. Taken as an argument for the IPH, this line of reasoning falls prey to a now-familiar objection: No one doubts that pragmatic resources come to play a role in comprehension *at some point*. What must be established is that they come into play from the very outset of the process. The studies that examine differences between direct and indirect discourse do not bear on this question.

will be worthwhile to examine the way in which these models incorporate the integrated processing hypothesis. Understanding the inner workings of such models, as well as their limitations, will help us to see why the construction of mental phrase markers is an absolutely crucial step for any effective comprehension system.

## §2.2 Integrated Processing Models: Inner Workings and Local Achievements

In laying out more precisely the technical features of their proposal, Schank *et al.* draw a helpful distinction between three aspects of a language processing system: (i) the system's control structure, (ii) the representations it generates, and (iii) the knowledge-base that it draws on. Specifying a control structure involves saying what inputs the system will receive, what outputs it will produce, and what processes and sub-processes are needed to effect these input-output transitions. Questions then arise about what information is encoded in the representations that the system generates in the intermediate stages of processing, as well as what background information the system needs to have at its disposal in order to compute those representations.

With regard to control structures, Schank and Birnbaum distinguish four positions.

- [a] Syntax and semantics are *completely separable*: Syntactic structure is computed first, without any input from semantic analysis, or from higher-level sources of knowledge. (Schank, *et al.* treat this as a naïve interpretation of Chomskyan grammar as a processing model.<sup>13</sup>)
- [b] Syntax and semantics are *nearly decomposable*: Syntactic structure is computed first, but the parser sometimes calls for information from the semantic processor. The semantic processor cannot, on this view, *autonomously* initiate a syntactic process.
- [c] Syntax and semantics have a *heterarchical relationship*: They are separate processes, but each can call on the other at any stage of processing. (Schank *et al.* cautiously attribute this position to Terry Winograd.)

---

<sup>13</sup> I provided textual evidence, in ch. 1, for the claim that Chomsky does *not* intend his syntactic theory as a processing model. Nevertheless, the claim that syntactic and semantic processing are completely separable, in the sense that Schank *et al.* intend, may very well be true of the HSPM. Indeed, I argued above for the psychological plausibility of a syntax-first model, which is compatible with claims [a] and [b].

[d] Syntax and semantics are used in an *integrated control structure*. There's a single control structure that decides whether to use syntactic or semantic resources at any given time. (Schank *et al.* attribute this position to Marslen-Wilson.)

They note that the key difference between [c] and [d] is that the former, unlike the latter, constructs distinctly *syntactic* representations in the intermediate processing stages. With this in mind, Schank *et al.* opt for position [d], which eschews such representations.

With regard to representations, Schank and Birnbaum identify three possible positions.

- I. The view of Fodor, Bever, and Garrett (1974), according to which “the structural analyses to be recovered are ... precisely the trees generated by the grammar.” (quoted on p. 219 of Schank and Birnbaum, 1984)
- II. There is no independent level of purely syntactic representation. Syntax and semantics are represented together, in one data structure, but the syntactic aspects of that structure are used for purposes that are distinct from those that require the semantic aspects.
- III. There is no independent level of purely syntactic representation. The only legitimate reason to posit syntactic aspects of the representations is if they serve some semantic or pragmatic purpose.

Schank *et al.* opt for position (III), which denies the existence of mental phrase markers.

Finally, with regard to the issues concerning the comprehension system's knowledge base, Schank and Birnbaum note that the range of options depends on one's prior commitments regarding the need for syntactic representations. If, for instance, one adopts the position of Fodor, Bever, and Garrett—i.e., position (I), above—then one is thereby committed to endowing the comprehension system with distinctly syntactic rules.<sup>14</sup> On the other hand, if one adopts positions (II) or (III), then the following options open up:

- (i) Syntactic rules exist, but they serve no distinct function from any of the other rules.
- (ii) There are no specifically syntactic rules at all.

---

<sup>14</sup> In the language of Devitt (2006), this claim can be rendered as follows: As soon as one buys into the psychological reality of mental phrase markers, one is thereby committed to the existence of (possibly implicit) mental representations of syntactic rules or principles—something akin to the Representational Thesis (RT). The inference embodied in this claim is one that we will explore in ch. V. Devitt, of course, rejects it.

Schank and Birnbaum opt for (i), the weaker of these two claims. In discussing their models, they provide examples of rules that make reference to both syntactic and semantic features of the utterance. It should be noted, however, that their conception of syntax is anemic—restricted, in essence, to precedence relations in linear word order and, in the extreme case, the notions of subject and object. They summarize the core claims of their integrated processing models:

We are now in a position to state exactly what the integrated processing hypothesis claims. First, it claims that language analysis proceeds as a unitary process, integrating all kinds of knowledge, rather than as a collection of separate processes, one for each kind of knowledge. This is contrast to the models of Fodor, Bever, and Garrett, Woods, and Marcus, among others. Second, it claims that no independent level of syntactic representation is constructed, operated on, or output by the language analysis process. This is in contrast to all of the above models, as well as the model proposed by Winograd. Third, it claims that, although there are rules that are some sense purely syntactic, such rules are not used differentially from other sorts of rule; that is, they are functionally integrated in processing and play no privileged role. This follows from the first two claims. (pp. 220-21)<sup>15</sup>

Let us now see how the integrated processing models operate and what capacities they exhibit. Following the presentation in Schank and Birnbaum (1984), I will discuss several concrete examples, in order of increasing complexity.

To begin with, let us take a toy example, where the input is the ordered set of words *Fred ate an apple*, individuated orthographically. The system scans from left to right, activating the concepts associated with each word. In cases of ambiguity, the system activates all of the concepts and attempts to satisfy, in parallel, each of their semantic and pragmatic requirements. When the requirements of a concept cannot be met given the rest of the input, the concept is dropped from the running. Here is a walk-through of the system's high-level activities:

---

<sup>15</sup> We note that Schank *et al.* endorse the idea that the structure rules of a language must be represented in a comprehension system. On this point, they differ from Devitt (2006), who denies this claim, in the guise of the Representational Thesis (RT).

1. 'Fred' activates a concept of an animate human male.
2. 'ate' activates the concept INGESTION, which requires an animate agent and an edible object.
3. 'Fred' is bound to the agent position of INGESTION.
4. 'an' activates the concept of indefiniteness.
5. The system predicts the presence of a concept for some type of edible object.
6. 'apple' activates the concept of an edible fruit.
7. 'an apple' is bound to the object position of INGESTION

The final product can be represented thus: [INGEST ACTOR (FRED) OBJECT (APPLE REF (INDEF))]

On account of their almost exclusive focus on semantic and pragmatic processing, the central feature of such models is how they encode high-level information in their knowledge base. It is here that the technical devices known as “scripts,” “cases,” and “frames” (among others) come into play.

One of the first ... memory structures devised to capture this high-level semantic knowledge ... and pragmatic knowledge necessary for language understanding ... was the script. A script is a temporally and casually linked set of low-level concepts describing a time-ordered stereotypical sequence ... paradigmatic[ally], going to a restaurant” (213).

For instance, when the system scans the input and recognizes the word ‘restaurant’ (or words that express semantically related concepts) it immediately activates the concepts that are typically relevant for understanding a restaurant visit—the concept of a host, a waiter, ordering, eating, china, utensils, the bill, the tip, and so forth. Each of these concepts encodes relatively sophisticated world knowledge, regarding which items are animate, which are edible, and so on.

Schank and Birnbaum describe techniques for scanning input text to find the “most important” words, and using them to activate the relevant concepts. Other input items are either discarded as “uninteresting” or kept in a buffer, awaiting “expansion,” which occurs if and when the information that they contain becomes needed. This strategy saves on processing time and computational resources. A piece of information comes to be needed when an activated concept is “expanded” and requests its “slot fillers”—i.e., concepts that would provide the information that is typically associated with the activated concept. For instance, given that a restaurant visit

typically involves human participants and food they eat, the concepts PARTICIPANT and FOOD would be slot fillers for the higher-level concept RESTAURANT. This guides the system in scanning the rest of the sentence for specific kinds of slot fillers, as dictated by the “important” concepts that have already been activated and expanded. In this way, the system can prioritize its activities and save time and resources by ignoring those words that provide no immediate slot fillers for the most important currently active concepts.

The notion of “most important concept/word” is bound to seem obscure at first. It is not clear what principled grounds there can be for determining whether a word is likely to provide important information. Offhand, it seems like *any* word can do that, given the right context. Schank *et al.* do provide a formal characterization, in terms what they call memory organization packets (MOPs), but the details of these data structures are beyond the scope of the present discussion. (See Schank, 1979b.) However, an example will serve to convey the impressive results that such models can achieve. Schank *et al.* programmed their SAM and FRUMP systems to use MOPs for the task of summarizing stories and newspaper articles. Consider the following input text and the system’s final output, generated in the space of several milliseconds.



**Input:** *An Arabic-speaking gunman shot his way into the Iraqi embassy here this morning, held hostages throughout most of the day before surrendering to French policemen and then was shot by Iraqi security officials as he was led away by French officers. (From the New York Times)*

**Output:**

**\*\* MAIN EVENT \*\***

EVENT 1 =

SCRIPT: *TERRORISM*  
ACTOR: *ARAB GUNMAN*  
PLACE: *IRAQI EMBASSY*  
INTEREST: *9 OUT OF 10*  
CITY: *PARIS*  
TIME: *MORNING*

**SCENES**

EVENT 2 =

SCENE: *HOSTAGES*  
PLACE: *IRAQI EMBASSY*  
ACTOR: *ARAB GUNMAN*  
INTEREST: *7 OUT OF 10*  
TIME: *DAY*

EVENT 3 =

SCENE: *CAPTURE*  
PLACE: *IRAQI EMBASSY*  
OBJECT: *ARAB GUNMAN*  
ACTOR: *POLICEMEN*  
INTEREST: *6 OUT OF 10*  
AFTER: *EVENT 2*

**\*\* UNEXPECTED EVENTS \*\***

EVENT 4 =

ACTION: *PROPEL*  
ACTOR: *IRAQI OFFICIALS*  
OBJECT: *ARAB GUNMAN*  
ITEM: *BULLETS*  
DIR. FROM: *GUN*  
INTEREST: *5 OUT OF 10*  
AFTER: *EVENT 3*

**RESULT**

EVENT 5 =

STATE: *DEAD*  
ACTOR: *ARAB GUNMAN*  
INTEREST: *4 OUT OF 10*

*Source: Schank and Birnbaum (1984: pp. 234-5)*

As the output of the system demonstrates, we are dealing with a very sophisticated natural language processing device.<sup>16</sup> Unfortunately, at the time that Schank and Birnbaum reported these results, the example given above was not indicative of the system's performance in the general case. In actual fact, the successful outputs constituted only 10% of the total.<sup>17</sup> Schank *et al.* attributed this low success rate to the impoverished theoretical and encyclopedic

---

<sup>16</sup> Note that the output is a representation of events that bear temporal and hierarchical relations to each other. This representation is itself couched in a linguistic medium, and plainly has a syntax of its own. It is, therefore, one way of implementing the Language of Thought Hypothesis (LOTH). However, the syntax of this representation is radically different from the syntax of English. This observation will play a crucial role in §3.

<sup>17</sup> This estimate is taken from Walter Kinsch, "Approaches to the Study of the Psychology of Language," in *Talking Minds: The Study of Language in Cognitive Science* Bever, T.G., Carroll, J.M., and Miller, L.A., eds., MIT Press: Cambridge, MA, pp. 111-145.

knowledge encoded in the system's conceptual framework. Accordingly, they undertook the project of building a sophisticated knowledge base—an effort that continues to drive impressive advances in what, today, we call “expert systems.”

We have seen that the integrated processing models operate largely on the basis of semantic and pragmatic knowledge. What about their syntactic resources? As noted above, Schank *et al.* are skeptical that the concepts from contemporary syntactic theories will play a crucial role in a comprehension system.

Traditional notions of syntax use categories like ‘part of speech’ and ‘phrase structure’ in discussing the structure of a sentence. ... [W]e would like to claim that these notions of syntax are inappropriate to attempts to describe and use syntactic knowledge in a language understanding process. (p. 228)

They maintain that syntax is only of use when local semantic information and background pragmatic knowledge are not sufficient to establish the relations between concepts. They believe that such cases will be relatively rare and that, when they arise, information about the linear order of words in the input string will be sufficient to pick up the slack. For instance, the rule that Schank and Birnbaum propose for handling relative clauses—a rule triggered by the occurrence of a complementizer phrase (‘which’, ‘who’, ‘that’, etc.)—states that

[t]o the right [of the complementizer phrase] will be found a concept with some unfilled slots. Use the concept to the left [of the complementizer phrase] to fill one of the slots, in accordance with semantic requirements. Then take the resulting conceptualization and subordinate it to the concept on the left [of the complementizer phrase]. (p. 230)

Schank and Birnbaum point out that while this rule is stated in terms of precedence relations—e.g., ‘to the left of’—it ultimately has to appeal to semantically defined positions. Their point is that “syntactic and semantic knowledge cannot be distinguished by use in this model of language analysis” (231).

With this sketch in mind, we are now ready to examine the limitations of integrated processing models. The discussion in the following section draws heavily on Marcus (1984). Following Marcus, I will argue that such models, while impressive and potentially illuminating, are not adequate *even in principle* to handle any but the simplest linguistic inputs.

### §2.3 Criticisms of the “Mostly-Semantic” Models

Marcus begins his discussion by agreeing with his opponents that there are reasons to *hope* that a comprehension model can make do without fancy syntactic resources. Such models would be theoretically parsimonious, in that they make use of no information beyond that which is inherent in problem domain—i.e., *linguistic meaning*. Moreover, he concedes that contemporary syntactic theories fail to either explain or to even address many interesting natural-language constructions. Nevertheless, Marcus argues that a model that doesn’t take into account *hierarchical* syntactic relationships will encounter pervasive failures, even on typical inputs.

Many workers in both psychology and artificial intelligence hold some form of [what] I call the *no explicit syntax* hypothesis. They therefore dismiss the claims of linguists that a language has specific additional structure, which each speaker of that language knows tacitly but explicitly. ... Often these workers ... insist ... that language can be understood by applying a system armed with only the minimum of information explicitly about the syntax of the language. ... I believe that these researchers are wrong. Furthermore, I believe that the models presented to date in support of such claims, to the extent that they remain true to the claims, are and will remain fundamentally inadequate to handle the range of grammatical phenomena well known and understood within the linguistics community for the last ten years. Perhaps surprisingly, these phenomena are part of the core of the syntax of English... In short, I argue that any model based on the hypothesis that language understanding involves no explicit syntactic structure is fundamentally inadequate to process the full range of natural language. If this is true, such models are simply not candidates for models of the process of understanding natural human language. ... [And] if these models fail the prior test of being computationally sufficient, questions of psychological reality simply don’t apply.” (Marcus, 1984: pp. 253-255, underlining added; italics in the original)

Following the presentation of Marcus (1984) we will first look briefly at models that use only semantic information. We then turn to those that also use linear order, and examine some simple extensions of such models, which are designed to handle problem cases. As will we will see,

these extensions, which are essentially those incorporated in the published models, are based on limited subcases of [the problematic] phenomena, and fail to handle any of their more complex manifestations. ... [T]hese methods suffice to capture only the subcases of these phenomena that involve local syntactic structure (that is, in which all relevant aspects of structure are contained in a single clause), but that these phenomena can affect global structure.” (p. 255)

As noted above, systems that make use of only semantic information are able to rule out some incorrect analyses by using semantic and pragmatic information—the meanings of the words in the input, combined with background encyclopedic knowledge. For instance, in *John bit the apple*, the system can tell that John did the biting—not the other way around—because it knows that inanimate things don’t bite. Marcus points out, however, that these resources will not always be adequate. Indeed, in the worst cases, reliance on them can lead a system to make serious errors. For example, confronted with the input *The postman bit the dog*, such a system would confidently issue the analysis, THE DOG BIT THE POSTMAN. But, while this analysis matches the system’s pragmatic knowledge to the effect that dogs bite more often than postmen do, the system’s output is obviously wrong. There will also be cases where pragmatic knowledge is either unavailable or irrelevant. For instance, a system that doesn’t know much about John or Bill, will incorrectly take the sentence *John insulted Bill* to be ambiguous. Marcus draws two morals from this. First, *linear order matters*. Second, and more important:

Examples like [*The postman bit the dog*] rule out as cognitive models *semantics-mainly* approaches that use syntax only *where necessary*—that is, where a semantics-only analyzer runs into trouble. The problem, of course, is that semantics and pragmatic expectations can sometimes give a straightforward, and wrong, analysis. Wrong but highly plausible analyses will always cause such a semantics-mainly approach to overlook a right but implausible analysis. This seems to directly rule out positions like the following, which is “Principle III” in Riesbeck and Schank: “A

parser must take care of syntactic considerations only when required to do so by semantic considerations... We simply believe that syntactic considerations should be done only when they are needed, that is, after other more highly ranked considerations are used. We have turned the syntactic approach of ‘semantics only when needed’ around... (p. 9).” Examples like [*The postman bit the dog*] show that one cannot always tell whether syntax was needed or not.” (p. 261)

These considerations give rise to models that are enriched with representations of the linear order of the input string, as well as some rudimentary syntactic notions. Such models use a technique that Marcus calls “Positional Template Filling” (henceforth, PTF). To illustrate how PTF works, imagine a user of a computer system querying the system in ordinary English. Suppose the user is asking about the history a specific file, viz., file123. The input to the comprehension system is as follows: *Has anyone accessed file123 in the last week?*

The PTF method draws on semantic entries for predicates, which contain some basic syntactic information. This includes the typical positions of the predicate’s arguments—e.g., subject and object positions. For the input sentence we are currently considering, the most important semantic entry is for the predicate ‘ACCESS’:

(9) ACCESS (subject: USER, object: FILE, {“on,” “before,” “in,” ...}: TIME)

A program employing the PTF method begins by scanning the input to find a verb. Having found ‘access’, the semantic entry in (9) guides the system to scan rightward for the object and the prepositional phrase, and leftward for the subject, resolving any ambiguities it encounters by drawing on the semantic and pragmatic resources at its disposal.

Passive sentences, like *Has file123 been accessed by anyone in the last week?*, are handled by drawing on *separate* semantic entries for the passive form of each verb. These differ from the active-form semantic entries in that the subject position occupies a different place in the linear order and the object position is optional. The passive counterpart of (9) would thus be:

(10) ACCESS (“by”: USER, subject: FILE, {“on,” “before,” “in,” ...}: TIME)

The PTF algorithm begins by finding the verb and scanning rightward, to see if the verb carries any passive morphology (e.g., ‘ed’). If so, then it scans leftward, to see if there is a passive marker (e.g., ‘been’). If these are both found, the passive semantic entry for the predicate ‘ACCESS’ is activated. This, in turn, instructs the system to scan forward from the verb to find ‘by’, and to scan backward from the verb to find the FILE in subject position. This strategy requires explicitly representing a bit of syntax—the order of the words and the passive morphology of the verb—but not very much.

Marcus discusses three problems with the PTF method. The first has to do with passive constructions. The second problem concerns constructions that have gaps. Finally, he considers the phenomenon of pronominal binding. We will consider these in turn.

The problem with the PTF strategy for handling passive sentences is that it incorrectly treats the passive form as a simple clause-bound construction. Consequently, the only cases it can handle are ones in which there is only a single clause and only one verb to be passivized. More complex cases immediately lead the strategy to fail. Consider, for instance:

(11) Is the disk crash suspected to have damaged my files?

When the system encounters (12), it finds the verb ‘suspected’ and activates the passive semantic entry for this predicate:

(12) SUSPECT (“by”: ANIMATE-ENTITY, subject: MENTAL-OBJECT)

Searching for the subject of the suspicion, the system scans leftward and finds ‘the disk crash’.

One can’t very well suspect a disk crash, so this is not a plausible candidate for the category

MENTAL-OBJECT.<sup>18</sup> No other candidates are available, so the system fails. Multiple passivization, as in (13), gives rise to still further difficulties.

(13) Were the files that were suspected to have been deleted important?

First, the system will fail to fill in the passive semantic entries for *both* verbs, for the very same reasons that it fails in cases like (11). Second, even if we imagine this problem to have been solved, the system would *still* fail, on account of the presence of a “gap” after ‘deleted’. Let us look more closely at the problems associated with sentences that contain gaps.

Consider the following input: *Are any files I accessed last week over 2 months old?*

From the perspective of formal syntax, this sentence has a gap, as represented in (14).

(14) Are any files that I accessed \_\_\_\_ last week over 2 weeks old?

The gap is a position in which the object of ‘accessed’—viz., ‘any files’—would normally be found. However, ‘any files’ has been “displaced” to the front of the sentence, in accordance with the syntax of yes/no questions in English. Despite its displacement, ‘any files’ nevertheless serves as the thematic object of ‘access’. Presented with (14), the PTF algorithm would first find the verb ‘access’ and activate its semantic entry. It would then scan leftwards for a subject, whereupon it will correctly locate “I”. But the rightward scan for the object will not find ‘any files’, because ‘any files’ has been displaced to the left of the verb. Having found no item to fill the object slot of ‘ACCESS’, the algorithm would simply fail.

A proponent of the PTF strategy might suggest that this can be fixed by searching leftward for a complementizer phrase (“which,” “who,” or “that”) and treating the material to the

---

<sup>18</sup> One can, of course, suspect a disk crash *of* something—e.g., of causing some further calamity—but that’s different from simply suspecting a disk crash. The event *disk crash causes calamity* would be a perfectly fine MENTAL-OBJECT; the entity *disk crash* is not.

left as the object of ‘access’. There are two problems with this suggestion. First, as (14\*) illustrates there won’t always be a complementizer phrase for the system to use in this fashion.

(14\*) Are any files I accessed \_\_\_\_ last week over 2 weeks old?

Second, and more important, the object of ‘access’ can be arbitrarily far back. Take, for instance, the following input:

(15a) Is the file, which the directory that I created contains, more than 2 weeks old?

(15b) Is the file, which the directory that I created \_\_\_\_ contains \_\_\_\_, more than 2 wks old?

In this case, a leftward scan from ‘that’ will yield ‘the directory’ and a leftward scan from ‘which’ will yield ‘the file’. The PTF system, even if it could be enriched with resources for locating gaps—especially multiple gaps in a single sentence—still has no way of knowing which phrases to bind to which gaps.<sup>19</sup> Largely the same point holds with respect to the example that Marcus provides:

(16a) Have the announcements I queued to have the mail system send out gone out yet?

(16b) Have the announcements I queued \_\_\_\_ to have the mail system send out \_\_\_\_ gone out yet?

Consider next the fact that a gap in a *wh*-question cannot occur just *anywhere*. As we observed in ch. III, there are complex constraints on noun movement. Marcus presents the following case, which illustrates that movement of a noun out of a clause, as in (18), is acceptable, whereas movement out of a noun phrase, as in (20), is not.

(17) John claimed that Bill visited Sue.

(18) Who did John claim that Bill visited?

(19) John made the claim that Bill visited Sue.

(20) \*Who did John make the claim that Bill visited?

---

<sup>19</sup> Notice that world knowledge won’t help in this case; files and directories are both the types of things that can be created and that can contain other things.



Now, in order to correctly process wh-questions, a system must know where gaps can appear. And in order to know *this*, it must *at the very least* mark the difference between noun phrases and clauses.<sup>20</sup> When combined with our earlier observations, we see that the system must also explicitly mark the orthogonal distinction between unembedded, singly-embedded and doubly-embedded phrases. The idea that semantic and pragmatic considerations alone will be sufficient to mark these distinctions, or to render them otiose, is very implausible. Marcus concludes that

[w]hen we turn to the implemented systems [of the Yale school], we see that without exception they handle the clause-bound special case of passive. All [these] systems, which fail to handle relative clauses and “wh-” questions, also handle only the special case of clause-bound passives. Thus, no one has yet managed to make the ‘semantics-mainly’ idea actually work.” (p. 272)

These points lead one to suspect that *any* syntactic phenomenon whose explanation requires appealing to hierarchical structure, as against mere precedence relations, will cause problems for the PTF model. This includes quantifiers, negation, adverbs, and a great deal more. Capitalizing on this insight, Marcus focuses on binding relations, which have been known since the late 1970s to be sensitive to the hierarchical dependence of the anaphor (e.g., a pronoun) on its antecedent. He asks us to consider the following set of data.

- (21) File123 disappeared when it was copied.
- (22) When file123 was copied, it disappeared.
- (23) When it was copied, file123 disappeared.
- (24) It disappeared when file123 was copied.

Note that that ‘file123’ can be co-referential with ‘it’ in all of these cases, except the last. The formal explanation, briefly, is that in (21)-(23), the antecedent does not c-command the pronoun in its governing domain, whereas in (24) it does. It has been demonstrated experimentally that the HSPM is sensitive to the hierarchical structures that underwrite c-command relations. We

---

<sup>20</sup> In actuality, of course, the distinctions are much more subtle. But even the crude distinction between noun phrases and clauses is sufficient to demonstrate the point.

considered some of the evidence in ch. III, particularly the study by Nicol and Swinney (1989). Focusing specifically on binding and anaphora, Frank, Mathis, and Badecker (2008) provide references to recent studies that confirm the same point.

[W]ith respect to the anaphoric phenomena under discussion the empirical patterns of coreference possibilities that are found in studies of sentence processing are precisely aligned with those that are predicted by the abstract conditions on co-reference [in (9) and (14) below]. Experimental results from Nicol and Swinney (1989), Gordon and Hendrick (1997), Asudeh and Keller (2001), Badecker and Straub (2002) and Sturt (2003) uniformly confirm people's sensitivity to a c-command-based locality condition in pronoun and reflexive interpretation, along the lines of those in (9) [A reflexive must have as its antecedent a c-commanding noun phrase in the same clause] and (14) [A pronoun may not take as its antecedent a c-commanding noun phrase within its own clause]. (p. 11)<sup>21</sup>

What this shows is that, in order to correctly handle sentences like (21)-(24), an effective (let alone a psychologically plausible) comprehension system must know when various terms can co-refer. And this requires, at a minimum, keeping track of c-command relations.<sup>22</sup> Lacking such distinctly syntactic information, the system will err frequently. And, again, there is just no reason to think that semantic or pragmatic knowledge can make up the difference in all, or even a significant number, of cases. The *only* way of *systematically* ruling out the incorrect interpretations is by reference to the hierarchical relations like c-command, which are defined in terms of both precedence *and* dominance.<sup>23</sup>

---

<sup>21</sup> Frank, Mathis, and Badecker are speaking somewhat informally when they state the conditions on co-reference in terms of clauses. A more precise statement would appeal to the notion of “governing domain.” See Haegeman (1994) for extended discussion.

<sup>22</sup> Needless to say, in the general case, syntax alone does not *determine* which phrases are co-referential. Judgments of co-reference are typically made in light of semantic or pragmatic knowledge. As noted in §2.2, syntactic processes merely *prune* the binding options, in advance of semantic processing.

<sup>23</sup> Carnie (2010) provides a clear and detailed discussion of various formal definitions of command in terms of precedence and dominance (see esp. chs. 3-4). I discuss precedence and dominance relations in ch. V.

## §2.4 Conclusion

We saw in §2.1 that the arguments Schank and Birnbaum (1984) advance in favor of their integrated processing hypothesis (IPH) are not at all compelling. Moreover, we reviewed a range of experimental evidence against the IPH. Still, the models that these theorists constructed on the basis of the IPH were capable of rather impressive feats (§2.2). We can now see that the fundamental problem with these models is not their so much commitment to the IPH, but rather their insistence on the stronger claim, “that no independent level of syntactic representation is constructed, operated on, or output by the language analysis process” (220). On the basis of the considerations reviewed in §2.3, Marcus (1984) concludes that “[t]he range of possibilities for the interaction between syntactic and semantic components of the process remains an open question, but there is no question that an independent syntactic component is justified and necessary” (277). I take his arguments to this effect to be decisive.

As effective as they are, Marcus’s arguments do not, I think, address the concerns raised in Devitt (2006). By way of introducing the topic of §3, let me elaborate on this claim.

Schank *et al.* eschew representations of syntactic structure because they believe that other resources—primarily, semantic and pragmatic information—can be used for all of the purposes for which syntactic information might have been thought to be necessary. Now, Devitt agrees with Schank *et al.* that the HSPM does not construct mental phrase markers. However, his reasons for this appear to be somewhat different from theirs. In particular, although Devitt does occasionally flirt with the idea that semantic and pragmatic resources render mental phrase markers otiose,<sup>24</sup> he nevertheless seems to concede that the “clues” which enable the HSPM to

---

<sup>24</sup> “We can go further down this brute-causal path ... Constraint-based theories, with their links to connectionism, certainly give support to a brute-causal associationist picture of ... the speedy automatic language comprehension. ... On all these theories, we can see an ambiguous word of an utterance as activating the several concepts with

treat a word as, e.g., an adjective, may well be purely syntactic (*Ignorance*, p. 225).<sup>25</sup> His rejection of mental phrase markers is, therefore, most likely grounded in rather different concerns than those that animate the models developed by Schank *et al.*

In §3, we will see that Devitt's motivations have more to do with his notion of *representation*—in particular, his distinction between *responding* to something as, e.g., an adjective and *representing* it as being an adjective. I believe that this distinction is the single most important piece of Devitt's case against mental phrase markers. Without invoking it in the present context, his brute-causal model cannot answer even the most basic questions about how linguistic expressions of any complexity are disambiguated. I will discuss this distinction at some length in §3.3, arguing that the operations of the HSPM cannot be seen as mere “responses,” nor even as “responses to something *as* something” (if this is not simply a covert way of invoking representations). My discussion of this issue will also serve another purpose—viz., to place constraints on when it is appropriate to posit representations, and hence to give more substance to the notion of representation that is at play in this discussion. The payoff that we derive from being explicit about these constraints will emerge in ch. V, when we turn to questions regarding the very nature of mental representation.

---

which it is linguistically associated, and an ambiguous structure as activating the several conceptual structures with which it is linguistically associated. At the same time multiple factors arising from general knowledge and context suppress some associations and enhance others, all in a fairly brute-causal way” (*Ignorance*, p. 239).

<sup>25</sup> See also his brief discussion of adverbs on p. 185.

### §3. The Brute-Causal Model of Language Processing

#### §3.0 Introduction

As noted above, the defining claim of the position tentatively endorsed by Devitt (2006) is that the HSPM does not construct mental phrase markers. My main target will be the strongest version of this view—what I’ll call the *Strong Brute-Causal* (SBC) model. In addition to rejecting the psychological reality of mental phrase markers, the SBC model is committed to what I’ll call the Strengthened Language of Thought Hypothesis (SLOTH). According to SLOTH, mental representations (particularly, thoughts) are couched in a linguistic medium—the hypothesized language, Mentalese—whose syntax bears a strong resemblance to that of the hearer’s public language.<sup>26</sup> The SBC model treats comprehension as a matter of mapping acoustic representations of the incoming linguistic input *directly* into thoughts—i.e., expressions in Mentalese—whose syntactic structure is assumed to be either similar or identical to that of the public-language stimuli.

The discussion that follows is aimed at casting doubt on the core commitments of the SBC model. I begin in §3.1 by examining Devitt’s argument for SLOTH. I raise problems for this argument—particularly, its clash with Devitt’s other commitments regarding the explanatory ground of syntactic properties and the nature of brute-causal processing. In §3.2, I note a number of qualifications that Devitt makes regarding his proposal, and present Devitt’s “official” statement of the SBC model. In §3.3, I bring together Devitt’s remarks about brute-causal processing, from which I distill three defining properties: (i) associationist transitions, (ii) reflex-like simplicity and rigidity, and (iii) the absence of structure-sensitive operations on mental representations. I then assess whether language processing is brute-causal in this sense.

---

<sup>26</sup> Throughout the remainder of this discussion, my use of the term ‘public language’ will be neutral as between dialects and idiolects. See ch. II for a discussion of the issues pertaining to the notion of a public language.

My claim will be that the operations of the HSPM exhibit none of these three characteristics. Finally, I address Devitt's appeal to the distinction between *representing* something as having a certain property and *responding* to it as having that property. I conclude that the operations of the HSPM cannot be viewed as mere "responses" in the sense that Devitt seems to intend.

### §3.1.1 Background Commitments: RTM, CTM and the Language of Thought Hypothesis

The Representational Theory of Mind is a version of what Jerry Fodor calls *Intentional Realism*. According to this view, the scientific study of cognition would do well to co-opt some of the notions that feature in our everyday folk-psychological explanations—paradigmatically, the notions of belief and desire—and to tentatively adopt the basic commitments of folk psychology. These commitments include the idea that rational agents are in causally potent internal states that modulate their behavior in accordance with environmental changes. Intentional Realism is founded on the claim that these states—perceptual judgments, beliefs, desires, hopes, worries, and so forth—have *propositional contents*, much like the sentences of a natural language, or the speech acts in which those sentences are employed.<sup>27</sup>

Appending to Intentional Realism the minimal physicalist claim that token mental events are identical with token neural events, we arrive at the following view: Mental states like belief and desire are states of the brain that have semantic values and are implicated in the causation of behavior. This is the Representational Theory of Mind (RTM). The dual burden on RTM is to explain how mental states can have both semantic and nonsemantic properties, and to give an account of what role their properties play in explaining our behavioral and cognitive capacities.

---

<sup>27</sup> Terminological note: In an effort to stay neutral on the issue of which public-language entities are the bearers of propositional content—token utterances, speech acts, sentence types, logical forms, etc.—I use the term 'public-language expression' to deliberately elide a number of distinctions that would be useful and relevant in other contexts.

Fodor argues that mental states, *qua* neural states, can serve to modulate behavior in response to stimuli, and that the same states, *qua* bearers of intentional content, can appear in the psychological laws that quantify over semantic properties. On this view, cognition is an orderly procession of semantically evaluable neural states, governed by the “laws of intentionality,” which are themselves implemented by neural mechanisms.<sup>28</sup>

Fodor takes for granted the standard (though by no means uncontroversial) view that propositions are structured entities, which have conditions of truth or satisfaction. As formulated above, however, RTM it is *not* committed to an analogous claim concerning the *vehicles* of those states. For instance, while the theory entails that a neural state (N) can have the same content as the English sentence ‘Space-time is curved’, it does *not* entail that some component of N has the same content as the English term ‘space-time’, and that another component of N has the same content as the predicate ‘is curved’. Nor does it entail that such components bear distinctly *syntactic* relations to one another. These additional claims constitute Fodor’s Language of Thought Hypothesis (LOTH).<sup>29</sup>

According to LOTH, all propositional attitudes involve *syntactically structured* mental representations. Complex mental representations are said to have atomic symbols as their constituents, and the contents of the atomic constituents, as well as the syntactic relations they bear to one another, are said to exhaustively determine the contents of the complex

---

<sup>28</sup> It is important to note that Fodor rejects the idea of purely intentional mechanisms. He stipulates that the causal chains in which mental states figure—the causal chains that begin with sensory stimuli and eventuate in behavior—must be specifiable in purely physical (presumably neurophysiological) terms.

<sup>29</sup> Notice that my statement of LOTH made reference to two constituents—a subject and a predicate. If these are the *atomic* constituents of the thought, then the syntax of thought is more like that of first-order logic than that of English—a routine assumption in discussion of LOTH. In order for the syntax of thought to resemble that of English, the constituent ‘is curved’ would have to be an inflectional phrase (IP) composed of ‘is’ and ‘curved’, the former bearing number-agreement relations to the subject. As we will see below, this is the view that Devitt defends.

representations in which they figure. Moreover, the atomic constituents can appear in an indefinite number of complex representations, each time contributing their individual contents to the total content of the complex state. Viewed in this way, cognition consists in sequences of neurally inscribed sentences of “Mentalese”. According to Fodor’s Computational Theory of Mind (CTM), the formal properties of such sentences determine the course of cognitive processing, and their semantic properties are implicated in psychological laws and explanations.

### §3.1.2 *The Standard Arguments for LOTH*

Fodor (1975, 1987) presents a number of arguments in favor of LOTH. We look first at the argument that has gained the most attention from philosophers and psychologists—the argument from systematicity.<sup>30</sup> This argument begins with the observation that, as a matter of contingent empirical fact, our competence in processing natural language expressions exhibits a kind of *systematicity*. Roughly, the idea is that our ability to produce and comprehend certain sentences is *inherently* or *intrinsically* tied to our ability to produce and comprehend others.<sup>31</sup> Fodor takes the systematicity of natural languages as a theory-neutral datum, and argues that the most natural explanation of it is that such languages exhibit another feature, known as

---

<sup>30</sup> This argument is especially significant because of the role it plays in Fodor’s ongoing critique of the connectionist approach to modeling cognition. (See Fodor & Pylyshyn, 1988; Fodor & McLaughlin, 1990; Fodor, 1997.) Systems that have a “classical” architecture—i.e. the sort of architecture characteristic of Turing machines and von Neumann systems—make essential use of syntactically structured representations, like the ones proposed by the LOTH theorist. Fodor has argued that connectionist models cannot make use of such representations without thereby implementing a classical model of computational cognition. Though his argument for that claim has been challenged in a number of ways, (Bechtel and Abrahamsen, 2002; Chalmers 1990, 1993; Cummins, 1996b; Smolensky, 1988, 1995) it remains influential in philosophical psychology (Hadley, 1994; Frank, Mathis, and Badecker, 2008). In any event, were it to be established, the truth of LOTH may go a long way toward securing the supremacy of the classical approach.

<sup>31</sup> As Johnson (2004) points out, it is important in this context to distinguish between the systematicity of a representational system—e.g., natural language—and the systematicity of a creature’s processing competence with respect to that system. Fodor elides this distinction, in effect equating a language with competence in that language. As we saw in chs. I and II, this equation is characteristic of the cognitive conception of language, popularized by Chomsky. Devitt (2006) argues against this conception. In what follows, I present the argument in Fodor’s own terms. The reader is invited to make whatever adjustments are needed in order to accommodate Devitt’s position.



*compositionality*. Compositionality is a property of a representational scheme just in case that scheme supports a distinction between atomic and complex representations, and imposes part-whole relations on those representations. Hence, for compositionality to hold of a representational scheme is for all complex expressions in that scheme to be constructed from atomic expressions, which are their constituents. Semantic and syntactic compositionality, in particular, require that the contents of the atomic constituents, as well as their syntactic relations, determine the contents of the complex states in which they figure. Moreover, the atomic expressions must be semantically evaluable, and able to appear in an indefinite number of complex expressions, each time contributing their individual contents to the total contents of the complex expressions. It is plain that compositionality is the essence of LOTH.<sup>32</sup>

If compositionality is a feature of natural languages, then it is easy to explain why anyone who is capable of comprehending certain sentences, e.g. ‘Newton admired Leibniz’, is very likely to be able to comprehend certain others, e.g. ‘Leibniz admired Newton’. In other words, compositionality makes it possible to explain why natural languages exhibit systematicity. The explanation is that competent speaker/hearers have access to a stock of primitive (i.e. atomic) expressions, which they can assemble or arrange in a variety of ways, in order to yield more

---

<sup>32</sup> Indeed, the claim that thought (or, more broadly, cognition) is compositional may be *logically equivalent* to LOTH. Whether the claims are logically equivalent depends, on the one hand, on whether or not certain picture-based representational schemes satisfy the compositionality constraints, and, on the other hand, on whether or not such schemes are genuine languages, in the sense that LOTH requires. For a discussion of these issues, see Cummins (1996a: ch. 6; 1996b). Aydede (2010) denies that compositionality is essential to LOTH. He writes: “Although I described the line above as official and presented it as requiring a compositional semantics, and although almost all the defenders of LOTH conceive of it in this way because they think that is what empirical facts about thought and language demand, nevertheless it is perhaps important to be pedantic about exactly what LOTH is minimally committed to. Minimally, it is *not* committed to regarding the internal code as having a compositional semantics, namely a semantics where the meaning of complex sentences are determined by the meanings of its constituents together with their syntax; this, in effect, requires that the atomic expressions always make (approximately) the same semantic contributions to the whole of which they are constituents (idioms excepted). But strictly speaking LOTH can live without having a strictly compositional semantics if it turns out that there are other ways of explaining those empirical facts about the mind to which I will come below. Admittedly, in such a case LOTH would lose some portion of its appeal and interest. But even if this scenario turns out to be the case, there are still a lot of facts for LOTH to explain. Having said this, however, I will simply forget it in what follows” (fn. 18). I must confess that I have not been able to make sense of this passage, though the fault may very be my own.

complex expressions. The capacity to perform combinatorial operations on the atomic elements of the language, e.g. ‘Newton’, ‘Leibniz’ and ‘admire’, allows a competent speaker to form and comprehend a number of grammatical strings, among which one finds the target sentences. Hence, compositionality seems to provide the inherent or intrinsic connection between our ability to understand pairs like ‘Newton admired Leibniz’ and ‘Leibniz admired Newton’.

Having argued for the compositionality of public language, Fodor turns his attention to cognition. He reasons that if the medium of cognition is systematic in the same way as natural language, then complex mental states must *also* have a constituent structure and a combinatorial semantics—precisely the Language of Thought Hypothesis. Since Fodor assumes that no other explanation for the putative systematicity is available, all that remains to be shown is that cognition really *is* systematic. What follows is Fodor’s argument for this claim.

- i. One of the main functions of language is the communication—i.e. the expression and comprehension of thoughts.
- ii. In order to be able to express and understand the sentence ‘Newton admired Leibniz’ one must be able to have the thought that is standardly expressed using that sentence; same for ‘Leibniz admired Newton’.
- iii. Everyone who understands the first sentence understands the second.
- iv. Everyone who understands both sentences can think the thoughts that each of the sentences expresses.
- v. Everyone who understands both sentences can think both thoughts.

Premise (i) is taken for granted by just about every theorist who has ever considered these issues.

Premise (ii) follows trivially from (i). Fodor regards (iii) as true on purely empirical grounds, and to my knowledge no one has ever challenged it. Premise (iv) follows from premises (i) and (ii), and (v) follows from (iii) and (iv). Moreover, the argument is perfectly general. If it is sound, then it can be used to show that for *any* pair of sentences whose propositional contents are systematic variants of one another, everyone who understands both of the sentences can think

both of the thoughts they express.<sup>33</sup> Thus, cognition would seem to be at least as systematic as natural language.

We now have before us Fodor's master argument for the compositionality of thought and, hence, for LOTH. To recap, Fodor begins by assuming that natural languages are systematic, and arguing that they must also be compositional. He then argues that in order to speak and comprehend a systematic and compositional language, we must have compositional thoughts.

LOTH remains a contentious hypothesis. All of Fodor's arguments for it have been challenged. For instance, Johnson (2004) argues that public languages do not exhibit systematicity. Johnson's argument rests on the following definition: A language *L* is systematic *iff* condition (S) holds for all *A*.

- (S) "*A* is a constituent of the language only if for all *B* of the same linguistic kind as *A*, and all things *C*, *C* can compose with *A* (in a certain way) to form a grammatical sentence if and only if *C* can compose with *B* (in that same way) to form a grammatical sentence." (p. 14)

Johnson points out that linguistic natural kinds cut across each other at different levels of analysis (phonology, syntax, and semantics). From this, he argues, it follows that we can find contexts *C*, such that for two members *x* and *y* of a single linguistic kind *K*, *x* can compose with *C* to make a grammatical sentence, but *y* can't. He concludes that natural language is, therefore, *not* systematic.<sup>34</sup>

Fodor's other arguments for LOTH are also open to substantive challenge. Consider his argument from productivity, which takes the same basic shape as the systematicity argument.

---

<sup>33</sup> The notion of a *systematic variant* is due to Cummins (1996), whose formal characterization is as follows: content *c*\* is a systematic variant of content *c* if *c*\* is a content that you can express by permuting the constituents of the expression whose content is *c*. Fodor (personal communication) does not endorse Cummins' definition, but does not provide an alternative. For a useful discussion of these issues, see Pullum and Scholtz (2007).

<sup>34</sup> Fodor (p.c.) rejects Johnson's formal definition of systematicity, but does not provide an alternative. He also maintains that *any* amount of systematicity cries out for explanation. His opponents would presumably deny this.

Fodor observes that public language is productive—i.e., that every natural language constitutes an infinite domain.<sup>35</sup> On the basis of the tight relationship between language and thought, Fodor argues that thought, too, is productive. And, as with systematicity, the only plausible explanation seems to be that thoughts have a compositional syntax and semantics. This underwrites another abductive argument for LOTH. Fodor concedes, however, that this argument is weaker than the argument from systematicity, as it relies on an idealization that an opponent of LOTH has no reason to honor. While this is a notable difficulty, there is, I think, an even bigger problem with the productivity argument.

Fodor's claim that public language is productive is grounded in the fact that an optimal grammar for any natural language is recursive. The possibility of embedding sentences within sentences, prepositional phrases within prepositional phrases, and so on, gives rise to a denumerable infinity of sentences (ch. I). However, despite the tight relationship between thought and language, it simply does *not* follow that *thought* is productive. For, it is quite possible that the optimal grammar for Mentalese is *not* recursive. For instance, it may turn out that we comprehend sentences by tokening thoughts that contain no embedding at all. For instance, upon encountering a sentence with a relative clause, e.g., (25), a hearer might token two distinct thoughts, (26) and (27).<sup>36</sup>

- (25) The boy who loves the girl is sick.
- (26) THE BOY LOVES THE GIRL.
- (27) THE BOY IS SICK.

---

<sup>35</sup> See also Pullum and Scholtz (forthcoming).

<sup>36</sup> For clarity of exposition, I am postponing discussion of difficult issues regarding the outputs of the comprehension process. Plainly, we do not literally think—i.e., judge, believe, or *endorse*—everything that we hear. Accommodating this point requires some care. I will return to the issue later in this chapter.

Other types of embedding may receive a similar treatment. Consider embedded prepositions.

An input sentence like (28) may give rise to two distinct thoughts, (29) and (30).

(28) The toy is in the box in the attic.

(29) THE TOY IS IN THE BOX.

(30) THE BOX IS IN THE ATTIC.

What this shows is that, unlike the syntax of public language, the syntax of thought may be non-recursive. Unless this possibility can be ruled out—and I do not believe that it has been—the inference from the productivity of language to the productivity of thought is unwarranted.

Devitt (2006: §8.1) endorses Fodor's central doctrines—Intentional Realism and RTM. He is, however, more hesitant about LOTH, recognizing explicitly that the case for it is far from airtight—indeed, downright shaky. Still, Devitt is clearly attracted to LOTH. He rehearses the standard arguments for it and, in the following passage, endorses Fodor's argument from inferential coherence.

I think that we should prefer LOTH [over the view that] a mental representation has some nonlinguistic complexity—for example, the complexity of a map. Perhaps [such a view] could meet the requirement of Fodor's first methodological argument<sup>37</sup> and could account for the systematicity that is the concern of his third argument, but it is difficult to see how it could account for the mental process of thinking that is the concern of his second argument. Formal logic gives us a very good idea of how thinking might proceed if thoughts are represented linguistically. From its very beginning, computer science has used this idea to build machines that do process linguistic representations. In recent years, computer science has developed "connectionist" machines that use representations of a very different sort, if they use representations at all. Despite the striking success of these machines with some forms of problem solving, connectionist processes seem rather far from capturing anything like human inference. We still have very little idea how thinking could proceed if thoughts were not language-like. (Devitt, 2006: 146-147)

---

<sup>37</sup> For reasons of space, I have omitted a detailed discussion Fodor's "methodological argument" above. Briefly, the argument rests on a methodological principle that Fodor calls "Principle P", which states (roughly) that effects structured from constituents have causes that are also structured from constituents. This principle has come in for a fair share of criticism. For a convincing argument to the effect that Principle P is not a general principle of explanation, see Egan, F. (1991). "Propositional Attitudes and the Language of Thought," *Canadian Journal of Philosophy*, 21, 3, pp. 379-88. For a less convincing argument, see Clapin, H. (1997). "Problems with Principle P," *Pacific Philosophical Quarterly*, 78, pp. 261-77.

This line of reasoning, like the other arguments for LOTH that we considered above, has been challenged from a number of directions. First, as Fodor (1983, 2000) recognizes, the computational theory of mental processing may be applicable only to modular systems. When applied to the non-modular, “Quinean,” and “isotropic” processes of practical and theoretical reasoning—the domain of what Devitt, following Fodor, calls the “central processor”—the theory quickly runs into seemingly insuperable difficulties. These may be loosely collected under the label “the frame problem.” The widely publicized failure of efforts in what Haugeland calls “Good Old-Fashioned AI” is, I think, a testament to the truth of “Fodor’s First Law”

I should like to propose a generalization; one which I fondly hope will some day come to be known as ‘Fodor’s First Law of the Nonexistence of Cognitive Science’. It goes like this: the more global (e.g. the more isotropic) a cognitive process is, the less anybody understands it. (J. A. Fodor, 1983: p. 107).

Second, Harman (1973, 1986, 2002) has argued—to my mind, persuasively—that there probably is no such thing as “deductive inference.” If that is so, then Devitt is wrong to say that “[f]ormal logic gives us a very good idea of how thinking might proceed if thoughts are represented linguistically.” Harman’s point is that logic gives us *very little* in the way of a theory of the psychological process of inference. Reasoning, he suggests, may very well be abductive *all the way down*—a process that we have barely begun to understand, despite our impressive advances in formal logic. Harman’s position is, I think, an echo of a view that was first articulated in Sellars (1953)—now popularized by Robert Brandom<sup>38</sup>—according to which the most basic and pervasive kind of reasoning is *not* constructed from “formal” inferences—i.e., ones that are sensitive to logical form—but, rather, from what Sellars calls “material” inferences, such as “Lightning now, so thunder soon,” or “This strawberry is red, so it is ripe.” I do not have

---

<sup>38</sup> See Brandom (1994), pp. 94-108.

the space here to elaborate the worries that Fodor, Harman, and Sellars present. Suffice it to say that the argument from inferential coherence cannot, at present, be said to fare better than the other standard arguments for LOTH. All in all, we have not, I think, been given any overwhelming reason to adopt LOTH. This should make us doubly suspicious of any claim that builds on LOTH, such as the Strengthened Language of Thought Hypothesis (SLOTH), or the Strong Brute-Causal model of language processing that Devitt seeks to establish.

### §3.1.3 *The Public LOTH vs. the Strengthened LOTH*

For present purposes, the paramount significance of LOTH consists in the role that it plays in Devitt's discussion of the brute-causal model of language processing. Devitt sees LOTH as providing support for this model.

In considering the brute-causal [model], I shall assume the Language-of-Thought Hypothesis (LOTH) and [that] the structure rules of a speaker's language are similar to the structure rules of her thought. Some of the case for the [brute-causal model] may survive if we assume only RTM, the minimal theory of thoughts that allows that thoughts might be map-like. But the [brute-causal model] certainly seems most plausible assuming LOTH *because that assumption lessens what has to be done in language processing*. So the case for the [brute-causal model] rests a bit on LOTH. (p. 223; emphasis added)

It is important to note that the assumptions Devitt lays out in this passage go well beyond the standard construal of LOTH. Ordinarily, proponents of LOTH take sentences of Mentalese to have *some structure or other*—a syntax that is as-yet-unknown, awaiting discovery by means of empirical (presumably neurocognitive) research. Devitt, by contrast, takes the Mentalese sentences of, say, an English speaker's thoughts to have roughly the same syntactic structure as the English sentences that express those thoughts. He writes:

What I think we can conclude ... is that if LOTH is correct, the speaker syntax of public-language sentences will be the same as, or similar to, the syntax of the mental sentences they express. So let us be cautious and say that they are similar. (p. 157)

With this in mind, let us distinguish between the standard version of LOTH and Devitt's stronger

version by introducing the label ‘SLOTH’ to stand for the *Strengthened* Language of Thought Hypothesis. We will examine Devitt’s arguments for SLOTH momentarily. Before doing so, however, let us briefly contrast SLOTH with a closely related view.

In §9.4 of *Ignorance*, Devitt discusses a view that he calls “the Public Language of Thought Hypothesis” (henceforth, PLOTH). According to PLOTH, a speaker’s language of thought *just is* her public language. French speakers think in French, English speakers think in English, and so forth. (Presumably, polyglots have the option of thinking in any of their languages.) The difference between SLOTH and PLOTH is this: SLOTH is committed to there being significant similarities between a speaker’s public language and her language of thought *in respect of syntax and semantics*. PLOTH goes further, requiring that the similarities extend to “lower” levels of analysis. (I provide examples below.) Devitt does not accept PLOTH, for he finds obscure the idea that the *vehicles* of thought belong to, say, French rather than English. On this point, Devitt is surely correct. Let us extend his remarks on the issue.

Expressions in spoken language have phonological properties; in written language, orthographic properties. A neural language presumably has some analogue of phonology and orthography, in the sense that its expressions have properties that are *neither* brute-physical/formal *nor* morpho-syntactic, but something “in between.” Let us coin the term ‘neuro-phonology’, to label the level at which such properties appear.<sup>39</sup> The neuro-phonological properties of Mentalese sentences are described in what Kirsch (1990) calls the “notational component” of our theory of a language, which

---

<sup>39</sup> I avoid ‘neuro-orthography’ because, unlike the phonology of any natural language, its orthography is less “natural,” in the sense that it is partially a product of deliberate design and maintained by explicit social conventions. Moreover, orthographic competence is acquired separately from the rest of the language, often painstakingly, and only under conditions of explicit instruction. (See Fernandez and Cairns, 2011: p. 93-4) None of this can be true of thought, which is a *natural* system of representation, acquired spontaneously.



tells us what to accept as an allowable variation in the structure of a token. It is the theory of symbol separation. So, for example, although my writing changes from page to page as I change my posture, a good notational component would specify enough variability in tokens to cluster my written words into correct equivalence classes. It gives the identity conditions of atomic symbols. (p. 352)

With this in mind, we can ask how one should go about deciding whether to count the neuro-phonological properties of thought vehicles as belonging to, e.g., French rather than English. At first blush, the answer may seem to depend on whether the thoughts in question have the syntactic and semantic properties that are characteristic of French or English sentences. After all, we often take a piece of speech or writing to have the phonology and orthography of French when it also has French syntax and semantics. If this were so, then there would be no real difference between PLOTH and SLOTH. But this is not correct in the general case. French speakers who acquire English late in life may very well employ a great deal of French phonology in their English speech, giving rise to a distinctive accent. Phonology thus comes apart from syntax and semantics.

Even more obscure is the claim that the *orthography* of, say, English writing is somehow mirrored in the neuro-phonology of an English-speaker's thought. For, although English-speakers have a relatively standard way of encoding their thoughts orthographically, there is, in truth, no unique thing that we can call *the* orthography of English. I do not have in mind superficial differences, such as those that distinguish fonts. Rather, my claim is based on the possibility of writing, e.g., a *Russian* sentence, like (31), using English letters, as in (32).

(31) Я пошел домой. (*I went home.*)

(32) Ya poshol domoi.

Plainly, the reverse can also be done—a sentence of English rendered, however ineptly, in Russian. Moreover, the sentences of *either* language can be rendered in the International Phonetic Alphabet—a notational system that, by its very design, is *not* tied to the syntax or

semantics of any particular natural language.

Making sense of PLOTH is complicated even further by the following considerations:

- Of the natural languages that have writing systems, each one has at least one characteristic mapping from orthography to phonology, which can (and often does) fail to be transparent. (English, for instance, is notoriously opaque, as the famous *ghoti/fish* example illustrates.)
- Representations of phonological properties always enter into the processing of written language, and representations of orthographic properties sometimes enter into the processing of spoken language. (See Fernandez and Cairns, 2011: ch. 7.)
- There is a distinction between phonetics and phonology, the latter dealing far less directly with the articulation of the speech signal than the former.
- While phonology was traditionally thought of as being concerned with something akin to the acoustic properties of spoken linguistic expressions, modern-day theorists treat the *non*-acoustic medium of sign language as *also* having phonological features.<sup>40</sup>

It seems to me that these difficulties scotch any hope of making clear sense of PLOTH. We simply have no grip on the proposal that the neuro-phonological properties of the vehicles of thought are similar, let alone identical, to the phonological or orthographic properties of public-language expressions. For this reason, I follow Devitt's suggestion that we "set [this view] aside, ... treat the language of thought as if it were a distinct Mentalese, and focus instead on what seems a more theoretically productive issue: How similar is a person's Mentalese to her public language? In particular, how syntactically similar is it?" (Devitt, 2006: p. 152) In other words, we should put aside PLOTH and focus instead on determining whether SLOTH is true.

---

<sup>40</sup> See Stokoe, W. C., (1960/1978). *Sign Language Structure*. Silver Spring: Linstok Press

Before turning to Devitt's argument for SLOTH, I want to briefly note an important point that Devitt brings out in his discussion of syntactic properties. These are often thought of as being constituted by the syntactic categories of their constituent parts and the hierarchical relations that hold between those constituents. But syntactic properties are also often characterized as the "shapes" of such states—i.e., their *formal* properties, which determine their computational roles. Devitt (2006) points out that these two characterizations are contradictory.

[I]t is important to distinguish a representation's brute-physical formal properties from its syntactic properties. Whereas the formal properties are intrinsic to the representation, the latter are relational. A formal property of the inscription 'Tex likes exciting sheep', in the intended brute-physical sense of 'formal', is that of beginning with an inscription shaped such and such (replace 'such and such' with a description of the shape of 'Tex'). A formal relation between 'Tex likes exciting sheep' and 'Tex loves Mary-Lou' is that of both beginning with an inscription of the same shape. A formal property of a symbol in a computer is that of being a certain pattern of on-off switches. A formal property of a symbol in the brain is that of being a certain array of neurons.<sup>[41]</sup> Manifestly, the written token of ['Tex likes exciting sheep'] above shares very few formal properties with any spoken token of ['Tex likes exciting sheep']. Similarly, it shares very few with a mental token that ['Tex likes exciting sheep'] expresses. ... Syntactic properties and relations are ones that bear on the construction of sentences in a language. A syntactic property of 'Tex' is that of being a noun; of 'likes', that of being a verb; of 'Tex likes exciting sheep', that of being a sentence. A syntactic relation between one interpretation of that sentence and the sentence, 'Exciting sheep are liked by Tex', is that of the latter sentence being the passive of the former. Syntactic properties are ones that reflect a token's relations to other tokens in the language; they are functional properties and extrinsic to the representation. Although formally so different, a written and spoken token of ['Tex likes exciting sheep'] might share all their syntactic properties. Sentences that "look different" can be syntactically alike. Formal differences are one thing, syntactic differences, another.

---

<sup>41</sup> Devitt's claim here is somewhat loose. In all likelihood, the mere existence of an array of neurons does not, by itself, constitute a symbol in the brain. Rather, mental representations are realized by the *activations* of an array of neurons, or the weights of the connections between them. Needless to say, alternative possibilities are conceivable. For an argument to the effect that memories are encoded in sub-neuronal *molecular* processes, see Gallistel, C.R. and King, A. P., (2010). *Memory and the Computational Brain: Why Cognitive Science will Transform Neuroscience*, Wiley-Blackwell.

Devitt's distinction between formal and syntactic properties is important to keep in mind, both for the purposes of the present discussion, but also in connection with broader issues.<sup>42</sup> While we have the distinction firmly in mind, I would like to make use of it in clarifying the relationship between two types of syntactic property—linear precedence structure and hierarchical dominance structure.

The properties that Devitt classifies as formal are characteristically superficial, having to do with the specific physical medium in which information is encoded—e.g., the acoustic and phonetic properties of speech sounds, the shapes and orthographic properties of inscriptions, and the textures of Braille dots. I argued above that no sense can be made of the claim that thoughts have the same superficial formal properties as public-language expressions. What about properties that are not quite so superficial, but still not as abstract as hierarchical structure? I have in mind properties that are defined solely in terms of the precedence relations that hold between the words or morphemes that make up a public-language expression—i.e., the linear order of items in a string. I will refer to this as the string's *precedence structure*.

Precedence structure can be shared by speech, writing, Braille, Morse code, sign language, and other media. For instance, the word 'dog' precedes the word 'man' in the string *The dog bit the man*, regardless of whether the string is spoken, written, or signed. In a word, precedence structure is *medium-neutral*. In speech and sign language, it is encoded by the temporal order of acoustic blasts and hand gestures; in writing, by the spatial order—left-to-right

---

<sup>42</sup> I believe that this distinction carries as-yet-unexplored consequences for the debate surrounding the doctrine of meaning holism. Briefly, suppose that an expression's syntactic properties are constituted by the relations that it bears to all of the other expressions in its language. Now combine this with the further claim that the semantic properties of an expression are partially constituted by its syntactic properties. From these two premises, we can derive the following: *The semantic properties of an expression are partially constituted by the relations that it bears to all of the other expressions in its language.* This is certainly a kind of meaning holism, though perhaps not the standard kind. Needless to say, this argument requires more careful discussion, which would take us beyond the scope of our principal concerns. See Pereplyotchik (in progress) for a fuller treatment of this issue.

in English, right-to-left in Hebrew, and top-to-bottom in Chinese. (Notice that precedence structure is relational, i.e., defined in terms of the string's membership in a particular language.)

The medium-neutrality of precedence structure suggests that it is not a superficial formal property, but a *syntactic* one, alongside dominance, scope, and various other hierarchical relations. Nevertheless, there are important differences between precedence and dominance. The hierarchical syntactic properties of a public-language expression—e.g., its logical form—are, in general, *not* a function of its precedence structure. This is especially easy to see in morphologically rich languages that have a relatively free word order, e.g., Russian, Turkish, and Warlpiri. But, even in morphologically poor languages like English, which encode quite a bit of information in the linear order, it is equally clear, upon reflection, that the scopes of quantifiers, wh-phrases, negation markers, tense markers, and adverbs are not a function of linear order. (This is what causes much of the trouble for models like the ones we explored in §2.) We can put this uncontroversial point as follows: In public language, precedence structure does not, in general, “track,” “reflect,” or “encode” scope relations or other elements of logical form.

Although taking thoughts to have the same superficial formal properties as public-language expressions is not a serious option—the claim is simply too obscure—there is a perfectly legitimate question about whether thoughts have the same (or even a similar) precedence structure as public-language expressions. For, as noted above, this property can be shared by messages in vastly different media—so why not a neural medium? There is, of course, a prior question as to whether thoughts even *have* a precedence structure, in the relevant sense. Not all complex representational systems do. For instance, in his discussion of the tensor-product representations in connectionist networks, van Gelder (1990) points out that these representations exhibit hierarchical syntactic structure, *without concatenating component*

*symbols* either spatially or temporally. So, while it is plain that neural events are spatially and temporally related to one another, it is not at all clear that these relations have any systematic bearing on their hierarchical syntactic properties (if they have any), or their semantic properties.

Let us suppose for a moment that thoughts do have precedence structure—i.e., some neural analogue of the temporal relations between speech sounds or the spatial relations between inscriptions. If the precedence structure of thoughts is the same as the precedence structure of public-language expressions, then it will fail to track the hierarchical relations that are constitutive of logical form. On the other hand, the precedence structure of thoughts might be different from that of public-language expressions. For instance, it might be more like the precedence structure of the LF tree-diagrams that one finds in today’s linguistics textbooks, which, in a certain sense, make logical and syntactic properties explicit.

I do not know whether we can, at present, settle the issue of whether thoughts have the same precedence structure as public-language expressions, nor even whether they have any precedence structure at all. Perhaps future neurophysiological findings will shed light on this issue. The most we can say right now is this: If thoughts *do* have a precedence structure and it differs from that of public-language expressions, then we will need a substantive account of *how*, in language comprehension, the HSPM manages to compute a thought from a sentence.

#### §3.1.4 Devitt’s Argument for SLOTH

Recognizing Devitt’s commitment to SLOTH is, in my view, absolutely crucial to interpreting much of what he says in favor of the brute-causal model of language processing. His argument for SLOTH emerges in a discussion of the example in (33)-(35). I quote at length.

(33) Tex likes exciting sheep

(34) [<sub>S</sub>[<sub>NP</sub>[<sub>N</sub> Tex]] [<sub>VP</sub>[<sub>V</sub> likes] [<sub>VP</sub>[<sub>V</sub> exciting] [<sub>N</sub> sheep]]]]

(35) [<sub>S</sub>[<sub>NP</sub>[<sub>N</sub> Tex]] [<sub>VP</sub>[<sub>V</sub> likes] [<sub>NP</sub>[<sub>A</sub> exciting] [<sub>N</sub> sheep]]]]

[I]n virtue of what does a sentence have its syntactic structure, whether explicit or not? What makes it the case that a particular token, for example a token of [33], has the structure it has, perhaps the structure revealed by [34] or (35)? This question does not seek a description of the structure that the sentence has—a description that linguists have been so successful at providing—but rather an explanation of its having that structure. Since this structure along with the sentence’s word meanings determines the sentence’s meaning, this explanation is part of an explanation of the sentence’s meaning. It is part of an explanation of in virtue of what the sentence has whatever meaning it has. ... One of our earlier priority claim points the way to an answer. This was the *Gricean claim that thought meanings explain speaker meanings*; and speaker meanings explain conventional meanings. In this way, *thought is prior to language*. We now wed this Gricean story to LOTH, the hypothesis that is central to the case for [SLOTH]. So we should answer questions about public language meanings by first addressing questions about Mentalese meanings. The meaning of a mental sentence is determined by the meanings of its words and the syntactic structure that contains them. ... In virtue of what does a mental sentence have its syntax? Presumably, in virtue of the way in which its meaning depends on the meanings of its parts and in virtue of the structure’s role in determining a sentence’s possible inferential interactions with other sentences. Thus, the mental sentences <Reagan is wrinkled> and <Thatcher is tough> share a syntactic structure in virtue of a similarity in these respects. Similarly, <All politicians are rich> and <All police are corrupt>, which share a different structure in virtue of their similarities. In light of this, in virtue of what does a public-language sentence have its meaning? There are two steps to the answer, reflecting the distinction between the speaker meaning of a sentence on an occasion of utterance and its literal or conventional meaning on that occasion. (i) The speaker meaning of a sentence on an occasion of utterance must be determined by the causal story of its production. Thus the speaker meaning of a word in that sentence is determined by the meaning of the mental word it expresses on that occasion. The “speaker syntax” of a sentence on an occasion of utterance is determined by the structure of the underlying mental sentence that the utterance expresses on that occasion and by the way in which the utterance was produced from that mental sentence. (ii) The conventional meaning of a sentence on an occasion of utterance—what it conventionally means in a language given the context—is explained somehow in terms of regularities in speaker meanings: Speakers of the language regularly use one linguistic word to express one mental word, another, another; they regularly use linguistic sentences with one speaker syntax to express one mental syntactic structure, another, another. Wedding the Gricean story of the general explanatory priority of thought over language to LOTH has yielded a sketch of an explanation of the syntactic properties of public-language sentences. (pp. 155-156, emphases added)

In short form, Devitt’s abductive argument for SLOTH is this: If SLOTH were true, then it would provide us with a satisfying explanation of why public-language expressions have the syntactic and semantic properties that they do.

How did we reach these conclusions? First, we adopted the controversial LOTH. Second, we noted, what seems scarcely deniable, that public language sentences can have syntactic properties that are not explicit. Third, we pointed to the role of mental sentences in explaining the implicit and explicit syntactic properties of the public sentences that express them. These three assumptions seem to make [SLOTH] inevitable. The importance of LOTH to this argument is enormous. (*Ignorance of Language*, p. 157)

There are, in my view, several related problems with this line of argument. First, the argument for SLOTH relies on the truth of LOTH. And, as we saw above (and as Devitt is fully aware), the arguments for LOTH are tenuous at best. Still, even if LOTH is true, I believe that the argument for SLOTH faces further difficulties.

To begin with, there is reason to be skeptical of the Gricean view that thought is explanatorily prior to language. I have argued against this view elsewhere and will not do so again here.<sup>43</sup> Suffice it to say that the uncontroversial priorities of thought over speech—i.e., the historical, ontogenetic, and causal priorities—are fully compatible with there being no *explanatory* priority of either over the other. This view has been ably defended by Wilfrid Sellars, in his famous debate with Roderick Chisholm.<sup>44</sup> (Chisholm championed what Devitt calls the Gricean view.) There is an unfortunate lack of consensus about the outcome of this debate, but I believe that Sellars emerges victorious. Rosenthal (1968, 1986/2005) examines the Chisholm-Sellars correspondence at length and reinforces Sellars' points by casting doubt on the claim that public-language expressions—speech acts, in his terminology—somehow *derive* their linguistic properties (both syntactic and semantic) from the analogous properties of thoughts. Rosenthal also argues against several ways of drawing the putative distinction between intrinsic and derived content as well as the closely related idea that thoughts “confer” their syntactic and

---

<sup>43</sup> Pereplyotchik, D. (2008). “Language Before Thought?” (MS)

<sup>44</sup> “The Chisholm-Sellars Correspondence on Intentionality,” in *Minnesota Studies in the Philosophy of Science*, II, ed. H. Feigl, M. Scriven, and G. Maxwell, Minneapolis: University of Minnesota Press, 1958, pp. 521–539.



semantic properties onto the speech acts that express them. I take Rosenthal's arguments to be decisive. (See also Dennett, 1987: ch. 8.) Plainly, though, Devitt remains unconvinced. This is not the place to resolve the issue. See Pereplyotchik (2008) for detailed discussion.

A second, though related problem with Devitt's argument is that it is inconsistent with what he says earlier in *Ignorance* about why public-language expressions have the syntactic properties that they do. In two separate passages, he endorses the structuralist view that a public-language expression has its syntactic properties in virtue of its relations *to other public-language expressions*.

The outputs of a linguistic competence, physical sentence tokens, are governed by a system of rules.... Something counts as a sentence only if it has a place in the linguistic structure defined by these structure rules. Something counts as a particular sentence, has its particular syntactic structure, in virtue of the particular structure rules that govern it, in virtue of its particular place in the linguistic structure. (*Ignorance*, p. 24)

A syntactic relation between one interpretation of ['Tex likes exciting sheep'] and the sentence, 'Exciting sheep are liked by Tex', is that of the latter sentence being the passive of the former. *Syntactic properties are ones that reflect a token's relations to other tokens in the language.* (p. 155, emphasis added)

This, incidentally, is very much akin to the picture that Sellars and Rosenthal endorse. On their decidedly un-Gricean view, public-language expressions—whether speech act tokens or sentence types—have their syntactic and semantic properties in virtue of their relations to other public-language expressions, as well as to a person's other actions, and to her environment. Rosenthal (p.c.) thinks of these relations as constituting a “restricted functional role”—restricted in the sense that it does not take into account the relations that speech acts have to thoughts. On his view, taking these relations into account would yield a *full-blown* functional role, which would *sharpen* the semantic properties of public-language expressions, by exhibiting their inferential

relations to a much broader set of contentful entities.<sup>45</sup> But sharpening properties is not the same as constituting them. Rosenthal's view is that the public-language expressions (i.e., speech acts) can be said to have semantic and syntactic properties *even if we put aside their relations to mental phenomena*. These properties, again, are constituted by the relations that public-language expressions bear to each other—much as Devitt's remarks suggest—as well as to the environment and to nonverbal behavior.

Needless to say, this view does not entail that speech acts can take place in the *absence* of thought, or that speech is *ontogenetically* or *historically* prior to thought. As noted above, Sellars and Rosenthal accept these uncontroversial priorities. That is, they accept that thoughts always precede and cause speech acts and that *only* thoughts can cause speech acts, so that if there were no thoughts, there would be no speech acts. They also accept that non-linguistic animals—including birds, dolphins, canines, primates, early humans, and prelinguistic babies—have thoughts. Their target is solely the *explanatory* priority that Devitt uses to argue for SLOTH. In contrast with the intuitively obvious priorities mentioned above, Sellars and Rosenthal argue that have no reason to believe in the kind of priority that Chisholm, Grice, and Searle have popularized. That is, we have no grounds for the claim that speech acts somehow “derive” their semantic and syntactic properties from thoughts, or that thoughts somehow “confer” these properties onto them via a “special kind of causation” (Devitt, p.c.).

In place of this explanatory priority thesis, Sellars and Rosenthal urge that neither speech nor thought has explanatory priority over the other. The intentional content of each can be *illuminated* in terms of the other—again, by enriching the set of causal and inferential relations

---

<sup>45</sup> As may be obvious from the terminology, Rosenthal is working with a “long-arm” or “wide” functional-role semantics, which treats an item's inferential role as a major determiner of its overall intentional content. (Connections with behavior and the environment are the other major determiners.) Devitt (1996) rejects the extreme holism that Rosenthal endorses. But, as far as I can see, Devitt's molecularist position is compatible with the core of Rosenthal's take on the priority issue. Thus, in *this* context, the issue of holism is a red herring.

that are constitutive of intentional content. And if both are found (on independent grounds) to have *syntactic* properties, then the same applies to those properties as well. But it does not follow that the syntactic properties of thought—if there be such—would be either similar or identical to those of public language. “We tend even to attribute to our thoughts the syntactic structure of the speech acts that express them, though it is plain on reflection that we have no reason to suppose that thoughts have much of the complex compositional structure that sentences exhibit” (Rosenthal, 2005: p. 98).

Given Devitt’s commitment to the explanatory priority of thought over language, he will surely seek to reinterpret the structuralist passages from *Ignorance* that I quoted above. And it may well be that there is a way to square his remarks in those passages with his commitment to the explanatory priority of thought over language. My point is merely that an alternative view, which rejects the explanatory priority thesis, is perfectly in line with those passages as they stand, and, arguably, preferable to Devitt’s Gricean position.<sup>46</sup>

The third and, I think, most serious problem with Devitt’s argument for SLOTH is that it is inconsistent with a central aspect of the brute-causal model of comprehension that he endorses. As we will see in §3.3, a large part of what Devitt means when he describes a process as ‘brute-causal’ is that the process is *not sensitive to the syntactic structure of the representations that are involved*. Moreover, Devitt takes the connectionist network to be a good example of how the

---

<sup>46</sup> As Devitt (2006) admits, the Gricean position faces a serious difficulty: “Gricean explanations of speaker meanings are seriously incomplete because they rest on unexplained thought contents: the speaker means the thought content that he intends to convey but no account is given of the nature of that content” (p. 133). The Sellars-Rosenthal position does not share this problem. On that view, thoughts have their intentional contents in virtue of their functional roles—their inferential relations to one another, as well as to the environment and to behavior (including, of course, verbal behavior). Needless to say, the Gricean can appeal to any of the extant theories of intentional content. It is not clear that any of these is satisfactory. The Gricean can, of course, also adopt the functional-role explanation. But then it’s not clear why the very same explanation cannot be given for the intentional contents of speech acts. See Rosenthal (1986/2005) for a forceful statement of this point.

brute-causal model might be implemented. Finally, he thinks that connectionist models implement associationist principles and, for this reason, often refers to his model of language processing as ‘brute-causal associationist’.

But consider how this characterization of brute-causal processing bears on the present issue. If language processing is simply brute-causal—i.e., consisting of associationist or simple dispositional links between a thought and its expression in speech—*then there is no reason why the thoughts and sentences could not be radically different in respect of their syntactic properties*. After all, if language processing is not sensitive to the syntactic structure of the representations that are involved, then *it would not matter if thoughts and public-language expressions differed in their syntax*; the process would be, as it were, “blind” to these differences.

We can render this point vivid by reference to a concrete example. Consider a prominent connectionist model of comprehension, due to Rohde (1999). The figures below illustrate the syntactic and semantic representations that Rohde’s model generates when it is fed the input sentence: *The dogs that were chasing a car chased the boy with some fast cars in a big house*. As we can see in fig. 2, the “thought” in Rohde’s model is language-like and has a *syntactic* structure, consistent with LOTH. That is to say, the thought has constituent entities—i.e., the concepts PROPERTY, ENTITY, CHASE, and CAR, *inter alia*—which are bound to one another and grouped into larger expressions that denote events and their relations. However—and this is the crucial point—*the syntax of the thought is radically different from the syntax of the sentence*.<sup>47</sup>

---

<sup>47</sup> In this connection, consider the following passage from Devitt (2006). “How indeed could the speaker syntax of the public sentence differ from the syntax of the mental sentence it expresses? Perhaps it could do so if production does more than reduce explicitness. Thus suppose that a mental sentence has an SOV syntax and production transforms this into what appears to be an SVO public sentence; or that the mental sentence is active and production transforms this into what appears to be a passive sentence. (*Ignorance*, p. 157).” In limiting his discussion to these possibilities, Devitt seems to be overlooking much more radical alternatives, like the one currently under discussion, as well as the model due to Schank *et al.* (about which more below).

As Rohde (1999) puts it:

The notion of the meaning of a sentence, or message, used here is based on what one might consider its *semantic parse tree*. Although this is a recursively composed structure, it should not be confused with the syntactic parse tree. The semantic parse tree is not based on the words of the sentence but the constituent concepts and their relationships. A message has a single representation regardless of the language in which it is expressed.

What's more, the transition between the thought and the sentence is accomplished, in Rohde's model, in a way that Devitt would call "associationist" and hence, to some extent, brute-causal.

The process of comprehension is, according to Rohde, sensitive only to the local properties of the nodes in the network and the weighted connections between them—*not* the syntactic properties of the representations that these things implement.<sup>48</sup>

Another example of a computational processing model that is committed to LOTH but not to SLOTH is the model that we considered in §2, due to Schank and his associates. This model, too, is brute-causal, in the sense that it does not posit metalinguistic representations of syntactic structure. But as noted earlier (fn. 15), the output of the model is syntactically structured in a way that is radically different from the sentences of a public language. Like Rhodes' model, it represents events and their temporal, causal, and hierarchical relations.

---

<sup>48</sup> "The model cannot profitably be viewed as merely an implementation of a symbolic system in neural network hardware" (Rohde, 2002: p. 152).

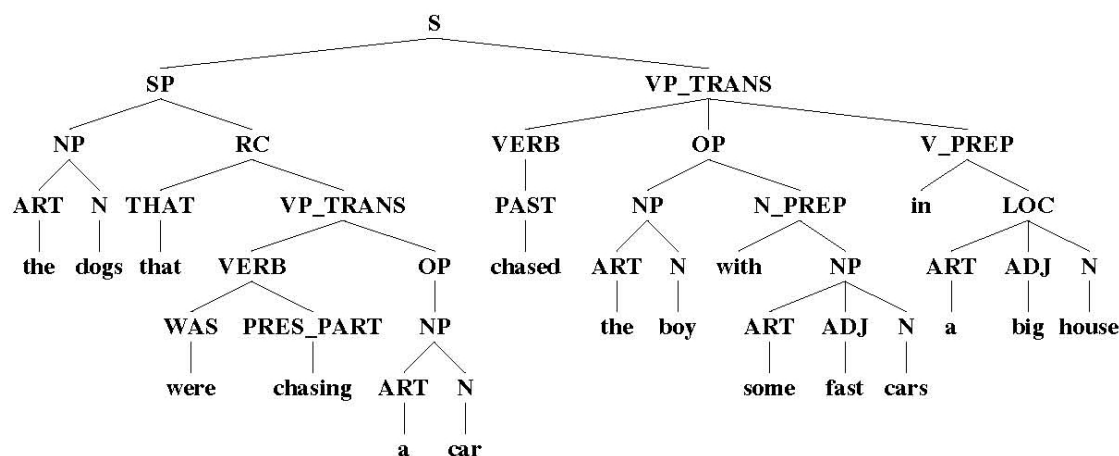


Figure 6: A Syntactic Parse Tree.

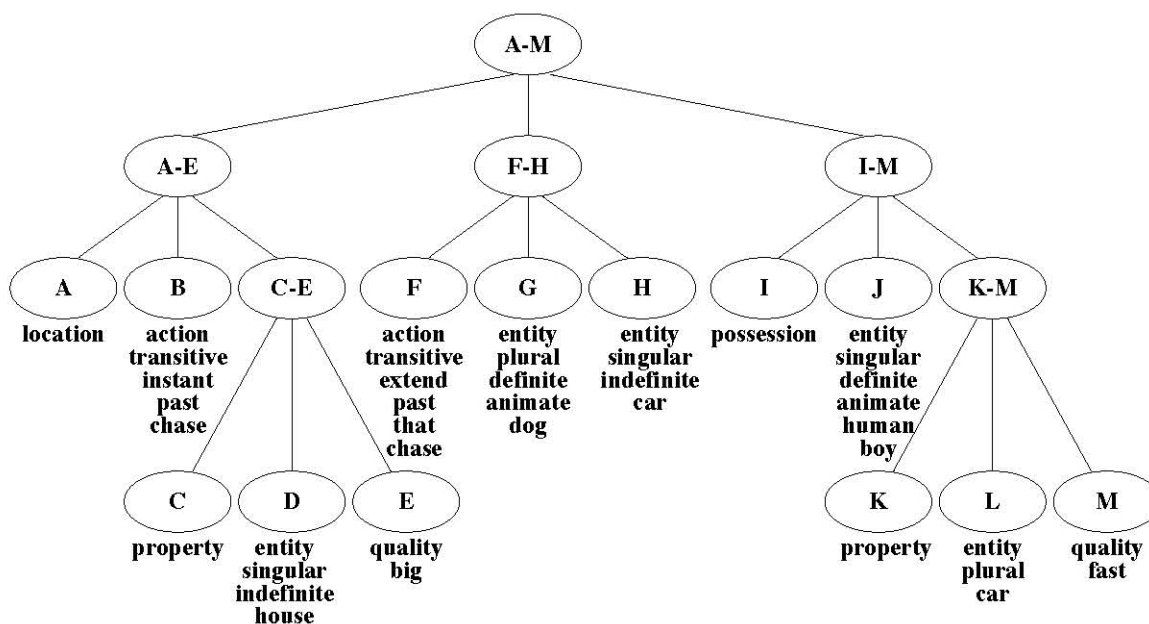


Figure 7: A Semantic Parse Tree.

**Figure 2:** Parse trees for the sentence *The dogs that were chasing a car chased the boy with some fast cars in a big house*. Source: Rohde (1999).

Let us sum up this objection to Devitt's argument for SLOTH. A process is brute-causal when it is blind to the syntactic properties of the representations that are involved in it. Thus, if language production—i.e., the process of translating from thoughts to speech acts—is brute-causal, then it is blind to the syntactic properties of the thoughts that are its inputs. If that is so,

then it is difficult to see how thoughts can “confer” their syntactic structures onto the sentences used in speech, or even to explain them. For, a thought can have entirely *different* structure than the sentence does—e.g., the structure proposed by Rohde (1999) or by Schank and Birnbaum (1984)—and yet still give rise to the sentence via a brute-causal process.

Devitt observes, in other places (§§11.5, 11.6), that language processing is relatively speedy and automatic. To explain this, he invokes the transparency between thought and speech that SLOTH affords. The idea is that the HSPM would have to do more work to turn a thought into a speech act, or vice versa, if the thought and speech have divergent syntactic structures. But Rohde’s model demonstrates that this is not the case. In general, a suitably large, properly organized, and well-trained connectionist network can automatically map one complex representation onto another in a very small amount of time.<sup>49</sup> Speed is, after all, a cardinal virtue of parallel distributed processing. So Devitt’s appeal to the speed and automaticity of language processing cuts no ice in the present context.

### §3.1.5 *Pinker’s Objections to SLOTH*

The objections that I have raised here to Devitt’s argument for SLOTH are closely related to an objection due to Pinker (1994). As it happens, Pinker develops two distinct lines of argument. Devitt discusses one of these, but not the other. Here, I propose to discuss both. I think that Pinker’s first argument simply fails; Devitt’s refutation of it is clear and effective. I include it in the present discussion only because it brings out some important aspects of Devitt’s

---

<sup>49</sup> Frank, Mathis, and Badecker (2008) say the following: “Rodriguez et al. (1999) and Rodriguez (2001) have shown that an SRN trained in [a certain] fashion is capable of inducing the kind of structurally-based generalizations representable by the class of context-free grammars. ... Lewis and Elman (2001) ... show that SRNs trained to do word prediction learn the conditions on subject-verb inversion, apparently inducing the correct structure-based rather than linear-order based generalization (cf. Crain and Nakayama 1987). Rohde (1999a) demonstrates that the SRNs can acquire knowledge of the conditions under which *want to* can contract to *wanna*, conditions which have been argued to require reference to abstract entities such as traces of wh-movement” (pp. 4-5).

position, to which we will return later in the chapter. Pinker’s second argument, which Devitt does not discuss, fares somewhat better. I will summarize it, indicate the relations it bears to the arguments that I presented above, and conclude that it presents a *potential* difficulty for Devitt’s commitment to SLOTH.

Let us call Pinker’s first argument ‘the argument from ambiguity’. Pinker describes some ways in which, according to him, public language differs from thought. Ultimately, these differences are all due to one or another type of ambiguity. The first is structural ambiguity, which Devitt illustrates with the now-familiar example:

- (33) Tex likes exciting sheep  
 (34) [S[NP[N Tex]][VP[V likes][VP[V exciting][N sheep]]]]  
 (35) [S[NP[N Tex]][VP[V likes][NP[A exciting][N sheep]]]]

On Pinker’s view, the English sentence (33) is ambiguous as between (34) and (35). As he points out, however, thoughts *cannot* be ambiguous in this way, for the simple reason that we individuate thoughts (both tokens and types) by their *meanings*—i.e., their *propositional* or *intentional contents*. Since (34) and (35) differ in *that* respect, they must correspond to two distinct thoughts. In short, Pinker’s point is that, unlike English, which is ambiguous, thought must be free of semantic ambiguity. Pinker goes on to mention other sources of ambiguity. He cites pronouns and deictic terms (*here, there, this, that, now, then, I, me, my, her, we, and you*), all of which introduce a certain kind of referential ambiguity into natural language. But, again, because thoughts are individuated by their meanings or propositional contents, any difference in referential or truth-conditional properties necessitates an explicit distinction in thought.

Devitt’s reply to this argument is, I believe, decisive. He begins with the observation that our best linguistic theory of public-language expressions individuates them by *all* of their properties—phonological, syntactic, and semantic—whether those properties are “explicit” in the surface string or “implicit” in their underlying structure. This holds for both sentences and



subsentential expressions. Linguists characterize public language words and morphemes by reference to a lexicon, in which each word and each morpheme is defined by *all* of its properties—a sort of “bundle theory,” to use Humean terms. It follows that a descriptively adequate grammar will contain no ambiguous English *words*, strictly speaking. The only things that are ambiguous, in any theoretically interesting sense, are *sounds and marks*. The point extends easily to the sentential case. When a theorist puts forward a *full* analysis of an English sentence, she includes in her analysis a level of representation where all of the syntactic and logical properties of the sentence are made explicit. Hence, there is, strictly speaking, no such thing as sentence (33). What appears in (33) is an ambiguous *inscription*, which can be used to express either of two distinct thoughts, by deploying either of *two distinct sentences*. Devitt makes this point in the following passage, and provides broad textual support from many leading figures in linguistics.

The differences between Mentalese and English in ambiguity and explicitness seem clear enough. But do these differences really show that Mentalese is not syntactically similar to English? There is persuasive reason for thinking that they do not. A triumph of generative grammar has been to make us appreciate how much of the syntax of a sentence is not explicit, “how unrevealing surface structure may be as to underlying deep structure”; “surface similarities may hide underlying distinctions of a fundamental nature” (Chomsky 1965: 24). Consider empty categories, for example: “An empty category ... is a constituent that has no phonological substance associated with it; it is inaudible in speech, and invisible in the standard orthography” (J. D. Fodor 1989: 156). Consider also the scope of quantifiers. This scope is often not explicit on the surface and yet is revealed at the level of logical form or LF, the level of syntactic structure that most concerns us here (1.3). As Gennaro Chierchia points out: “in the interpretation of sentences with quantified NPs, we apply scoping to such NPs. Scoping of quantifiers in English is a covert movement, part of the mental computation of meaning ... The result of scoping ... is what gets semantically interpreted and is called Logical Form”.<sup>[50]</sup> It “provides an explicit representation of scope, anaphoric links, and the relevant lexical information” (1999: c).

---

<sup>50</sup> Note that Chierchia’s claims rest on a processing model that posits mental phrase markers and mental computations over them—indeed, computations that involve shuffling around various constituents of those representations. This is incompatible with Devitt’s brute-causal model.

Considerations of this sort lead Stephen Neale to go further: “rather than saying that (30) [“Every poet respects some sculptor”] is an ambiguous sentence, really we should say that (30) is the surface representation of two distinct sentences ... that share an S-Structure representation and in fact look and sound alike” (1993: 119). In sum, once one takes note of the implicit structure of an English sentence it is far from obvious that this structure differs from that of the Mentalese sentence it expresses. Indeed, Chierchia goes on to say: “It is highly tempting to speculate that logical form actually is the language of thought” (1999: ci). ... In light of this, we can see that the objection moves far too swiftly from a difference in explicitness and ambiguity to a difference in syntax. (Devitt, 2006: pp. 153-157, underlining added)

Pinker might complain that Devitt is cooking the books by characterizing public language expressions in a more thought-like fashion. But the reply is clear: All Devitt is doing is insisting that we get suitably precise about what public language actually *is*. And his characterization of it is *independently supported*. Indeed, if anyone can be accused of cheating here, it’s Pinker, who relies on an underspecified and unmotivated conception of public language expressions.

Having refuted Pinker’s objection, Devitt’s puts forward his most considered formulation of SLOTH in the following passage, throughout which he refers to (33)-(35) above.

Suppose that a public sentence was not ambiguous either in word or structure and was produced as directly as could be from a mental sentence. *Then surely the speaker syntax of that public sentence would be the same as that of the mental sentence.* But perhaps there is no such public sentence. So now suppose that the public sentence is ambiguous in word and structure and that the production process simply reduces the explicitness of the mental sentence it expresses. Consider (33), for example. (a) We have noted that the mental sentence underlying a particular token of (33) must make explicit whether the structure is of the sort captured by (34) or (35). So, what happens in production is that this explicitness is lost. (b) The speaker may know several people named ‘Tex’ and so that name is ambiguous for him. The mental correlates of this name will explicitly mark the distinction between these various uses of the name: for example, <Tex1>, <Tex2>, and so on. This explicitness is also lost in production. Now if the only effect of the production of a public language sentence on the syntax of a mental sentence is to reduce its explicitness, it seems as if we must say that it has the same syntax as the mental sentence, albeit less explicitly. How indeed could the speaker syntax of the public sentence differ from the syntax of the mental sentence it expresses? Perhaps it could do so if production does more than reduce explicitness. Thus suppose that a mental sentence has an SOV syntax and production transforms this into what appears to be an SVO public sentence; or that the mental sentence is active and production transforms this into what appears to be a passive sentence. It seems rather unlikely to me that production would make these sorts of changes but if it did then perhaps we

should say that the syntax of the mental and the public differ. But even that isn't obvious. Perhaps we should say the sentences are implicitly SOV and active, respectively, and so they really have the same underlying syntax as the mental sentence they express. What I think we can conclude from all this is that if LOTH is correct, the speaker syntax of public-language sentences will be the same as, or similar to, the syntax of the mental sentences they express. So let us be cautious and say that they are similar. (*Ignorance*, pp. 156-157)

Pinker's argument from the alleged ambiguity of public language expressions was designed to show that there is a one-many relation between public language expressions and thoughts. We have seen why this argument does not work. However, Pinker has a *separate* argument to the effect that the relation is also *many-one*. That is, on his view, not only can each sentence express multiple distinct thoughts, but *each thought can be expressed by multiple distinct sentences*. To argue for this latter claim, Pinker supplies four public-language sentences, which he claims are synonymous—hence expressions of *one and the same thought*. However, on account of their syntactic differences, the sentences are plainly distinct.

- (36) Sam sprayed paint onto the wall.
- (37) Sam sprayed the wall with paint.
- (38) Paint was sprayed onto the wall by Sam.
- (39) The wall was sprayed with paint by Sam.

One wonders, of course, why we should treat these sentences as *strictly* synonymous, and hence expressions of *one and the same* thought. After all, it is perfectly conceivable that the individuation of thoughts is as fine-grained as the individuation of sentences (36)-(39). In that case (36)-(39) can be seen as expressions of *subtly different* thoughts. Pinker addresses this worry in the following remarks.

[Sentences (36)-(39)] refer to the same event and therefore license many of the same inferences. For example, in all four cases, one may conclude that the wall has paint on it. But they are four distinct arrangements of words. You know that they mean the same thing, but no simple processor, crawling over them as marks, would know that. Something else that is not one of those arrangements of words must be representing the single event that you know is common to all four. For example, the event might be represented as something like

(Sam spray paint<sub>i</sub>) cause (paint<sub>i</sub> go to (on wall))

which, assuming we don't take the English words seriously, is not too far from one of the leading proposals about what Mentalese looks like. (Pinker, 1994: pp. 80-81)

Let us carefully extract the various points of interest that Pinker raises in this pregnant passage.<sup>51</sup>

First, Pinker seems to be committed to the view that thoughts are individuated, at least in part, by the inferences that they license. This fits well with the functional-role semantics that I earlier attributed to Sellars and Rosenthal, as well as the view of semantic representation that was embodied in the models of Schank and his associates, discussed in §2. Whether it is compatible with Devitt's "molecularist" semantics is, I think, a trickier issue. Ultimately, though, I think that there is little if any tension between these views, especially when we put aside the red herring of holism.

Second, Pinker hedges when he says that sentences (36)-(39) "refer to the same event and therefore license *many* of the same inferences" (*ibid*, emphasis added). A radical holist about meaning will be quick to point out that licensing *many* of the same inferences is not the same as licensing *all* of the same inferences. Such a theorist would conclude, then, that the thoughts underlying (36)-(39) are *close* in meaning, but not strictly synonymous. And since only *strict* synonymy can warrant Pinker's claim that these sentences express *one and the same thought*, such a theorist would conclude that Pinker's argument fails.

Is Devitt entitled to this response? I think not. For, Devitt's molecularist view is precisely a *rejection* of radical holism. He holds that only *some* of the inferences in which a thought participates are constitutive of that thought's intentional content.<sup>52</sup> And the differences between the inferential roles of (36)-(39) are so slight that it seems highly implausible that they

---

<sup>51</sup> Note that Pinker's semantic representation, i.e., '(Sam spray paint<sub>i</sub>) cause (paint<sub>i</sub> go to (on wall))' is itself syntactically structured—consistent with LOTH—but its syntax is distinct from that of the sentences in (36)-(39). This reinforces my earlier objection to Devitt's argument for SLOTH.

should constitute a distinction in intentional content, on Devitt's own view. Rather, being a fan of a firm semantics/pragmatics distinction, Devitt would presumably consider the differences between (36)-(39) as being merely *pragmatic*, not semantic.

How, then, might Devitt reply to Pinker's argument? There are, I think, two possibilities, neither of which strikes me as wholly compelling. The first is to seize upon Pinker's toy proposal regarding the structure of the allegedly unique thought that underlies (36)-(39). The proposal crucially relies on the analysis of the transitive verb 'paint' according to which, *x paints y* has the semantic structure *x causes paint to go on y*. From the very outset of his career, Jerry Fodor has argued forcefully against this sort of proposal.<sup>53</sup> For instance, he points out that when an artist dips her brush into a can of paint before committing the paint to a canvas, she causes paint to go on the brush, but, intuitively, she does *not* thereby paint the brush! On this basis, Fodor argues that only a *specific kind* of causal transaction involving the transfer of paint can count as an instance of painting. Thus, Pinker's toy analysis is inadequate, and will remain so until his remark that we not "take the English words [e.g., 'cause'] seriously" is more fully cashed out. Fodor doubts that this can be done and concludes from this that the concept underlying the transitive verb 'paint' is just PAINT. This conclusion bodes well for Devitt's view.

How well this reply to Pinker's objection fares depends very much on how the details of Pinker's favorite semantic theory are cashed out. Pinker's suggestion is, after all, very informal, in accordance with the general tenor of *The Language Instinct*—a pop-science bestseller. It is by no means clear that *every* semantic theory that is compatible with Pinker's objection will

---

<sup>52</sup> Consider, for instance, the following passage: "I think, though Fodor (1987) does not, that some words will surely not be basic but rather covered by a description theory; <bachelor> is a likely example. Meanings for these words will come from inferential associations with others; for example, <bachelor>'s association with <adult>, <unmarried>, and <male>" (*Ignorance*, p. 156).

<sup>53</sup> See J. A. Fodor (1970, 1998).

succumb to Fodor's critique. This, however, is not an issue that can be settled *en passant*. Let us then move to a second possibility that is open to Devitt in this dialectic.

Devitt might urge that Pinker's individuation conditions on thought are simply inadequate. Just as Pinker's earlier argument made the mistake of individuating sentences solely on the basis of their superficial features, it may well be that the argument we are now considering makes an analogous mistake in assuming that thoughts are individuated solely by their intentional contents. Perhaps Devitt would insist that we should individuate thoughts by not only by their semantic properties, but by their syntactic properties as well. And if we are prepared to adopt these richer individuation conditions, then we open up the possibility that (36)-(39) express thoughts that differ, but *only* in respect of their syntactic properties.

This line of response has some initial appeal, but it also faces a potential difficulty. For, the proposal relies on the existence of *synonymous thoughts*, which differ *only* in their syntactic properties. One might wonder what explanatory purpose could ever require us to posit distinct-but-nevertheless-synonymous thoughts. It would be question-begging in this context for Devitt to insist that the explanatory purpose is precisely the one that Pinker has identified—i.e., the explaining the existence of distinct-but-nevertheless-synonymous *sentences*. And no other explanatory purposes come to mind that would license the posit in question. Still, it is clearly an empirical issue whether distinct-but-nevertheless-synonymous thoughts exist. That being the case, we cannot rely on the mere intuition that positing such thoughts is somehow suspicious. Pending a serious argument to that effect, we should conclude that an appeal to distinct-but-nevertheless-synonymous thoughts may well serve as the core of Devitt's most effective reply to Pinker's second argument against SLOTH.

### *§3.1.6 Concluding Remarks About SLOTH*

We began this section by introducing RTM, CTM, and LOTH. We then discussed some of the most popular arguments for LOTH and found them wanting. Turning to Devitt's argument for SLOTH, we found that it, too, faces serious difficulties. First, I argued that Devitt's explanation of public-language syntax in terms of Mentalese syntax is beholden to a dubious commitment to a view that accords explanatory priority to thought over language. Second, I pointed out that Devitt's explanation appears to be inconsistent with his other remarks concerning the explanatory ground of public-language syntax. I briefly indicated how these remarks, combined with the insights of Wilfrid Sellars and the arguments of his followers (e.g., Rosenthal, 1986/2005), can be used to articulate an alternative explanatory ground for public-language syntax. Third, I showed—partly by appeal to a concrete proposal from the connectionist literature—that Devitt's explanation is rendered otiose by his own commitment to a brute-causal model of language processing. Finally, I reviewed an objection to the conclusion of Devitt's argument—i.e., SLOTH itself—due to Pinker (1994). While the objection may not be decisive, it does require Devitt to bite the bullet and posit distinct-but-nevertheless-synonymous thoughts.

We can conclude from all this that SLOTH, which constitutes the first major pillar of Devitt's Strong Brute-Causal (SBC) model of language processing, is without support and quite possibly false. Let us now articulate the other major commitments of the SBC model and examine whether language processing is brute-causal in Devitt's sense of the term.

### *§3.2.1 The Brute-Casual Model of Language Processing*

Devitt (2006: §§3.4, 11.6) tentatively endorses what he calls the “brute-causal” model of language processing, according to which “processing rules [i.e., comprehension routines] do not

operate on metalinguistic representations of the syntactic and semantic properties of linguistic items but are directly responsive, in a fairly brute-causal associationist way, to these properties” (p. 275).<sup>54</sup> In this section, I spell out many of the details of this model and present several preliminary criticisms along the way. I reserve my most serious criticisms of the model for §3.3.

Before proceeding to this task, however, let me take a moment to make clear the bearing that this discussion has on the Representational Thesis (RT), according to which syntactic rules or principles are mentally represented in the minds of competent speakers. Devitt discusses the relationship between his brute-causal model and RT in the following passage.

My *third* goal [in chapter 11 of *Ignorance of Language*] concerns the metalinguistic representations of the syntactic and semantic properties of linguistic entities. According to [the rational-causal view], the processing rules operate on these representations; according to [the brute-causal view] they do not. If RT is true then [the rational-causal view] must be, but even if RT is false, [the rational-causal view] might be true: processing might be a rational information flow even if the rules governing it are not represented. [The rational-causal view] seems to be the received view even among many who reject, or do not clearly hold, RT. My goal is to cast doubt on this view, trying to make [the brute-causal alternative] seem plausible. I conclude that we have good, although very far from conclusive, reasons to think that the speedy automatic part of language processing is indeed fairly brute-causal associationist. (p.198)

The upshot of these remarks is that a vindication of the brute-causal model of language processing would effectively preclude any hope of sustaining RT. By contrast, a vindication of the competing *rational*-causal models would leave RT in the running. Having argued against the brute-causal model in this chapter, I will consider the case for RT in ch. V.

---

<sup>54</sup> Devitt often characterizes the processes as being not only “brute-causal” but also “associationist.” In §3.3, I will discuss the import of this additional descriptor. Two more points of terminological significance: First, what Devitt calls *metalinguistic representations* is just what I have been *calling mental phrase markers* (MPMs). Second, Devitt sometimes refers to the brute-causal model as “version (b)” of many popular positions on the psychological reality issue. Here is his explicit statement of the distinction between versions (a) and (b) of those positions: “language use [is] a fairly brute-causal associationist process—version (b)—rather than a process involving metalinguistic representations of the syntactic and semantic properties of linguistic expressions—version (a)” (*Ignorance*, p. 220).



At the outset of his defense of the brute-causal model, Devitt issues four important clarifications and qualifications. I propose to discuss each of these at some length. The first appears in the following passage:

Some parts of language use involve abilities other than those strictly of linguistic competence, abilities that we have been calling “pragmatic.” Thus, as we have already noted (§8.2), language comprehension involves assigning referents to indexicals. It also involves using contextual clues to remove ambiguities. And language use in general involves Gricean processes: in production, the process of tailoring the expression of thought to the hearers and circumstances; in comprehension, the process of figuring out what thought the speaker is likely to be expressing. Some of these processes may well involve central-processor reflections on linguistic items. Version (b) [i.e., the brute-causal proposal] is not concerned with any such process. The processes that are brute-causal according to version (b) are the *speedy automatic* ones which may arise solely from a person’s linguistic competence, from her ability to speak and understand her language, or may arise partly from that ability and partly from pragmatic abilities like that of assigning referents. My tentative proposal is that these processes are fairly brute-causal. (p. 220)

Recall that we saw a number of arguments in ch. III that demonstrate the existence of mental phrase markers (MPMs). Much of the work in contemporary psycholinguistics is premised on the idea that the HSPM quickly constructs MPMs in the initial stages of comprehension, thereby representing the syntactic properties of incoming linguistic stimuli. These representations are then refined and modified in various ways, on the basis of information about lexical items, syntactic constraints, as well as discourse context and encyclopedic world knowledge. The passage above illustrates that, on the brute-causal processing model that Devitt is proposing, this does *not* occur. That is, the comprehension process, on this model, does *not* involve the construction of MPMs or the explicit encoding of information about the syntactic or thematic structure of the input. Rather, it involves the automatic tokening of thoughts that are appropriate to incoming linguistic stimuli.

This raises a question about what exactly it means for an output of the comprehension process to be appropriate to the stimulus. In the passage above, Devitt says, quite plausibly, that

“comprehension [is] the process of figuring out what thought the speaker is likely to be expressing” (p. 220). However, his considered view, expressed in the following passage, departs from this in an important respect.

[W]hat we have to end up with in “the message-in box” is not a mental metalinguistic statement about the linguistic input, however detailed, but a mental statement that means the same as the input. That is, on hearing [a sentence] what we end up with is a translation of [that sentence]... That is what understanding the sentence amounts to. Understanding does not amount to metalinguistic analyses ... which tell us about the syntactic and semantic properties of [a sentence]. The task is not to generate a theory of the linguistic input but to generate a synonymous mental representation. (pp. 223-4).

Motivated by the basic commitments of the brute-causal model, Devitt characterizes the output of the comprehension process as a “mental *statement* that means the same as the input” (emphasis added). This, I think, cannot be correct. On the usual conception of speech acts, a statement is an *assertion*. Thus, a *mental* statement is one that amounts to a *mental assertion*—i.e., an *endorsement* of the propositional content of the input. But this *cannot* be what we arrive at in the final stage of the comprehension process, for we do not automatically come to *believe* whatever we hear or read. Endorsement cannot be necessary condition for understanding; the two come apart routinely.

The correct view, I think, is that the speaker constructs a personal-level (though not necessarily conscious) *perceptual judgment* to the effect that linguistic input has such-and-such speaker-meaning—i.e., a judgment to the effect that a speaker said *that p*. This proposal does incorporate a mental representation that is, in the ideal, synonymous with the linguistic input. That representation is what provides the value of ‘*p*’ in the comprehension schema. Unfortunately, to my knowledge, there is no consensus among psycholinguists about the end-products of comprehension. Alternative views are certainly conceivable. For instance, Devitt might opt for an account on which the hearer constructs a mental representation that is

synonymous with the linguistic input, and then adopts a *special* mental attitude—what he might call ‘understanding that *p*’—to this representation. His remarks in the passage above suggest just such a view.

There is, however, reason to avoid positing a special mental attitude of understanding—a “message-in” box, in addition to a “belief-box,” to use Schiffer’s terms. Parsimony dictates that we draw, as much as possible, on the stock of mental attitudes to which we are antecedently committed. At a minimum, this includes (occurrent) judgment and (dispositional) belief. The view that I endorsed above makes do with just these. To what extent does that view cohere with the brute-causal model? That depends on one’s commitments regarding the nature of attributions of speaker meaning—i.e., indirect discourse. Insofar as this is taken to be *metalinguistic*, in the sense that Devitt intends, then the view I proposed above is at odds with the brute-causal model. I confess, however, that I do not know how to determine whether this is so.

Let us return to our principal theme: Devitt’s rejection of mental phrase markers (MPMs). The work done by MPMs in standard psycholinguistic models will have to be offloaded, on the brute-causal model, to either “brute-causal associations” or what Devitt thinks of as “pragmatic abilities.” From one of the passages quote above, we can glean that these include: (i) the ability to assign referents to indexicals or, more generally, to determine who or what the speaker is likely to be talking about, (ii) personal-level (perhaps conscious) reflection on linguistic items, and (iii) the use of general world knowledge and of clues about the discourse context, in the service of ambiguity resolution.

In connection with the models proposed by Schank and his colleagues, I argued in §2.3 that pragmatic abilities are simply not sufficient to account for our capacity to understand a wide range of sentences, particularly those that contain anaphora or exhibit movement or embedding.

I shall argue in §3.3 that brute associations likewise *cannot* play the explanatory role that Devitt's model requires of them. This will be especially clear in those cases where hearers resolve local syntactic ambiguities on the basis of *the global syntactic context* of an expression. I will argue that this is because syntactic contexts can extend *arbitrarily far back in time*, and encompass *discontinuous* fragments of the prior stimulus stream.

Devitt issues a second qualification in the following passage:

[T]he *received* view in linguistics and psycholinguistics seems to be the version (a) view that processing operates on metalinguistic representations. ... I shall present some reasons for doubting this received view and for favoring a brute-causal alternative, particularly version (b)... The point is not, of course, to offer the alternative as a complete explanation of language use. Like the received view, it is far far too lacking in details for that. The point is rather to suggest that the best explanation is more likely to comply with the brute-causal alternative than the received view. And the considerations favoring the alternative are, it goes without saying, far from decisive. The fourth proposal [i.e., that "the speedy automatic language processes arising wholly or, at least, partly from linguistic competence are fairly brute-causal associationist processes that do not operate on metalinguistic representations of the syntactic and semantic properties of linguistic expressions"] really is tentative. (p. 221)

Several points of interest emerge here. First, he acknowledges that the parsing models on which the HSPM constructs mental phrase markers currently enjoy widespread acceptance among linguists and psycholinguists—a fact that I stressed in ch. III. Second, Devitt concedes that the arguments he offers for the brute-causal model are far from decisive, and that his confidence in that proposal is, consequently, rather tentative. Finally, Devitt makes clear that he does not intend to provide a detailed theory of comprehension. Rather, his aim is to give reasons for supposing that a certain *type* of parsing model is superior to the currently well-received proposals.

Needless to say, the incompleteness of a comprehension model does not, by itself, count against its plausibility. *Every* theory, in *every* field, is to *some* extent incomplete—but no less

plausible for all that. This applies not only to the dominant theories of parsing and reanalysis in mainstream psycholinguistics,<sup>55</sup> but also the brute-causal model that Devitt tentatively advances. That this model is incomplete should not, by itself, count against it. However, even incomplete theories must be constrained by empirical findings. And here, I suggest, the brute-causal model falls short, on account of its inability to explain the behavioral and neurocognitive data discussed in ch. III and below.<sup>56</sup>

Let's turn now to Devitt's third qualification:

We are not concerned here with what seem intuitively to be performance failures. Thus, there will be sentences of the language that a person cannot translate satisfactorily into thought in the brute-causal way. And there will be sentences she cannot produce because she cannot think the corresponding thought. (But will there be any thoughts that she can think but cannot translate into sentences of the language?) Presumably the central processor helps deal with the sentences that she cannot immediately translate. If the "message-in box" receives an unacceptable message, then the central processor orders a rerun: using memory of the starting point, another attempt at translation is made. If even this fails to yield a satisfactory message in comprehension, then the central processor may have to take a serious hand. Perhaps this will involve deploying such declarative linguistic knowledge as the person has acquired at school and elsewhere to try to analyze the input. But it may mostly consist in trying several times to get "a feel" for the troublesome sentence by comparing it to others until the syntactic structure is recognized by the associative memory. And this may fail: there may be sentences of the language that a speaker simply cannot translate into Mentalese or, at least, cannot without a great deal of help from a teacher. Thus, I haven't yet managed to parse Fodor's example, 'Bulldogs bulldogs bulldogs fight fight fight' as a sentence despite the clue "Take the first two verbs as transitive" (1975: 168). (*Ignorance*, pp. 227-228)

In ch. III, we saw that a surprisingly wide range of linguistic stimuli involve "performance

---

<sup>55</sup> Fodor and Ferreira, eds. (1998); Fodor and Inoue (1994, 1998, 2000a, 2000b); Sturt, Pickering, and Crocker (2001); Meng and Bader (2000a, 2000b); Stevenson (1993/1998, 1999).

<sup>56</sup> Devitt (2006) surveys a great deal of empirical literature and discusses an astounding range of experimental and theoretical findings. However, *Ignorance of Language* contains strikingly little discussion of the priming, eye-tracking, reaction-time, and ERP studies of parsing and comprehension that psycholinguists have painstakingly garnered over the past 50 years. Devitt does discuss the infamous click experiments (pp. 72-3), but as I pointed out in ch. III (§3), these are no longer considered to be reliable. Jurafsky and Martin (2007: pp. 424-5) point out that these studies failed to control for semantic biases that correlate with syntactic structure. Clark and Clark (1977) discuss several other methodological problems.

errors” in Devitt’s sense.<sup>57</sup> This raises the question of whether all such cases can be subsumed by Devitt’s proposal in this passage. I think they cannot. True, in some cases, there is a mismatch between a semantically or pragmatically sensible reading and the one that the HSPM initially delivers, prior to reanalysis. These cases fit the description that Devitt gives when he says that “[i]f the ‘message-in box’ receives an *unacceptable message*, then the central processor orders a rerun: using memory of the starting point, another attempt at translation is made” (*ibid*, emphasis added). However, we saw that there are also cases in which a linguistic input is both semantically and pragmatically kosher, but causes difficulties *solely on account of syntax*.

Consider two examples of this phenomenon: First, the well-known main-clause/relative clause ambiguity (40) and a high/low attachment ambiguity (41).

- (40) The dealer sold the forgeries complained.
- (41) I didn’t put the dead mouse that I found outside my office.

In these cases, the brute-causal model seems to predict, contrary to fact, that there will be *no* processing difficulties. For the messages in (40) and (41) sentence are perfectly clear—i.e., semantically and pragmatically flawless—assuming, of course, that the syntax “comes for free,” as it were. Of, course, contrary to the strong brute-causal model, the syntax does *not* come for free—it must be *computed* on the basis of the clues in the prior or posterior regions of the stimulus stream, which I will argue in §3.3 must be *explicitly represented* by the HSPM.

It is also unclear how the brute-causal model—as articulated in the passage quoted

---

<sup>57</sup> As always, we must bear in mind that processing difficulties *need not be conscious*; we now know that they extend *well* beyond the conscious case. The behavioral and neurophysiological indications of processing difficulty need only be accessible from the third-person point of view. No first-person report need be elicited, though it may well constitute additional (albeit, typically *superfluous*) evidence for a processing model. There are, in any case, outstanding issues about how reliable such reports can really be. If the case of sentence processing is anything like that of other perceptual faculties (e.g., color vision or pain) or even of more “cognitive” processes, such as self-evaluative judgment-formation, then we can expect the reliability of first-person reports to be *quite low* in a strikingly wide range of conditions. See Rosenthal (2005) for a discussion of the reliability of first-person reports.

above—can account for all of the subtleties of the behavioral and neurophysiological phenomena associated with the process of reanalysis. The theory of reanalysis that one finds in Devitt’s account of comprehension is limited to the following remarks: “the central processor orders a rerun,” “using memory of the starting point, another attempt at translation is made,” and “[reanalysis] may mostly consist in trying several times to get ‘a feel’ for the troublesome sentence by comparing it to others until the syntactic structure is recognized by the associative memory” (*ibid.*) This rerun-and-get-a-feel account of reanalysis fares rather poorly in comparison to the most sophisticated garden-path models that have been articulated to date (see fn. 56), all of which posit mental phrase markers.<sup>58</sup> Contrary to Devitt’s suggestion (pp. 12, 221), these models are far more detailed than the brute-causal alternative, and have been explicitly constructed to account for fine-grained behavioral and neurocognitive data. Here are three examples of hard-won generalizations that such models can now explain:

- (1) For German speakers, reanalysis routines that are triggered by a mismatch between a noun and a verb in respect of the (syntactic) feature of number-agreement engender *greater* processing difficulties than those triggered by a mismatch between Case features.

---

<sup>58</sup> The connectionist “constraint-satisfaction” model of Rohde (2002) does not posit mental phrase markers. He writes: “[T]he current model certainly does not involve a separate all syntactic phase of parsing. Lexical frequency and semantic information pervade all stages of processing. While this is true of many recent theories of sentence processing (MacDonald et al., 1994a; Tabossi et al., 1994; Trueswell & Tanenhaus, 1994), the current model makes the additional claim that the human language system does not necessarily engage in parsing per se. Certainly the model must develop some knowledge of the syntactic structure of the language to do a reasonable job in comprehension and production. But performing these tasks, mapping from a surface form to a message and back again, does not necessarily require the model ever to construct an explicit representation of a sentence’s syntactic structure” (p. 152). Rohde goes on to admit, however, that “[t]he model is unable to perform reanalysis. It must process a sentence sequentially and cannot return to earlier material. A more complete system should be able to self-monitor its progress and, when it gets lost, jump back to *an appropriate point*. This could be done visually in the case of normal reading. During listening or during masked reading, this must be done with some sort of short term, perhaps phonological, memory buffer. Reanalysis of this sort is actually quite complex as it requires the ability to discard those aspects of the message that derived from the incorrectly processed portion of the sentence. Thus, the comprehender must maintain *some sort of connection* between the surface form of the sentence and the meaning that has been extracted from it” (p. 155, emphasis added). Small wonder that a system with no representations of syntactic structure fails to maintain the relevant “sort of connection” and to return to the “appropriate point” in the sentence. For, the relevant “sort of connection” is plainly a syntactic one, and the “appropriate point” in the sentence is typically definable only in terms of the syntactic relations between constituents of the sentence.

- (2) For English speakers, the generalization in (1) does not hold.
- (3) Recovery from a low-attachment error—the kind due to an application of the processing principle known as Late Closure—is typically much more difficult than recovery from other types of purely syntactic errors.<sup>59</sup>

It is, at present, not clear how a brute-causal model accounts for these (and countless other) data. We can call this the incompleteness problem, which, as we noted above, is, by itself, relatively benign. More serious is the *inadequacy* problem—i.e., that brute-causal models do not have the theoretical resources to explain such data *even in principle*. In §3.3, I will argue directly that brute-causal models face this problem, even in the case of first-pass parsing. Here, I offer a preliminary assessment, by illustrating how specific reanalysis effects are explained by competing “*rational-causal*” models, and gauging whether the brute-causal model can yield a substantive alternative to such explanations.

The Diagnosis Model of reanalysis, due to Janet Dean Fodor and Atsu Inoue (see references in fn. 56) has it that, in the face of an unexpected or deviant input, the HSPM follows a policy of “Attach Anyway.” That is,

the parser forcibly attaches the symptom word into the tree structure anyway, creating an ill-formed tree in which that word conflicts with some other constituent. The *grammar detects the conflict*,<sup>[60]</sup> and the parser changes the other constituent to eliminate the conflict. The changed element may now conflict with something else in the tree, so the something else now has to be altered; and so on. If all goes well, the parser will eventually work its way back to the original error that was the source of the problem, and will correct that. Whether the parser ultimately arrives at that point, and at what cost in processing effort, depends on how informative the symptom is about what is wrong with the structure and how to put it right; if the symptom is misleading, or if there are ambiguities within the revision procedure, the

---

<sup>59</sup> Fodor and Inoue (2000a) write: “This is one fact that almost all work on reanalysis agrees on: raising a low-attached constituent is hard (see papers in Fodor and Ferreira, 1998)” (p. 264)

<sup>60</sup> Note that Fodor and Inoue take the grammar to be playing a causal role in the detection of an error. In order to do this, they hold, the grammar must be mentally represented—i.e., the representational thesis (RT) must be true. This commitment goes well beyond what I seek to establish in this chapter—viz., that the HSPM constructs mental phrase markers. In the next chapter, however, we will examine the case for RT, by reference to a wide range of psychologically plausible parsing models that have been developed in computational linguistics.



repair attempt may fail. (Fodor and Inoue, 2000a: p. 262, emphasis added)<sup>61</sup>

To show this model in action, Fodor and Inoue discuss a *successful* case of reanalysis.

(42) Have the sick soldier walked to the first-aid tent.

Following Marcus (1980), Fodor and Inoue note that sentence (42) typically engenders a mild garden-path effect, because the HSPM has an initial bias of treating sentence initial ‘Have’ as an inverted auxiliary verb, whose function is to introduce a question. From this, it follows that the subsequent NP will be the subject of the clause. Moreover, the agreement features associated with ‘Have’ dictate that this NP must be plural. (If the auxiliary were ‘Has’, then the HSPM would have posited a *singular* NP in subject position.) This analysis works well for the HSPM as it integrates the next two words of the input—i.e., ‘the’ and ‘sick’—but it fails when the singular noun ‘soldier’ is encountered. At this point, according to Diagnosis Model, the principle of Attach Anyway is applied, causing a domino effect of revisions to the mental phrase marker.

Since there is nowhere else to attach ‘soldier’ on this misanalysis, we assume the parser attaches it to the plural NP anyway. There, it conflicts with the plural marking on the dominating NP node, so the plural marking is adjusted to singular. Now the singular NP node conflicts with the plural verb ‘have’ that it is supposed to be in an agreement relation with. So now ‘have’ must be altered. It cannot be made singular, so it must be expelled from the agreement position. Then the unattached ‘have’ and the NP ‘the sick soldier’ must be rejoined; and the only way to do so is with ‘have’ as a nonfinite imperative verb—just as in the sentence that begins ‘Have Thomas...’

Note that Fodor and Inoue’s accounts, of both the initial processing difficulty and its ultimate resolution, make ineliminable reference to mental phrase markers. The explanation of the processing difficulty appeals to the computational resources that the HSPM must expend in operating on these representations and altering their internal structure. *This* is what accounts for

---

<sup>61</sup> Terminological note: Fodor and Inoue use term ‘tree structure’ in the way that I have been using ‘mental phrase marker’.

the increased gaze times, reaction times, error rates, and distinctive ERP signatures. I find it difficult to see what an explanation of those phenomena could possibly look like if it eschewed reference to mental phrase markers. It looks as though the brute-causal model will not, *even in principle*, be able to account for the processing difficulties involved in such cases, let alone to describe in detail the operations that must be deployed in resolving them. (See also fn. 58 above.)

Putting aside the problems surrounding reanalysis, let us consider how Devitt elaborates the brute-causal model of the *earliest* stages of language processing:

My concern is not with whether language use involves representation of the phonetic and phonological properties of the sounds. Clearly, these properties have to be recovered somehow from the acoustic signal at the beginning of speech perception and bestowed somehow on the signal by the end of speech production. [fn. 24: “Speech perception processes...require a transformation of the auditory input from the peripheral auditory system to a spectral representation based on more generalized auditory patterns or properties, followed by the conversion of this spectral representation to a more abstract feature (phonological) representation, and ultimately the mapping of this sound structure onto its lexical representation.” (Blumstein 1999: 644)] I assume that this must involve, at least, representations of the phonetic properties, the “physical” properties of the sound, as a result of a transduction. Beyond that I shall (wisely) have nothing to say on the vexed question of what else it involves, including on the influence of lexical knowledge on the recovery process. I suppose, although I shall not argue, that my reasons for doubting that syntactic and semantic properties are represented would carry over to phonological properties. (p. 222)

Let us focus our attention on Devitt’s positive claim: At the very least, the *acoustic* properties of the signal must be represented by the hearer. Devitt does not discuss the reasons for this requirement, presumably because it is wholly uncontroversial. Still, I think that it is worth taking the time to spell out the role that acoustic representations play in his model, for doing so provides us with the means to evaluate proposals regarding what *else* must be represented in the course of comprehension. I believe that, when spelled out, the reasons for positing acoustic representations overlap with—though do not exhaust—the reasons for positing MPMs.

Acoustic representations are crucial for the recognition of lexical items or the activation of concepts.<sup>62</sup> If, for instance, a hearer is not sensitive to the distinction between the sounds /rake/ and /lake/, but *does* have two distinct lexical items or concepts, RAKE and LAKE, then the sounds would be, for her, *ambiguous*. This is the situation of many native Korean and Japanese speakers, who struggle to acquire English late in life. To many of them, sounds like /rake/ and /lake/ are nearly indistinguishable. Thus, in order for them to successfully comprehend a message that contains either ‘rake’ or ‘lake’, they must resolve the ambiguity in some other way—e.g., by using semantic or pragmatic clues to determine the mapping from the ambiguous acoustic representation into one of their two concepts, RAKE or LAKE. Unfortunately, such clues are not always available. And, in any case, it is plain that such a mapping is much easier to carry out when the relevant acoustic distinction is explicitly represented. Thus, to explain the native English speaker’s fast and relatively effortless comprehension of these sounds even in the absence of non-acoustic cues, we must posit representations of fine-grained acoustic properties.<sup>63</sup>

Removing ambiguities is, then, a crucial part of comprehension. But, as noted above (and discussed at length in ch. III), in addition to acoustic ambiguities of the *rake-lake* variety, there are countless other types of ambiguity, at every level of linguistic analysis. Fast and effective comprehension routines must be capable of resolving these as well. A key question for parsing theorists is how the HSPM accomplishes this feat. As Phillips (1996) puts it,

---

<sup>62</sup> Below, I address the issue of whether we should draw a distinction between lexical retrieval and concept activation.

<sup>63</sup> As Devitt’s quotation from Blumstein suggests, we must also posit *phonological* representations, which are more abstract than acoustic or phonetic ones, in the sense that their relationship to the acoustic properties of the stimulus is more complicated and indirect. Devitt supposes that his “reasons for doubting that syntactic and semantic properties are represented would carry over to phonological properties” (*ibid.*). By parity of reasoning, I suggest—though I will not argue—that the arguments for positing mental phrase markers can be parlayed into arguments for positing phonological representations as well.

Experimental studies of sentence processing have tended to focus on situations in which the incoming word and the existing representation can be combined in more than one way consistent with the grammar, i.e. structural ambiguities. This is a particularly rich area for experimental investigation, because it is possible to experimentally manipulate the various factors that might influence people in choosing between the different possible continuations of an incomplete sentence (structural, pragmatic, probabilistic, lexical etc.). In fact, the study of ambiguity resolution dominates research in sentence processing to such an extent that it can sometimes seem as if ambiguity resolution is the only topic in sentence processing. (*Order and Structure*, p. 107)<sup>64</sup>

Now, if our approach to the resolution of syntactic ambiguities is analogous to our approach to the resolution of *acoustic* ambiguities, then it seems that, by parity of reasoning, we must commit ourselves to the parser's generating and employing representations of *syntactic* properties.

Of course, as in the case of acoustic ambiguity, theoretical alternatives are available.

Devitt seems to favor the view that syntactic ambiguities are resolved in accordance with heuristics that require no representation of syntactic properties. For instance, he seems to applaud the following quotation from J. A. Fodor (1975).

English speakers can infer messages from representations of grammatical relations... But though they presumably can, they presumably don't. What apparently happens is that grammatical relations are computed only when all else fails. There exist heuristic procedures for sentence recognition which, in effect, ignore grammatical relations and infer messages directly from lexical content, accepting, thereby, the penalties of fallibility. (Fodor, 1975: 167–8).

Devitt goes on to wonder “why these heuristic procedures could not be associative and why they could not be all there is to the speedy automatic process of language processing” (p. 227). In §3.3, I will indicate why the process of comprehension should not be viewed as associative, and I will argue that the states of the HSPM really do have both structural and representational properties. Presently, however, I would like to address the “heuristic procedures” that Fodor and

---

<sup>64</sup> Note that, in appealing to an “existing representation” with which the incoming word can be combined, Phillips is already assuming version (a)—i.e., the rational-causal approach. Devitt would urge, no doubt, that the same point can be made without this presupposition.

Devitt mention in the quotations above. My claim is that these are simply not sufficient, in general, for successful comprehension.

The heuristics in question come in two varieties. The first variety consists of deploying only semantic and pragmatic resources. We saw that this was the approach favored by Schank and his associates at Yale. As I argued in §2.3, this approach cannot handle any but the simplest inputs. The second variety of heuristics—the one that Fodor (1975) probably has in mind in the passage quoted above—is drawn from Fodor, Bever, and Garrett (1974). These are prescriptions for what structure to assign to an incoming stimulus on the basis of very general clues. The paradigm example of such a heuristic is the so-called “Canonical Sentoid Strategy,” according to which the HSPM will interpret an N-V-N sequence as having an S-V-O structure. This proposal is now known to be untenable.<sup>65</sup> As we noted in ch. III, psycholinguists have summarily abandoned it, pursuing more promising avenues of research that stem, in large part, from the surface-structure parsing principles proposed by Kimball (1973)—close ancestors of Minimal Attachment, Late Closure, and the Minimal Chain Principle (ch. III), all of which make ineliminable reference to mental phrase markers.

Pritchett (1992: pp. 22-26) points out that both the Canonical Sentoid Strategy and the other heuristics that Fodor, Bever, and Garrett proposed were later found to be at variance with a wide range of experimental data. Moreover, from the standpoint of a psycholinguist, the heuristics-based approach is *inherently* unsatisfactory, because it posits different strategies for what are essentially the same linguistic phenomena. This complicates both parsing theory *and* the theory of language acquisition. Wanner and Maratsos (1978) express this point in the

---

<sup>65</sup> As J. D. Fodor (1998a: p. 287) observes, it wouldn’t work *at all* for verb-initial languages like Irish or verb-final languages like Japanese. That’s a lot of languages! Of course, it may be that parsing heuristics differ from language to language. But a uniform (i.e., language-neutral) account of parsing preferences would be vastly more preferable.

following important passage.

With the empirical failure of psychological models that incorporate transformational rules ... many psychologists have abandoned not only transformational grammar but also the assumption about the psychological reality of the generalizations on which transformational grammar is based. The heuristic strategies schemes advanced by Bever, Fodor, and others make no claim to capture linguistically significant generalizations. For example, Fodor and Garrett (1967; also Fodor, Garrett, and Bever 1968) propose one heuristic strategy to handle object relative clauses and an entirely different heuristic strategy to handle subject relative clauses (Fodor, Bever, and Garrett 1974). Presumably, their system of strategies can only be expanded to handle the full range of relative clauses by adding more heuristic strategies. Such a system does not class relative clauses together in any way, nor does it represent their relation to declarative clauses. If a child acquired this type of grammatical representation of relative clauses, he would face the formidable task of having to learn each relative clause strategy independently of all the others, and none of his knowledge of declarative clauses would transfer to relative clauses.

We thus have strong grounds for rejecting any parsing model that relies solely on the kinds of heuristics that Fodor, Bever, and Garrett proposed.

To sum up: We have seen that one of the main functions served by acoustic representations is the facilitation of lexical retrieval. And I have suggested that representations of *syntactic* properties play a similar role. In §3.3, I will give examples of this from the literature on first-pass parsing. For now, our earlier example of reanalysis will suffice to illustrate the point. Models that incorporate feedback loops between the syntactic level of analysis and the process of lexical retrieval offer powerful explanations of the behavioral and neurocognitive evidence of processing difficulty. Hence, pending a compelling alternative account of these data, our best theory is that, in reanalysis, representations of hierarchical syntactic relations play a role in *correcting* an initial retrieval of the wrong lexical entry. In the case we discussed (i.e., *Have the sick soldier walked to the first-aid tent*), the HSPM engages in a subpersonal analogue of a rational inferential process, at the end of which it “concludes”—on the basis of both the input and its internal representation of syntactic constraints—that ‘Have’ is not an inverted auxiliary

verb but, rather, a nonfinite imperative verb. This is a clear case in which representations of syntactic properties facilitate lexical selection. Until the brute-causal model yields an equally satisfying explanation, we have strong grounds for adopting the rational-causal alternative.<sup>66</sup>

### §3.2.2 *The Relation Between Lexical Items and Concepts*

We have touched on the topic of lexical retrieval. Let us examine it in some more detail, with an eye toward clarifying Devitt's position. Two interesting issues arise: The first concerns exactly what is represented in the lexicon. The second is whether Devitt's processing model distinguishes between lexical entries and "Mentalese words," i.e., *concepts*—the constituents of personal-level propositional attitudes. We will take up these issues in turn.

Psycholinguists differ about what information is stored in the lexicon, but there are several claims that are relatively uncontroversial and derive abundant support from the available experimental data. It has been firmly established, for instance, that the frequency of a lexical item has strong effects on subsequent processing, so we may assume that this information is stored in the lexicon.<sup>67</sup> More importantly, we've seen evidence that *syntactic* information is represented there as well. At a minimum, the top-level syntactic category of an item must be associated with it in the lexicon, distinguishing it as a noun, verb, adjective, preposition, or any of a host of more exotic parts of speech. Standard parsing models typically rely on the lexicon to provide more fine-grained information about a verb's subcategorization properties—i.e., which

---

<sup>66</sup> Note also that if the explanation appeals to SLOTH, then it is groundless, pending a response to the arguments that I advanced in §3.1.

<sup>67</sup> See Jurafsky (2003) for a discussion of the relevant findings.

complements it allows or requires. For instance, in our discussion of Traxler and Pickering (1996) in ch. III, we saw that the HSPM represents the verbs ‘recognize’ and ‘know’ as taking either a nominal or a sentential complement.<sup>68</sup> We also saw in that discussion that the HSPM represents the pronouns ‘she’ and ‘her’ as having Nominative and Accusative Case, respectively, and uses this information *immediately* upon accessing those lexical entries. So Case must be added to our list of syntactic features represented in the lexicon.

In addition, a number of parsing models rely heavily on the lexicon’s storage of a verb’s argument structure.<sup>69</sup> For instance, the verbs ‘sleep’, ‘wear’, and ‘put’ take one, two, and three arguments respectively. This is demonstrated by the following examples.<sup>70</sup>

(43) David is sleeping. / \*David is sleeping the blanket. / \*David is sleeping the blanket on the table.

(44) \*David is wearing. / David is wearing the shirt. / \*David is wearing the shirt on the table.

(45) \*David is putting. / \*David is putting the book. / David is putting the book on the table.

Other syntactic features, such as person (1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>), number (singular or plural), and gender (masculine, feminine, neuter, etc.) are likewise associated with lexical items. On the standard models, these facilitate the efficient construction of a mental phrase marker, which involves binding together items that share these properties. For instance, in processing the fragment *Which songs does...*, the HSPM recognizes that the NP ‘which songs’ is the antecedent of a *wh*-

---

<sup>68</sup> A number of studies have demonstrated the rapid access of subcategorization information during comprehension. Gorrell (1991), for instance, used a reaction time measure on a lexical-decision task to show that the HSPM makes a distinction between transitive and intransitive verbs within milliseconds of lexical access. Frenck-Mestre and Pynte (1995) obtained similar results in an eye tracking study.

<sup>69</sup> The best-known example is Pritchett (1992), though there are several similar proposals in the literature.

<sup>70</sup> The examples given here gloss over a great many of subtleties. In the interest of brevity, I shall not enter into a digression concerning the syntactic arguments for the distinction between arguments and adjuncts, nor the tests for deciding whether a seemingly complete sentence is elliptical for a one that contains additional arguments in its overt form. These are delicate matters and a full discussion would take us well beyond the scope of the present work.



trace and that it is marked [+plural]. It follows that the *wh*-trace must also be marked [+plural] and thus can appear only in positions in which a plural constituent is licensed.

The HSPM faces the task of determining where in the mental phrase marker to put a *wh*-trace. According to the grammar, the *wh*-trace cannot appear after the word ‘does’ because this is marked [+singular] and hence would not be in agreement with the [+plural] *wh*-trace. The HSPM is demonstrably sensitive to this constraint, in that it does not even consider the position immediately after ‘does’ as a plausible candidate for the *wh*-trace. This can be shown using the cross-modal priming paradigm discussed in J. D. Fodor (1989, 1991).

We have now assembled a list of information types that psycholinguists assume to be encoded in the lexicon: frequency, category, subcategory, Case, person, number, and gender. Let us now ask whether the brute-causal model shares this widely held conception of the mental lexicon. We expect, antecedently, that it will not. For, in standard psycholinguistic models, lexical retrieval is a necessary step precisely because lexical entries contain information about syntactic features, which the HSPM needs for the purpose of constructing mental phrase markers. On the brute-causal model, by contrast, the HSPM does not construct mental phrase markers, so lexical retrieval would seem to be an unnecessary processing step.

We can shed light on this issue by considering Devitt’s most detailed statement of the brute-causal model. (I have underlined the most relevant aspects of the passage.)

Let me fill out the brute-causal view a bit more, starting with words. The view is that there is a simple association stored in memory between a linguistic word—more accurately, a representation of its phonetic [i.e., acoustic] properties—and the corresponding mental word with the result that the one leads straight to the other without any analysis. This association is established by the regular use in the community of the linguistic word to express the mental word... The same story also applies to familiar sentences—for example, ‘What’s the time?’ and ‘How are you?’—and familiar expressions, including idioms—for example, ‘kick the bucket’, ‘butterflies in the stomach’. The view is not, of course, that these simple associations are all there is to the use of these expressions. Thus, pragmatic abilities will be called

on in language comprehension to determine the reference of ‘you’, to remove ambiguities, and so on... But the view is that simple associations are at the core of the processes. This story for words and familiar expressions seems very plausible. And there seems to be no reason why it cannot be extended to syntax. The basic idea again is of simple associations stored in memory, in this case of syntactic structures in thought with similar (implicit and explicit) syntactic structures in language so the presence of the one prompts the other. Each of the structural features of any linguistic sentence—the features captured by structural descriptions ... or by phrase-structure trees—is associated with a similar structural feature of mental sentences. This association is established by the regular use in the community of that linguistic structure to express the mental structure, a regularity that also establishes the conventional syntax of that linguistic structure. In comprehension, a person must identify the structure from clues provided by the syntactic category of words (which comes with word recognition), word order, and the like. (*Ignorance*, pp. 225-226)

It may appear, at first, that there is a tension between some of the claims that comprise this proposal. For, on the one hand, Devitt claims at the outset of this passage that a representation of the acoustic or phonetic properties of a word “leads straight to ... the corresponding mental word ... *without any analysis*” (*ibid.*, emphasis added). As noted above, this suggests that the model has no room for a process of lexical retrieval. On the other hand, by the end of the passage we are told that “[i]n comprehension, a person must *identify the structure* from clues provided by *the syntactic category of words (which comes with word recognition)*, word order, and the like.”

The tension between these remarks may, however, be illusory. Taking into account Devitt’s aforementioned commitment to the Strengthened Language of Thought Hypothesis (SLOTH), we might instead construe him as rejecting the commonly held distinction between, on the one hand, lexical entries and, on the other hand, “Mentalese words” or, less poetically, *concepts*. On this interpretation, Devitt holds that *there is no difference between activating a lexical item and activating a concept*. Rather than being distinct processes, on this view, lexical retrieval *just is* the activation of a concept.<sup>71</sup>

---

<sup>71</sup> See also the following remarks: “We can go further down this brute-causal path .... Constraint-based theories, with their links to connectionism, certainly give support to a brute-causal associationist picture of ... language

If correct, this interpretation would allow us to see how Devitt can hold two apparently conflicting commitments—i.e., (i) his claim that there is a direct, brute-causal association between phonetic representations and lexical items, and (ii) his claim that word recognition involves the activation of a syntactic category. For, in accordance with SLOTH, concepts have syntactic categories *just like words do*. The concept READ, for instance, is a “mental verb” that subcategorizes for two “mental nouns” and whose argument structure contains two free “mental theta-roles.” Moreover, unlike the closely related concept READS, the concept READ is marked [–3<sup>rd</sup>-person]. On Devitt’s view, then, the “activation” of a syntactic category is not the activation of a metalinguistic *description* of the input—e.g., a classification of the input word as a transitive verb. Rather, what is activated is a mental entity that, as a matter of fact, *belongs* to the category “mental transitive verb.”

While this tidy interpretation is certainly tempting, it becomes difficult to sustain when we take note of Devitt’s remarks about language *production*. He begins his discussion of production by citing the work of Kay Bock, as well as other theorists who endorse models that seem to be wholly committed to mental phrase markers.

Language production consists in processes “from mind to mouth”. These processes start with intentions to communicate messages, intentions to express thoughts. The processes go from these intentions to the articulation of linguistic expressions: “the cognitive processes that convert nonverbal communicative intentions into verbal actions. These processes must translate perceptions or thoughts into sounds, using the patterns and elements of a code that constitutes the grammar of a language.” There are two steps in language production, a grammatical one (producing a “lemma”) and a phonological one (producing a “lexeme”). “Grammatical encoding refers to the cognitive mechanisms for retrieving, ordering, and adjusting words for their grammatical environments, and phonological encoding refers to the mechanisms for retrieving, ordering, and adjusting sounds for their phonological

---

comprehension. ... On all these theories, *we can see an ambiguous word of an utterance as activating the several concepts with which it is linguistically associated, and an ambiguous structure as activating the several conceptual structures with which it is linguistically associated*” (*Ignorance*, p. 239; emphasis added).

environments” (Bock 1999: 453). At the grammatical stage words with the appropriate meanings have to be retrieved and put into a syntactically appropriate frame. It is “generally assumed that syntactic properties of words are retrieved as a consequence” of word retrieval.

Expanding on the process of lexical retrieval, Devitt approvingly cites another model that seems to be up to its ears in syntactic representations, both within the lexicon itself and in the stages of processing that follow lexical retrieval.

Assuming that the speaker wants to express the event ‘The ferocious dog bites the boy’, the stored meaning-based representations for ‘ferocious’, ‘dog’, ‘to bite’, and ‘boy’ are retrieved. These representations would further specify that ‘dog’ and ‘boy’ are nouns (hence usable as subject and object), ‘to bite’ is a verb (hence specifying a relation between the nouns), and ‘ferocious’ is an adjective (hence it can modify a noun). These representations also specify that ‘dog’ and ‘boy’ are common, countable nouns, and may be pluralized, and that ‘to bite’ may be used to refer to the past, present, or future, depending on the speaker’s intentions. The message-level information controls the assignment of the different lexical elements to their intended grammatical functions (e.g. ‘dog’ is assigned to be the subject, and ‘boy’ as the object and not vice versa, which would result in the unintended sentence: ‘The boy bites the ferocious dog’). Message-level information also ensures that the appropriate specification of number for the noun ‘dog’ (singular) and tense for the verb ‘to bite’ (present) are passed along. (Vigliocco and Vinson 2003: 185)

All of this seems to be in direct conflict with Devitt’s brute-causal model. But then Devitt goes on to say the following:

What goes on in this retrieval process? There is no suggestion that the process is governed by represented rules. Does it even involve representations of syntactic and semantic properties? Researchers talk of the process retrieving “stored information” (p. 184; see also Levelt et al. 1999: 2), but this storage may be in the form of implicit memory, about which much remains to be learnt (11.5). There seems to be nothing here that requires the representation of syntactic or semantic properties of ‘dog’. Indeed, what would be the use of retrieving the information that ‘dog’ means DOG? The account is compatible with the following brute-causal picture: ‘dog’ is selected because it is associated with <dog>; its subject position is associated with the place of <dog> in the message; its singularity is associated with the singularity of <dog>. What about the syntax of the sentence? The retrieval of ‘dog’ is the retrieval of something that is a noun and hence something that must be fitted into the sentence in the way appropriate for a noun. But this does not require a representation that it is a noun. Of course, more has to be done to fix the syntax of the sentence; for example, the singularity of the subject demands the singularity of the verb. But researchers see this sort of grammatical constraint as “procedural knowledge concerning how words

can be used in sentences” (Vigliocco and Vinson 2003: 185) and so there is no clear need for these constraints to be represented.

This passage raises problems for our earlier tidy interpretation of Devitt’s view, for Devitt repeatedly distinguishes here between the lexical entry ‘dog’ and the Mentalese concept <dog>. The passage also throws into the mix a number of thorny issues regarding “implicit memory” and “procedural knowledge.” We will examine these notions in ch. V.

Perhaps Devitt’s remarks regarding lexical retrieval can all be squared with one another, though I must confess that I do not, at present, see how. What is clear, though, is that Devitt’s notion of brute-causal processing is a bit of a moving target, not least on account of its association with the vexed notion of implicit memory. In order to evaluate the SBC model at its core, we will first need to gain a better understanding of what makes a process *brute-causal*. Let us now turn to this crucial topic.

### §3.3.1 *A Precise Characterization of Brute-Causal Processing*

Although the notion of brute-causal processing is central to his rejection of the psychological reality of mental phrase markers, Devitt is up-front about the fact that his characterization of this notion is a bit rough.

I appreciate that this characterization leaves [the] distinction [between brute-causal and rational-causal processing] somewhat vague,<sup>15</sup> particularly on the brute-causal side. But this cannot be avoided without an overcommitment to a particular view of fairly brute-causal processes. [fn 15: Note that we cannot characterize it as the distinction between computational and noncomputational processes because, although computational processes are frequently taken to be those of a “classical” architecture, there are a variety of views of computation: “It is surprisingly difficult to find an answer within the cognitive science community to the question of whether there is a univocal notion of computation that underlies the various different computational approaches to cognition on offer.” (Wilson 1999: xxix; see also Brian Smith 1999)] (*Ignorance of Language*, p. 56)<sup>72</sup>

---

<sup>72</sup> I would like to register that I do not share Devitt’s doubts about the univocality of our shared notion of computation. The notion is adequately captured by Chalmers (1993/forthcoming, 1995). “A physical system

There is a question about whether avoiding overcommitment to a specific characterization of brute-causal processing helps Devitt's argument. As noted above, I find that it presents something of a moving target. To fix ideas, then, let us reconstruct Devitt's discussion of this key notion, distilling out several defining features. In the quotations that follow, I will underline the remarks that will be most helpful in this regard.

Devitt begins by providing the following characterization of brute-causal processes:

“Here we have in mind operations that are arational and more ‘brute-causal’ and ‘primitive’. In such an operation, the processing rule is directly responsive to the input/output: the input directly triggers a fairly brute-causal processing operation and the output is the direct result of that brute-causal operation.” He goes on to supply a number of examples:

### 1. Traditional associationism

Associationism provides paradigm examples. The elements of classical associationism are mental, experiences and ideas. Learning is a matter of pairing these according to certain laws. Mental processes are simply a matter of moving from an experience to an associated idea, or from an idea to an associated idea. Association is a primitive relation in the theory. (p. 55)

### 2. Behaviorism

Behaviorism is nonmentalistic associationism: the mental elements are replaced by physical ones, environmental stimuli and behavioral responses. Mental processes are replaced by processes from stimuli to associated responses, processes that are like reflexes in their brutal simplicity. (p. 55)

### 3. The SpheX Wasp

---

implements a given computation when there exists a grouping of physical states of the system into state-types and a one-to-one mapping from formal states of the computation to physical state-types, such that formal states related by an abstract state-transition relation are mapped onto physical state-types related by a corresponding causal state-transition function” (Chalmers, 1995: p. 392). Chalmers goes on to replace this informal statement with a formal description of finite-state automata and a generalization to combinatorial-state automata. He then argues, quite effectively to my mind, that his characterization of computation is both appropriate and useful for the foundational role that this notion plays in cognitive science. In particular, it is neutral on whether the mind/brain is a “classical” computational device or whether it instantiates one of the nonclassical alternatives.

Daniel Dennett finds what seems to be a nice example in the “simple, rigid and mechanical” behavior of the wasp *Sphex* as it drags a previously paralyzed cricket to its burrow (1978: 65-6). The wasp leaves the cricket on the threshold of the burrow whilst “checking” inside. If the cricket is moved back from the threshold whilst the wasp is inside then, on emerging, the wasp will drag it back to the threshold and repeat the “checking” procedure. If the cricket is moved again, then the wasp will check again. And so on indefinitely. (p. 55)

#### 4. Hull-Tolman behaviorism

sensory stimuli and behavioral responses can be linked in [a] direct way not only to each other but also to internal states, perhaps representational ones. These are examples because the rules governing operations are not sensitive to the syntactic properties of representations. (p. 55)

#### 5. Moderate behaviorism

Moderate behaviorism is the first step toward positing a complex functional organization within the object that intervenes between the stimulus and the response. This organization, however complicated, might be governed by rules of the sort in question because its causal operations might not manipulate structured representations. (p. 55)

#### 6. Old-fashioned mechanical calculators

Perhaps the rules governing addition in an old-fashioned mechanical calculator are of this sort. (p. 55)

#### 7. Motor skills in humans and animals

Perhaps most of the rules governing motor skills like those of the diving heron, the bee dancing, and human bicycle riding, swimming, catching, and touch-typing are too. (p. 55)

#### 8. Transducers

Wherever a system is governed by rules for processing representations there have to be transducers that take physical inputs and transform them into representations for processing and take representations that are the result of processing and turn them into behaviour. The rules governing transducers are further examples of processing rules that do not manipulate structured representations. (p. 55)

## 9. Connectionist models

[T]he rules governing connectionist networks are notoriously of this sort, for connectionism is a complex form of associationism: “associative learning is what neural networks implement” (Macintosh 1999: 183); “simple associationism is a particularly impoverished and impotent corner of the connectionist universe” (Smolensky: 1991: 165). (p. 55)

## 10. Cognitive maps in insects and animals

Would the operations on cognitive maps that are alleged to explain the navigational behavior of animals (Gallistel 1990) count as rational-causal? I think not but I confess to being far from confident about this. These operations are sensitive to the “syntax” of the cognitive map but do not seem to involve any manipulation of symbols. Rather, given an appropriate stimulus—for example, an ant obtaining food—the map leads “automatically” to a certain behavior—heading straight back to the nest. (p. 54, fn. 13)

Having supplied an abundance of examples, Devitt goes on to clarify his characterization of brute-causal processes, and to apply it to the case of language comprehension. The first clarification is that representations *may* be involved in a brute-causal process.

We note that representations may be involved in a brute-causal process. Thus the processes of Hull-Tolman behaviorism and of connectionism may involve representations as surely do those of many skills. But these processes are not governed by rules that manipulate representations according to their syntactic structure in the rational-causal way. (p. 56)

Still, in the case of sentence processing, Devitt claims that representations are *not* involved.

If a rule for language use governs a brute-causal process then it may respond to and produce items that have properties like *being a VP* and *meaning CLINTON* without there being any complex metalinguistic representations of those items as having those properties. ... [T]he inputs/outputs of language use *are* representations of a certain sort, [but] they might be processed without being *represented as* representations of that sort. (p. 56)

In casting doubt on the psychological reality of these representations, Devitt invokes what he calls “Pylyshyn’s Razor,” which “requires us to find an explanatory need for supposing that language use is rational-causal before positing the metalinguistic representations.”



Putting all these remarks together (including the numbered quotations above), we can discern three core features of brute-causal processing. A process is brute-causal if

- i) it is associative; the only rules that govern the mental operation of moving from one mental state to the next are the rules of associationism, or sufficiently like them. (1), (9)
- ii) it is like a reflex, in respect of simplicity; simple, rigid, and mechanical. (2), (3)
- iii) its causal operations either do not involve structured representations at all (5), or the rules governing it are not sensitive to the syntactic properties of those representations. (4), (8), (10)

Having achieved a tolerable grip on this characterization, we can go on to ask whether any aspect of it is applicable to language processing. I believe that there are decisive difficulties for the view that the HSPM's comprehension and production routines exhibit any of these features. Let us now survey these difficulties, with an eye toward making clearer the notion of mental representation that is at the very core of the debate.

### *§3.3.2 Is Language Comprehension an Associationist Process?*

Devitt's characterization of brute-causal processing makes heavy use of notions from the associationist tradition. There is, however, an issue about what exactly counts as an associationist rule of processing. Fodor, Bever, and Garrett (1974) lament the "adherence to a terminological convention whereby *any* [process] is ipso facto called 'associative' whether or not there is reason to believe that it has the logical properties characteristic of associations and whether or not there is reason to believe it satisfies associative laws" (p. 76). Of course, their worry was much more pertinent in the mid-1970s than it is today. Nevertheless, their injunction to specify exactly what one means by 'association' still holds.

Unfortunately, Devitt does not say much about this. Mainly, the clues he provides are contained in his remarks about the link between associationism and connectionism. These remarks, however, are difficult to square with the practices of present-day connectionists,

especially those working on sentence processing. For instance, Devitt quotes Smolensky (1991) as saying that “simple associationism is a particularly impoverished and impotent corner of the connectionist universe.” He seems to ignore the fact that Smolensky, like many connectionists, is actually *resisting* the common characterization of connectionist networks as implementations of old-school associationism. Indeed, a look at Smolensky (1991) reveals that the quoted passage continues by pointing out that it is “*a serious mistake* to presume connectionism to be committed to simple associationist principles” (p. 165, emphasis mine). Rohde (2002) elaborates on this point as follows:

One complaint often leveled against connectionist networks is that they are only able to form stimulus-response associations. Therefore, they should have been discarded as a serious account of human language processing along with behaviorism. While this may be true of the simplest one-layer networks, it is not true of multi-layer backpropagation networks composed of units with non-linear response properties, which is usually what is meant by connectionist networks these days. Multi-layer networks, during training, have the ability to develop higher-level internal representations that represent abstract properties of the environment to which the network is exposed. ... The ability to develop abstract representations is ... one of the principal factors that sets modern connectionist networks apart from the earlier program of associationist behaviorism, which assumed that only observable stimuli, responses, and reinforcements play a causal role in behavior and learning. (pp. 10-11).

This is particularly true of the psychologically plausible connectionist models of sentence processing that are available today—including a constraint satisfaction model proposed by Smolensky and his colleague, Suzanne Stevenson.<sup>73</sup>

Thus far, the notion of association in Devitt’s remarks remains somewhat obscure. But, though Devitt does not do so, it is possible to clarify the notion, by identifying the characteristics of associationist processes that are at play in the work of the classical associationists—Aristotle,

---

<sup>73</sup> Stevenson, S. and Smolensky, P. (2006). “Optimality in Sentence Processing,” ch. 19 in *The Harmonic Mind*. Smolensky and Legendre, eds., Cambridge: MIT Press, pp. 307-337.

Hume, Locke, Mill, and others<sup>74</sup>—or, more pertinently, in the behaviorist psycholinguistics of the early 20<sup>th</sup> century. Fodor, Bever, and Garrett (1974: ch. 2) do precisely this. Their conclusions do not bode at all well for the claim that human sentence processing is governed by associationist principles. Let us briefly review two of their arguments.

The classical associationists had it that one mental representation ( $R_1$ ) is disposed to give rise to another ( $R_2$ ) to the extent that the things that  $R_1$  and  $R_2$  represent are contiguous in either space or time. The principle of association-by-temporal-contiguity shows up again in the behaviorist tradition—albeit in a modified form, stripped of commitment to mental representations. It serves as the foundation of conditioning theory, according to which the presentation of two stimuli that are contiguous in time (e.g., bell sounds and food) gives rise to a conditioned response (e.g., salivation triggered by bell sounds alone).

Fodor, Bever, and Garrett (henceforth, FBG) note that if language processing were governed by the principle of association-by-contiguity, then the recognition of “grammatical dependency between two items other than adjacent ones should occur very rarely in sentences” (p. 75). But, as they go on point out, “precisely the reverse is the case” (*ibid*). We saw in ch. III that the HSPM is highly sensitive to dependency relations between syntactic constituents that are *arbitrarily far apart in a sentence*. Indeed, it *must* be sensitive to such long-distance relations, since, as FBG note, “the interpolation of material between syntactically interdependent parts of sentences is a fundamental principle of every language” (*ibid*).

---

<sup>74</sup> See Harnish (2002: ch. 1) for an illuminating discussion of the historical development of associationist psychology. For original sources, see Hume (1740/2000, 1748/1999). J. A. Fodor (2003) provides interesting discussion of Humean associationism in the context of contemporary philosophy of mind and cognitive science.

FBG continue their assault on associationist psycholinguistics by examining the *optional* character of the syntactic relationships that the HSPM must handle. They consider the following pattern of data:

- (46) John phoned Mary up.
- (47) \*John divorced Mary up.
- (48) John phoned up Mary.
- (49) John phoned her up.
- (50) \*John phoned up her.

As the contrast between (46) and (47) illustrates, the particle ‘up’ is dependent on the verb ‘phoned’. But the dependence is of a very specific sort. In particular, if the object of the verb is a pronoun, then the particle *must* appear *after* that pronoun (49)-(50), though it *may* appear elsewhere in other cases (48). FBG argue at length that associative principles, such as association-by-contiguity, do not “provide mechanisms for reconstructing the intricate interplay of optional and mandatory decisions involved in specifying the order relations among elements of a sentence” (*ibid*).<sup>75</sup>

Having launched these criticisms—alongside several others, which I pass over in the interest of brevity—FBG draw the following moral:

[V]erbalization, like other forms of intelligent behavior, resists treatment of as a special case of a habit. Roughly speaking, the relation of association is locally transitive, probably irreflexive, and probably asymmetrical. In the case of any given organism, it holds between at most a finite number of pairs of behavioral elements, and its strength is highly responsive to the frequency with which the associated items cooccur in the organism’s experience. Grammatical relations, on the other hand, are of a variety of different logical types; they characteristically hold of infinite sets of elements, and their mastery appears to be largely independent of the experience of the speaker. In short, grammatical relations differ in practically every conceivable way from associative relations; *it is difficult to imagine a less likely candidate than association for reconstructing the psychological processes of language use*. The interaction between taxonomic grammar and Hulleian psychology failed to produce a viable account of either the structure of natural languages or the psychology of

---

<sup>75</sup> This is best seen as a problem for associationist theories of language *production*, though there are analogous difficulties in the case of comprehension. See FBG, ch. 2 for detailed discussion.

the speaker hearer. (*The Psychology of Language*, pp. 76-77)

Thus far, I have focused only on the first aspect of Devitt's characterization of brute-causal processing—i.e., his appeal to associationist state-transitions. I have not attempted to cast doubt on the existence of either brute-causal or purely associationist processes; indeed, it may well be that such processes are ubiquitous in the psychology of many creatures. The arguments I have surveyed establish only that *human language processing* does not have this character.

### §3.3.3 *Is Language Comprehension Simple, Rigid, or Reflex-Like?*

Let us now turn to the second aspect of Devitt's characterization of brute-causal processing, according to which it is like a reflex, in respect of its simplicity and rigidity. This too, I think, fits poorly with what we know about parsing and comprehension.

To begin with, we must get clear on what exactly it means to say that parsing is simple, rigid, or reflex-like. Here, Devitt's appeal to the example of the Sphex wasp is particularly instructive. Recall his description of the wasp's behavior. The wasp responds to a specific stimulus—its prey (a cricket) being some distance away from the threshold of its burrow—in the same exact way, over and over, *every single time*. The stimulus-response connection is both rigid and unique: same stimulus—same response. Consequently, extending our conception of the stimulus—for example, by taking the stimulus to be constituted by *multiple* consecutive encounters with the same cricket-threshold configuration—would *not* give us any additional predicative or explanatory leverage. For, even after having encountered that configuration *multiple* times, the wasp continues to engage in the very same behavior. Hence, taking the stimulus to be a single encounter with such a configuration is sufficient to predict what the wasp will do. The surrounding context seems to make no difference. Let us now see whether this characterization is applicable to the operations of the HSPM.

We have observed several times that the HSPM encounters stimuli that are almost exclusively ambiguous, typically at every level of analysis. In particular, the linguistic signals to which the hearer is exposed are rife with both syntactic and semantic ambiguity. Hence, a key question for parsing theorists concerns how the HSPM manages to cope with such rampant ambiguity.<sup>76</sup> As we saw in our survey of psycholinguistic studies in ch. III, behavioral and neurocognitive experiments have revealed that the HSPM engages in processes of rather impressive complexity. In particular, its processing routines are highly sensitive not only to the syntactic and semantic properties of each incoming morpheme, but also to the syntactic and semantic *context* within which that morpheme appears.

One immediate consequence of the fact that morphemes must be integrated into larger syntactic and semantic units is that *there is no unique response to a specific word or syntactic category*. In one context, a morpheme or a part of speech in the stimulus will be integrated in one way; in a different context the same material will typically be integrated in a different way. This feature of human parsing routines stands in stark contrast to the rigid behavior of the Sphex or, for that matter, to *any* reflex. Being struck in a particular spot on the knee *always* gives rise to the same knee-jerk response; and, as Jerry Fodor (1983) points out, if your best friend endeavors to poke you in the eye, you will *invariably* flinch. As with the Sphex, the context—prior stimuli or background knowledge—doesn't make a difference. This, again, is the essence of the S-R connections that Devitt sees as characteristic of brute-causal processes. But parsing is *not like that*. For the HSPM, *context always matters*.

Faced with the flexibility and variability of a system's responses to particular stimuli, a

---

<sup>76</sup> Recall the statement from Phillips (1996): "the study of ambiguity resolution dominates research in sentence processing to such an extent that it can sometimes seem as if ambiguity resolution is the only topic in sentence processing" (p. 107).

theorist sympathetic to Devitt's position might adopt a strategy that was popular in the heyday of behaviorism—indeed, one that still is popular in the study of animal behavior. The strategy is to increase, as it were, the “size” of the stimulus—i.e., to concede that there is no unique cognitive (let alone behavioral) response to a *single morpheme*, but to nevertheless insist that the response to *larger* linguistic units—sequences of words, phrases, sentences, etc.—*is*, in fact, reflex-like. This strategy will not, however, work in the present context; we simply know too much about language processing to take such claims seriously.

To begin with, we know that the HSPM processes linguistic stimuli *incrementally*. That is, rather than waiting for multiple words to come in before making its decisions regarding attachment and hierarchical dependency, the HSPM initiates syntactic analysis as soon as the first morpheme is received. Parsing is, to a first approximation, a morpheme-by-morpheme affair.<sup>77</sup> Second, the point about ambiguity and context-sensitivity applies *all the way up*, from phrases to sentences. Having encountered, say, a noun phrase in context  $C_1$ , the HSPM will perform one set of operations, while in  $C_2$  it will perform another. And, as the processing of globally ambiguous sentences illustrates, the same principle is operative one level higher. That is to say, variations in discourse context give rise to different interpretations of the very same sentence-sized stimulus. For instance, in one context, a speaker will naturally interpret the string ‘Visiting professors can be boring’ as a comment about a certain kind of activity; in another context, the same string will be interpreted as a comment about a certain kind of professor.

Needless to say, in any particular fixed context, there is *in fact* one and only one thing that the HSPM *does*—indeed, quite predictably, as we saw in ch. III. For instance, upon encountering the string ‘Janet knows the student’, the HSPM will (*ceteris paribus*) interpret ‘the

---

<sup>77</sup> Recall again the neurocognitive studies that we looked at in ch. III (§2), e.g., Hahne and Friederici (2002).

student’ as constituting the object of the matrix verb ‘know’. Hence, the HSPM will not “expect” the string to continue with material that reveals ‘the student’ to be the subject of a subordinate clause (as in ‘Janet knows the student is working’). Regularities such as this make the scientific study of comprehension routines possible. However, it is crucial to see that *it does not follow from this that parsing is a rigid reflex*. For, as noted above, the HSPM will *not* interpret ‘the student’ as the object of a matrix verb when the verb is something other than ‘know’ and will *not* interpret *every* NP as the object of ‘know’. The parser’s representation of prior context weighs crucially on such decisions.

Our earlier reflections on the Sphex wasp provide another way of putting largely the same point: In the case of the Sphex, “enlarging” our conception of the stimulus would *not* yield any predictive or explanatory gains. We predict that the Sphex will do the same thing, regardless of whether we take the stimulus to be one presentation of a cricket-threshold configuration or multiple such presentations. By contrast, in the case of the HSPM “enlarging” our conception of the stimulus *does* make a difference. A big one! For, in considering a *complex* phrase to be the stimulus, rather than a single word or morpheme, we take into account crucial bits of *context*, thereby gaining the predictive or explanatory leverage that such information typically affords.

#### §3.3.4 *The Distinction Between Responding and Representing*

The considerations discussed above bring us directly to the third aspect of Devitt’s characterization of brute-causal processing. Recall that Devitt takes a process to be brute-causal when its operations do not involve manipulating structured representations, or when the rules governing that process are not sensitive to the syntactic properties of those representations.<sup>78</sup>

---

<sup>78</sup> I am, at present, staying neutral on the question of whether the rules that govern language processing are mentally represented. For all that has been said thus far, it may well be that they are not. As I argued in ch. II, the



Applying this to language comprehension, Devitt claims that the rules governing the operations of the HSPM make no reference to the syntactic properties of mental phrase markers. He writes:

If a rule for language use governs a brute-causal process then it may respond to and produce items that have properties like *being a VP* and meaning *CLINTON* without there being any complex metalinguistic representations of those items as having those properties. (Devitt, 2006: p. 56)

In light of the discussion in the preceding pages, let consider a natural objection to Devitt's view, which Devitt (2006) anticipates in the following dialogue with a hypothetical opponent.

**Objection:** The way a word should be processed depends on its syntactic category, on whether it is a verb, noun, adjective, or whatever. What the processor must have then are rules for each syntactic category. It surely does not have a rule for each word of that category. So, it has rules for processing all adjectives, not a rule for each adjective. And such a general rule must operate on a representation of a word as a member of that category, on a representation like <That is an adjective>.

**Response:** Why must the general rule operate on a representation? To represent a word as an adjective, the processor must first recognize that it is an adjective. Whatever clue enables this recognition could simply trigger rules appropriate for adjectives: the processor responds to it as an adjective without representing it as one. Being an adjective is not a local physical property of a word, as Rey likes to emphasize (1997: 128; 2003b: 178–9), and so recognizing a word as having that property may not always be an easy matter (although it mostly is; see 10.6). But it is not made any easier by supposing that it involves representing the word as an adjective rather than simply responding to it as an adjective. (*Ignorance of Language*, p. 225)<sup>79</sup>

---

Representational Thesis (RT), according to which syntactic rules or principles are represented in the mind, is a substantive thesis, in need of empirical defense. Hence, in casting doubt on Devitt's claim that language processing is brute-causal, I do *not* take myself to be directly supporting RT. We will explore the case for RT in chs. V-VI.

<sup>79</sup> Following up Devitt's reference to §10.6 of *Ignorance*, we find the following: "Sometimes, it is easy to tell that an object has a certain relational property because that property is well correlated with superficial properties. This makes it quite easy to tell ... many English adverbs, the ones that end in 'ly'. It can also be easy to tell that an object has a certain relational property if learning to identify the object involves learning to identify it as having that property. This makes it quite easy to identify the other English adverbs; identification comes with word recognition. One way or another, it is quite easy to tell the explicit structural properties of utterances although sometimes hard to tell the implicit ones. (pp. 185-6)." This, I think, underestimates the complexity of the parsing task. J. D. Fodor (1989) provides examples of some quite common cases in which it is hard to tell the "implicit" structural properties of a stimulus expression: "[T]he antecedents of NP-trace and PRO are typically normal argument NPs, which cannot, when they are encountered, be recognized *as* antecedents. When the processor encounters the predicate ... it can tell that an empty category *may* be coming, but even then it is not certain. Nicol (1988, p. 108ff.) notes that a passive verb is often confusable with an adjective, and a passive *by*-phrase may be confusable with a temporal or locative phrase. Similarly, a raising verb may be homonymous with a non-raising verb (as in *John appeared NP-*

Throughout his response, Devitt explicitly draws a distinction between *representing* something as having a certain property and “*responding to it as*” having that property.<sup>80</sup> As I indicated above, I believe that this distinction plays a key role in Devitt’s rejection of the psychological reality of mental phrase markers.<sup>81</sup> Without appealing to it, Devitt would not be able to explain how non-local, high-level relational properties could have the systematic effects that they do on human language comprehension.

The distinction between representing and responding is, I think, quite real. Indeed, trivially so, since not everything is a representational system. When a billiard ball strikes the bumper of a billiard table, the bumper responds to the ball as a massive object. But it would be pointless to say that the bumper *represents* the ball as a massive object. No appeal to representation is necessary in explaining the outcome of the collision. Still, while the representing/responding distinction is quite real, there are problems with Devitt’s invocation of it in the present context. I shall now argue that language processing is *not* a mere “response,” in Devitt’s sense, but *must* involve representations—specifically, mental phrase markers.

Consider first the fact that there is simply no *unique* way of “responding to [a word] as an adjective” (*ibid.*) This is the now-familiar context-sensitivity point. The presence of an adjective in the stimulus does not, *by itself*, necessitate or invariably give rise to the same cognitive processes on every occasion. Assuming, for convenience, the GB framework, we can

---

*trace to dislike the cake* / *John appeared to Susan in a dream*); and a sentence containing PRO may be temporarily indistinguishable from a non-PRO construction (e.g., *John intended PRO to leave* / *John intended no harm*).” (197).

<sup>80</sup> At other times, Devitt speaks of hearers (or their behavior) as being “sensitive to” various syntactic properties. See, e.g., *Ignorance of Language*, pp. 32, 73. To avoid confusion, I will continue to work exclusively with the “responding” terminology.

<sup>81</sup> Devitt’s official statement of his view explicitly appeals to this distinction: Human language comprehension routines “do not operate on metalinguistic *representations* of the syntactic and semantic properties of linguistic items but are *directly responsive*, in a fairly brute-causal associationist way, to these properties” (*Ignorance*: p. 275, emphases added).

say that in the context of the string ‘The wall is’, the adjective ‘red’ will be attached below the inflectional phrase node (IP) or the tense phrase node (TP). However, if ‘red’ appears at the outset of a matrix clause—as in, e.g., ‘Red, she painted it!’—then quite different processing operations must apply, so that that HSPM can ascertain the relevant attachment options and determine that ‘Red’ was, in this case, topicalized. Needless to say, there are many other possible ways of “responding to” an adjective like ‘red’. Because parsing is not a reflex, there is no unique context-independent response.

Now, Devitt claims that “[w]hatever clue enables this recognition [of, e.g., ‘red’ as an adjective] could simply trigger rules appropriate for adjectives: *the processor responds to it as an adjective without representing it as one*” (*ibid.*, emphases added). This, however, points to a different and, I believe, deeper problem for Devitt’s view. To see the problem, note first that the “clues” which enable the HSPM to “respond to” an acoustic blast as an adjective are not in the *immediate* stimulus. Such clues are typically “implicit” in the structure of the incoming phrase, or, when they are “explicitly” marked in the surface form of the sentence, they can (and often do) appear arbitrarily far back in the sentence, or even in the broader discourse context. Consider some examples of the syntactic relations that a stimulus expression might bear to the other items in the stimulus stream, awareness of which would enable the HSPM to correctly “respond” to it:

- the stimulus is modifying a verb (for, then it cannot itself be verb)
- the stimulus is governed by a referring expression (for, then it cannot itself be verb)
- the stimulus is bound by a quantifier (for, then it cannot itself be quantifier)

These clues all pertain to relations that are relatively abstract and, crucially, *non-local*. To know whether, e.g., the current stimulus is bound by a quantifier, one must have already “noted” the

presence of a quantifier in prior stimuli and “kept track” of its hierarchical relation to the current stimulus.<sup>82</sup>

Now, as the billiard ball example that I gave above illustrates, cases of mere response—as against cases of representation—involve the sufficient conditions of the response being *wholly present* in the moment at which the response is caused. The impact of the ball’s force, upon contact with the bumper, is not distributed over various discontinuous regions of time. Nor does the response of the bumper—i.e., to resist the ball and bounce it away—rely crucially on events arbitrarily far back in time. The immediate impact of the ball is sufficient for that response; prior context is irrelevant.

Imagine now a billiards table that, for whatever reason, keeps track of the number of times a ball hits a bumper, displaying the tally on a digital readout. The first time the ball hits a bumper, the readout displays ‘1’; the next time, ‘2’, and so forth. In this case, the response—i.e., the number displayed on the readout—is different each time. To explain *this* behavior, we posit representations. Why? Because the stimulus alone is not sufficient to explain the variety of the system’s behaviors—i.e., its flexibility. The system’s sensitivity to prior context is plainly making an important difference to its responses. The number on the display, or perhaps the mechanisms that produce it, are “keeping track,” as we say, of some aspect of the prior context—in this case, collisions involving the bumper. In more sophisticated terminology, something in the system “stores information about prior context in memory.” Whether one chooses commonsense or information-theoretic terminology, the upshot is the same: the system is, in effect, bringing *representations* of prior context to bear on its own subsequent behavior.

---

<sup>82</sup> Note, by the way, that not all quantifiers are explicit in the stimulus. To take just one example, plural nouns sometimes introduce implicit quantification. The universal claim ‘Lions are carnivorous’ differs from the existential claim ‘Lions are in the building’ precisely in respect of implicit quantification. (The example is due to Ernie Lepore.)

Return now to our observations regarding the HSPM. The “clues” that a system needs in order to interpret an incoming stimulus as, e.g., an adjective, are *not* “wholly present” in the stimulus. For, the stimulus is only a morpheme, but the relevant clues are either arbitrarily far back in time, or not overtly present in the stimulus stream at all (ff. 80, 83). Thus, something in the HSPM has to be “storing” or “keeping track of” these clues, so as to enable the HSPM to *use them as clues* later in time. And because the HSPM is demonstrably sensitive to such a *wide variety* of interacting contextual factors, we can conclude that it must be keeping track of a *great deal* of information. Crucially, it must also make *intelligent use* of that information, accurately discerning the interactions of the contextual factors themselves. This requires the HSPM to contain many *interacting* representations. We can think of each of these as a complex multi-track disposition, which bears in intricate ways on the processing of “nearby” representations.

The point becomes even clearer when we consider a commonsense example of high-level relational property—say, the property of being an uncle. Suppose, for a moment, that a theorist were to propose the hypothesis that mature adults do not *represent* other people as uncles, but rather merely *respond* to them as such. At first, it is tempting to protest that this theorist’s notion of “responding to something as having property P” is *just code* for “*representing* something as having property P.” But, as I urged above, we must resist this temptation. The responding/representing distinction is quite real. The theorist’s proposal is thus *not* incoherent or self-contradictory. It’s simply *false*.

For one thing, there is no unique or general way of “responding to someone as an uncle.” How one responds to an uncle (or even how one “ought to”) is a matter of what other attitudes one has, both about uncles and about other things. Moreover, the mere presence of an uncle in a person’s environment is not sufficient for that person to respond to the uncle “as such.” (We will

examine the import of the ‘as such’ locution momentarily.) Rather, the person must first “have in mind” certain *clues* that enable her to engage in whatever responses are “appropriate to uncles.” For instance, she must know that the target individual is male and that he has had a sibling who, in turn, has a living child. Plainly, these properties of a person are not wholly present in the immediate stimulus. Indeed, the clues about *them* are typically *also* presented arbitrarily far back in time, and hence must be “stored in memory”—i.e., mentally represented—in order to be useful for guiding uncle-appropriate responses. Once again, all of these points apply directly to the case of the HSPM. There are no unique responses to adjectives, and whatever responses we do observe on any given occasion *cannot* be accounted for by the mere presence of an adjective in the stimulus. A suitably large and properly organized class of background mental representations must be wheeled in to pick up the explanatory slack.

### §3.3.5 *The Distinction Between Responding-as and Representing-as*

I want now to develop a related but somewhat different line of argument against Devitt’s claim that the operations of the HSPM are mere brute-causal “responses,” unmediated by mental representations of syntactic structure. The argument rests on two observations, both of which I take to be rather uncontroversial. First, one and the same linguistic stimulus can bear any number of properties that might be relevant to sentence processing. Second, at any point in processing, the HSPM may register the instantiation of any one of those properties, *without registering others*. To take just one example of this ubiquitous phenomenon, consider a case discussed by J. D. Fodor (1989), in which an expression is both a noun phrase and an antecedent of an NP-trace.

(51) These are the children who<sub>i</sub> *wh-trace*<sub>i</sub> were only very rarely observed *NP-trace*<sub>i</sub> in the classroom.

In (51), the noun phrase ‘the children’ is an antecedent of the NP-trace that appears after the passive verb, ‘observed’. Fodor notes that sentences like (51) present special difficulties for the HSPM, because “the antecedents of NP-trace and PRO are typically normal argument NPs, which cannot, when they are encountered, be recognized *as* antecedents” (p. 197, emphasis on ‘as’ in the original). According to Fodor, when the passive verb in (51) is integrated into the mental phrase marker, the HSPM posits an NP-trace and proceeds to search for a suitable antecedent. As discussed in ch. III (§5), the search takes time and computational resources, as reflected in a distinctive ERP signature (Featherston, 2001), as well as slower response times on cross-modal priming tasks (Nicol and Swinney, 1989) and visual probe recognition tasks (MacDonald, 1989; McElree and Bever, 1989).

Reflect now on the import of Fodor’s use of ‘as’. In order to explain the processing lags mentioned above, she takes the HSPM to initially represent ‘the children’ *as* belonging to one syntactic category, viz., NP, but not *as* belonging to another, i.e., antecedent of the NP-trace. Moments later, of course, the HSPM cottons on to the fact that ‘the children’ belongs to latter category as well. Since the stimulus expression is the same, the only way to describe this transition is as a change in the *representational content* of the HSPM’s classification of that expression. For, viewing the HSPM as merely *responding* to the stimulus in a brute-causal fashion would require sacrificing the explanatory and predictive leverage that we gain from positing representations.

Devitt might suggest, at this point, that we must distinguish responding to a stimulus *as* a noun phrase from responding to that very stimulus *as* an antecedent. This would not, however, rescue the brute-causal model, for, in drawing *this* distinction, we abandon the notion of “brute-

causal response” that is central to that model. When the ‘as’ locution is found to be playing a significant role, the term ‘response’ is revealed to be simply code for ‘represent’.<sup>83</sup> To see this, consider a theorist who concedes that mature adults respond to one and the same individual sometimes *as an uncle* and sometimes *as a father*, but insists that these are mere brute-causal responses to the properties of being an uncle and being a father. The proposal is either false—because such responses are plainly mediated by representations—or it is incoherent, on account of there being no intelligible distinction between representing-*as* and responding-*as* (provided that the ‘as’ locution is making a genuine contribution to our inferences, predictions, and explanations). I suspect that Devitt’s confidence in the brute-causal model is based on a conflation of the responding/representing distinction, which is quite real, and the responding-*as*/representing-*as* distinction, which I have argued is illusory.<sup>84</sup>

### §3.3.6 *Is the Comprehension Process Sensitive to the Structure of Mental Phrase Markers?*

I take the considerations presented above to decisively establish that mental phrase markers are psychologically real. There is, however, one last move that Devitt can make in an effort to save the brute-causal model of language processing. As noted above, Devitt holds that brute-causal processes *sometimes* involve representations.

We note that representations may be involved in a brute-causal process. Thus the processes of Hull-Tolman behaviorism and of connectionism may involve representations as surely do those of many skills. *But these processes are not governed by rules that manipulate representations according to their syntactic structure in the rational-causal way.* (*Ignorance*, p. 56, emphasis added)

---

<sup>83</sup> Here, I echo Chomsky and Fodor’s most important criticism of Skinner: “as Skinner uses it, (at least when he’s outside the laboratory), ‘response’ is really a crypto-intentional term” (Fodor, 1991: p. 55).

<sup>84</sup> I hasten to note that the analogy between representing uncles (as such) and representing noun phrases (as such) breaks down in the long run. In ch. V, I argue that the latter kind of representation, unlike the former, is *subpersonal* and hence *not* a constituent of any personal-level propositional attitudes. Needless to say, the earlier analogy with a sophisticated billiards table is likewise limited.



To accommodate this point, I rendered the final aspect of his characterization of brute-causal processing as a disjunction. A process is brute-causal, I said, if its operations *either* do not involve representations at all, *or* if the rules governing that process are not sensitive to the syntactic properties of those representations. This raises the question: What reason do we have to believe that the comprehension process is sensitive to the syntax of mental phrase markers? If it turns out that there are no grounds for believing this, then perhaps there is still a sense in which comprehension can be said to be “brute-causal associationist.”

I believe that we have at our disposal a ready response to this point. Recall that Devitt maintains that the process of *reasoning* is sensitive to the syntax of thoughts. He endorses Fodor’s arguments for LOTH and CTM, which appeal to the productivity and systematicity of thought, and the inferential coherence of the process of reasoning. Following Fodor, Devitt argues that the only satisfactory explanation of the rationality of human reasoning is one that appeals to computations that are sensitive to the formal and syntactic properties of thoughts.<sup>85</sup> A moment’s reflection reveals that these considerations carry over *directly* to the mental phrase markers. For one thing, it is plain that mental phrase markers must be exactly as productive and systematic as public language, *for that is their domain*. It follows, then, that mental phrase markers are *themselves* structured entities. Moreover, in ch. V, we will discuss a variety of psychologically plausible computational models of parsing that demonstrate how the operations

---

<sup>85</sup> “I think that we should prefer LOTH [over the view that] a mental representation has some nonlinguistic complexity—for example, the complexity of a map. ... it is difficult to see how [the latter] could account for the mental process of thinking .... Formal logic gives us a very good idea of how thinking might proceed if thoughts are represented linguistically. From its very beginning, computer science has used this idea to build machines that do process linguistic representations. In recent years, computer science has developed “connectionist” machines that use representations of a very different sort, if they use representations at all. Despite the striking success of these machines with some forms of problem solving, connectionist processes seem rather far from capturing anything like human inference. We still have very little idea how thinking could proceed if thoughts were not language-like.” (*Ignorance*: 146-147)

of the HSPM can exhibit a subpersonal analogue of inferential coherence. The HSPM is, in a certain qualified sense, a “rational” device.

In §3.1, I cast doubt on the standard arguments for LOTH and CTM. However, I think that these arguments *do* work, when they are applied to mental phrase markers and parsing routines, rather than to thoughts and reasoning. The reader is invited to verify that my objections to these arguments lose much (or all) of their force when the conclusion is taken to be the psychological reality of structured mental phrase markers and structure-sensitive operations defined over these.

### §3.3.7 *What Does All This Tell Us About the Nature of Mental Representation?*

I would like to wrap up this section with some reflections on how the conclusions that we’ve reached thus far bear on broader philosophical issues. Our recent discussion offers an opportunity to clarify the very notion of representation that is here at play. In ch. V, we will address the question “What is a mental representation?” head-on. In the service of that discussion, let us be explicit now about the conditions under which mental representations are an appropriate and *explanatory* posit.

As many have noted, one of the main reasons for positing mental representations is that doing so allows us to explain the *flexibility* of a creature’s (or a mechanism’s) behavior.<sup>86</sup> A creature’s behavior is flexible when it is contingent upon a large variety of potentially interrelated factors, not all of which are wholly present in the stimuli that give rise to its

---

<sup>86</sup> See, e.g., Pylyshyn (1986), Sterelny (1991).

behaviors. Varying one of the factors may influence the creature's behavior in a particular way, but only on the condition that a great many other factors are kept constant (or also varied, but only in *very* specific ways). If pressed to explain why such contingencies hold in the case of many biological systems—birds, mammals, and the like—the only answer seems to be that such creatures have *internal representations* of each of the environmental factors on which their behavior is contingent. They are, in effect, systems that *store memories* of those factors.<sup>87</sup> This, combined with our earlier remarks about the 'as' locution, suggests that it is appropriate to posit representations in those cases where we wish to explain why one and the same stimulus gives rise to different responses, depending on which of its high-level, relational properties are relevant to the system's behavior. For such an explanation to be at all satisfactory, the representations that we posit must be dynamically responsive to changes in the environment and they must be in constant interaction with one another, giving rise to behavior *jointly*, not individually.

## Conclusion

In this chapter, my aim has been to reinforce the claim that constructing mental phrase markers is a necessary step in the process of language comprehension. To that end, I examined models, due to Schank and his colleagues, that attempted to get by with only semantic and pragmatic resources (§2). I argued that the motivations for such models are weak and, in some cases, ruled out by available psycholinguistic data. Then, following Marcus (1984), I argued that these models are unable to process syntactically complex constructions, including embedded clauses, multiple passivization, wh-questions, and anaphora. Although I did not argue for it here,

---

<sup>87</sup> Note that Devitt and I are in agreement on the need to posit representations even for quite simple creatures. He cites approvingly the following passage from Gallistel (1990): "bees must represent the angles and distances of food sources not only with reference to the sun but also with reference to prominent features of the terrain surrounding the hive" (p. 132). Devitt notes that "Gallistel's account of the extraordinary navigational skills of insects makes this plausible" (p. 21, fn. 5).

I believe that similar problems are likely to bedevil more recent attempts, on the part of connectionists and “constraint-satisfaction” theorists, to get by without mental phrase markers. It would be interesting to examine how these newer models propose to avoid repeating the mistakes that Schank *et al.* made several decades ago. I leave this project for future work.

In §3, I turned to a position advanced by Devitt (2006), which combines a controversial conception of thought with a commitment to minimizing the role of representations in language processing. My aim was to cast doubt on both pillars of this position.

I began by suggesting that the standard arguments for LOTH are rather weak. The productivity argument relies on a wholly unsubstantiated assumption to the effect that the optimal grammar of thought is recursive, like that of natural language. The systematicity argument rests on the assumption that natural language exhibits a property that, for quite some time, was characterized only informally. We saw that more rigorous definition, due to Johnson (2004), reveals that natural language may not, after all, be systematic.<sup>88</sup> Finally, the argument from inferential coherence assumes that rational inference mirrors the syntax-driven transformations familiar from formal logic. Following Sellars and Harman, I suggested that this is a misleading picture of practical and theoretical reasoning—one that we would do well to abandon.

Turning to the *strengthened* version of LOTH, which I labeled ‘SLOTH’, I argued that there are additional difficulties. Devitt’s abductive argument for SLOTH rests on a putative explanation of the syntax of public language in terms of the syntactic properties of thought. I pointed out that this argument relies on a contentious view about the explanatory priority of thought over language, which has been forcefully challenged—and, I believe, effectively

---

<sup>88</sup> For a similar line of reasoning, see Cummins (1996b).

refuted—by Sellars (1958) and Rosenthal (1986/2005). Second, the explanation of the syntax of public language in terms of the syntactic properties of thought is not the only game in town. An alternative explanation, which Devitt occasionally seems to endorse, is both available and arguably superior.

Third, I pointed out that an explanation of the syntax of public language in terms of the syntactic properties of thought is in tension with Devitt's conception of language processing as brute-causal—i.e., blind to the syntactic properties of the representations that are involved. If language production—i.e., the process of translating from thoughts to speech acts—is brute-causal in this sense, then it is blind to the syntactic properties of the thoughts that are its inputs. And if that is so, then it's difficult to see how thoughts can “confer” their syntactic structures onto the sentences used in speech, or even to explain them. For, the thoughts could have entirely *different* structure than the sentences—e.g., the structures proposed by Schank and Birnbaum (1984) or Rohde (1999)—and yet give rise to those sentences just as quickly and automatically. Indeed, if the second argument that we examined from Pinker (1994) is cogent, then the structure of thought *is* different from that of public language. But, even if that argument does not succeed, I believe that we have no grounds, at present, to endorse the claim that thoughts are syntactically structured in a way that resembles public-language expressions.

In §§3.2 and 3.3, I argued that the best models of the parsing and comprehension process posit mental phrase markers. My argument rested on two observations. First, unlike the rigid reflexes of the Sphex wasp—the primary piece of theoretical machinery in the behaviorist toolkit—correctly parsing an input string requires keeping track of prior context. Moreover, the clues that are relevant for a successful parse may be arbitrarily far back and in discontinuous regions of the stimulus stream. When a system is responsive not only to the stimulus that is

*currently* impinging on it, but also to contextually relevant factors, we posit representations to explain the flexibility of its behavior. On these grounds, I argued that we should see the internal states of the HSPM as representations. This conclusion was reinforced by a second observation, namely, that clear, satisfying, and, as far as I know, *unrivaled* explanations of both first-pass parsing and reanalysis routinely posit internal states that are responsive to stimuli *as* bearing one high-level relational property rather than another. I argued that responding-*as* is the hallmark of a representational system, not a substantive alternative.

Finally, I argued that the representations employed by the HSPM must themselves have constituent structure, which mirrors that of public-language sentences. Applying the aforementioned productivity, systematicity, and inferential-coherence arguments, I concluded that the computational operations defined over mental phrase markers are sensitive to their internal structure. In the next chapter, I will examine the precise character of these operations, by surveying the computational models of language comprehension that have been advanced over the past several decades. My goal will be to show that some of those models are psychologically plausible and that they are committed to the representation of syntactic rules and principles. We will explore to what extent the success of these models can be used to bolster the Representational Thesis (RT), and, more generally, to shed light on the psychological import of syntactic theory.



## V

# A HISTORICAL SURVEY OF COMPUTATIONAL PARSING MODELS

---

### §1 Introduction

Our goal has been to explore the psychological import of syntactic theory. We have, thus far, found that the syntactic structure of linguistic expressions plays an important role in language processing. The previous two chapters constitute a sustained defense of the psychological reality of mental phrase markers (MPMs). In ch. III, we saw that ERP studies, structural priming experiments, and research into garden-path and filler-gap processing all point to the same conclusion: A psychologically plausible comprehension model must, at some stage in its processing, construct explicit representations of the syntactic structure of linguistic input. This conclusion was further supported in ch. IV, where we considered the prospects for the “mostly-semantics” models developed by Schank and his associates at Yale (Schank and Birnbaum, 1984), as well as the “brute-causal” model tentatively advanced by Devitt (2006). I argued that both types of model face difficulties, which can be traced to their eschewing MPMs for the purposes of parsing and comprehension.

Establishing the psychological reality of MPMs is important, but it does not get at what many see as the core of the debate over the psychological reality of syntax. For, what is ultimately at issue is whether, and in what sense, *the rules or principles of a grammar* are psychologically real. This is the issue I propose to explore here. The naturalistic methodology of chs. III and IV counsels us once again to look at the available empirical results and the viable models of language processing. In the pages that follow, we conduct a historical survey of the computational parsing models that have some claim to psychological plausibility. The bulk of this chapter (§§2-6) is devoted to laying out the details of a variety of promising proposals in



computational linguistics. In the final section (§7), I will explore the extent to which our best models support the Representational Thesis (RT), which Devitt (2006) formulates as follows:

**(RT)** A speaker of a language stands in an unconscious or tacit propositional attitude to the rules or principles of the language which are represented in her language faculty.

There is, I think, something to be said in favor of a suitably modified version of RT, which eschews reference to the personal-level phenomenon of propositional attitudes. As I will make clear in §7, the human sentence processing mechanism (HSPM) is a subpersonal computational system, whose internal states are not usefully assimilated to propositional attitudes. Nevertheless, while I am confident that syntactic rules or principles are psychologically real in an important sense, we will see in §§6-7 that the case for RT is rather weak. We have, at present, only rather tenuous grounds for thinking that grammars are explicitly represented in a separate data structure in the human brain. Still, the sense in which grammars *are* psychologically real is perhaps robust enough to sustain some of the more modest rhetoric in generative linguistics. As the discussion below illustrates, we have strong reasons to think that the HSPM *embodies* the rules or principles of a grammar, perhaps in the form of *procedural dispositions*.<sup>1</sup>

Why think of embodiment as a species of psychological reality? Most centrally, because embodied rules or principles share important features with explicitly represented ones, in respect of the roles that they play in our explanatory practices. In particular, just like explicit

---

<sup>1</sup> I discuss various notions of procedural rules in §4 and §7. For the moment, to fix ideas, we can point to the ACT-R system, discussed by Vasishth and Lewis (2006), as an example of a procedural implementation of a grammar. The system treats MPMs as declarative data structures, represented explicitly in the system's memory buffers. By contrast, it treats grammatical rules as "procedural knowledge," which is "represented as production rules— asymmetric associations specifying conditions and actions. Conditions are patterns to match against buffer contents, and actions are taken on buffer contents. All behavior arises from production rule firing; the order of behavior is not fixed in advance but emerges in response to the dynamically changing contents of the buffers" (p. 414). Though the word 'representation' is used, Devitt (2006) argues that it plays no useful role (§7). Likewise, ATN parsers can be thought of as embodying "condition-action" (or "pattern-action") procedural rules (§4). See also Marcus (1980).

representations, embodied rules/principles figure in algorithmic descriptions and explanations of the activities of the HSPM—pitched at David Marr’s second level of description (Marr, 1980). Moreover, we can appeal to embodied rules/principles in issuing predictions about the observable effects of these activities, in the absence of detailed knowledge of their physiological underpinnings. This powerful mode of description, prediction, and explanation typically does not require that we appeal solely to representations; in many cases, positing embodied rules and principles can serve just as well to illuminate the structure and function of the HSPM.

### *§1.1 A Preliminary Look at the Rules that Govern Language Processing*

The discussion that follows can be characterized as an exploration of what Devitt (2006) calls the “processing rules” of a comprehension system.<sup>2</sup> To clarify this technical notion, it will be helpful to begin with a general “anatomy” of language processors, due to Steedman (2000).

All language-processors can be viewed as made up of three elements. The first is a grammar, which defines how constituents combine to yield other constituents. The second is an algorithm for applying the rules of the grammar to a string. The third is an oracle, or mechanism for resolving nondeterminism. The oracle decides which rule of grammar to apply at points in the analysis where the nondeterministic algorithm allows more than one rule to apply. [fn. 1: The division of processing labor between a nondeterministic algorithm and an oracle is not always made explicit, particularly in implementations. However, all processors can be viewed in this way. A more extensive survey of the parsing literature in terms of a similar tripartite architecture can be found in Kay 1980.] (pp. 226, 280)

What Steedman calls ‘the grammar’ is a set of *structure rules*, which govern the language that constitutes a particular system’s domain. Talk of “applying rules” reflects Steedman’s commitment to what Bresnan and Kaplan (1982) called the “Strong Competence Hypothesis,”

---

<sup>2</sup> As with previous chapters, I focus exclusively on language comprehension, putting aside language production. This is largely on account of the fact that it’s unclear to what extent the production process must deal with ambiguity in the input. By contrast, it is *abundantly* clear that the resolution of ambiguities is a central feature of the comprehension process. My argument is that both mental phrase markers and (perhaps implicit) representations of syntactic principles play a crucial role in our explanations of how the HSPM can fare as well as it does in the face of ambiguity.

which Steedman formulates as follows: “the grammar that is used by *or implicit* in the human sentence processor is the competence grammar itself” (226, emphasis added).<sup>3</sup> Whether Steedman is committed to RT depends a great deal on what he means by a grammar’s being “implicit in the human sentence processor” (*ibid.*) As noted above, we must distinguish between a grammar’s being represented and its being embodied. If by ‘implicit’, Steedman means ‘embodied’, then taking a grammar to be implicit in the parser does not amount to a commitment to RT.

We cannot, in the present context, help ourselves to the claim that the structure rules of natural language serve also as processing rules that govern the HSPM. (We will examine arguments for this claim in §§3-6.) Thus, in order to proceed with our discussion of *processing* rules, we must put aside the grammar for the time being,<sup>4</sup> and focus on what Steedman calls “the algorithm” and “the oracle.” Let us begin with the latter.

### §1.2 *The Oracle*

Every parsing model that seeks to achieve psychological plausibility must ultimately face the problems engendered by the massive ambiguity of the input—a point that has come up repeatedly in our discussion of the evidence for MPMs. Steedman characterizes the oracle as that aspect of the processor that “decides which rule of grammar to apply at points in the analysis where the nondeterministic algorithm allows more than one rule to apply” (*ibid.*) Let us look at three competing accounts of how the oracle accomplishes this task.

---

<sup>3</sup> Steedman (2000) actually makes a stronger commitment: “It is important to note that the strong competence hypothesis as stated by Bresnan and Kaplan imposes no further constraint on the processor. In particular, it does not limit the structures built by the processor to fully instantiated constituents. However, the Strict Competence Hypothesis proposed in this book imposes this stronger condition” (228).

<sup>4</sup> We will return to the notion of a structure-rule in §1.4 and examine concrete examples throughout this chapter.

In ch. III, we encountered several parsing principles that emerge from recent psycholinguistic research: Minimal Attachment, Late Closure, and the Minimal Chain Principle.

- (MA) The parser will attach incoming material into the existing mental phrase marker in such a way as to minimize the number of nonterminal nodes in the resulting structure.<sup>5</sup>
- (LC) The parser will incorporate newly encountered material into the most recent phrase or clause of the mental phrase marker that it has already constructed, thus preferring attachments lower in the tree over attachments higher in the tree. (cf. the principle of Right Association in Kimball, 1973)
- (MCP) The parser will (i) avoid postulating unnecessary chain members (e.g., gaps/traces) and, (ii) having identified an item in the input as an antecedent, it will posit a gap/trace in the very first position at which it is licensed by the grammar.<sup>6</sup>

It is important to be clear about the status of MA, LC, and MCP. To my knowledge, it has never been suggested that these principles are *represented* in the HSPM. Rather, these are intended to be *descriptive* principles; they are *true of* the HSPM in much the same way that the principles of

---

<sup>5</sup> This formulation of MA follows the original proposal in Frazier (1978) and Fodor and Frazier (1978). Later formulations depart from this. For instance, McRoy and Hirst (1990) write: “MA has been defined as a preference for attachments that require adding the fewest nodes to attach a new word to the current structure; but ... this definition is too dependent on assumptions about the underlying grammar. In addition, this definition can also make the system’s account of the interactions among MA and other preferences implausibly rigid (Church, 1980; Wanner, 1980; Wilks et al., 1985) or force the system to delay attachment decisions (Shieber, 1983), which is also implausible. ... Thus, we use a more general statement of MA: The attachment that can be computed most quickly (i.e., with lowest time cost) is the one that will be favored. ... [T]o find the most minimal attachment of a constituent C is to find (in decreasing order of preference): 1. A place along the right edge of the current parse tree where C is expected (or a place where C expects the current parse tree); 2. A place along the right edge of the current parse tree where C is allowed; 3. A place along the right edge of the parse tree where C could be attached by first proposing some intermediate structure such as a phrase or a clause” (323). This definition requires an independent specification of which computations have the “lowest time cost.” McRoy and Hirst (1990) characterize the computational cost of an operation by listing a number of empirically confirmed generalizations: “Lexical information is faster to identify than syntactic information, presumably because lexical retrieval is easier than tree traversal. Simpler attachments are faster to build than more complex ones, because it takes more time to build more structure. More local hypotheses are faster to recognize than more distant ones, because it takes more time to traverse more of a parse tree. Syntactic information is faster to access than semantic information, because tree traversal is easier than knowledge-base inference (cf. Ferreira, 1986; Ferreira & Clifton, 1986). Possibilities that have been primed, either by semantics or syntax, are faster to identify and build than unprimed ones (Frazer, Taft, Roeper, Clifton, & Ehrlich, 1984; Hirst, 1987). Possibilities that are more highly expected are faster to identify than ones that are less expected” (322).

<sup>6</sup> This formulation is a paraphrase of the one found deVincenzi (1991). The second conjunct is equivalent to what Frazier and Clifton (1989) refer to as the “Active Filler Hypothesis.” I use the locution ‘gap/trace’ to avoid commitment to grammatical formalisms that have traces and movement operations in their theoretical toolkit.

celestial mechanics are *true of* our solar system. Moreover, admitting the descriptive accuracy of such principles does not automatically commit one to any view regarding RT—at least not without additional argument.<sup>7</sup>

I will refer to MA, LC, and MCP, collectively, as the *least-effort parsing principles*, for they reflect a tendency on the part of the HSPM to adopt whichever syntactic analysis requires expending the fewest computational resources.

In general, ‘garden-path’ explanations of processing difficulty account for asymmetries between sentences where pure memory load explanations do not. Our own model is of the garden path variety. The parser chooses to do whatever costs it the least effort; if this choice turns out to have been correct, the sentence will be relatively simple to parse, but if it should turn out to have been wrong, the sentence will need to be reparsed to arrive at the correct analysis (Frazier and Fodor, 1978: p. 295).

Appealing to least-effort parsing principles has allowed psycholinguists to derive both clear predictions and satisfying explanations of a wide range of ambiguity resolution preferences—hence, of a great deal of behavioral and neurocognitive data (ch. III). It will come as no surprise, then, that versions of MA, LC, and MCP are implemented in many of the models that we will examine below. These principles constitute the “oracle” in such models.

As is well-known, however, there are alternatives to this “resource-based” approach to ambiguity resolution. One competing approach can be characterized as “probabilistic” or “frequency-based”. Theorists who opt for this approach claim that the problems posed by ambiguity in linguistic signals can be met by a system that takes into account lexical, syntactic,

---

<sup>7</sup> J. D. Fodor and L. Frazier (1978, 1980) do supply an additional argument for RT. Fodor and Frazier claim that their “Sausage Machine” parsing model provides a *principled* explanation of why MA and LC are true of the parser, precisely in virtue of the model’s commitment to representing the grammar of a language in a separate data structure. They point out that the competing ATN models, which implement the grammar “procedurally,” can build in such principles only in an *ad hoc* way, if at all. From this, they conclude that it “seems unavoidable that *the well-formedness conditions on phrase markers are stored independently of the executive unit*, and are accessed by it as needed.” (Frazier and Fodor, 1978: 322n). We will return to this argument in §4. (See also fn. 10 below.)

and semantic frequency information, derived from an annotated corpus, such as the Penn Treebank.<sup>8</sup> It has also been hypothesized that such frequencies are psychologically real, in that they have an explanatory relation (as yet not fully understood) to the ambiguity resolution principles employed in human language comprehension.<sup>9</sup> The core idea is that, when faced with a lexical or structural ambiguity, the HSPM will prefer the “most probable” analysis—i.e., the analysis that has occurred most frequently in prior input.

For our purposes, it is important be clear on the fact that a parser whose oracle consists of frequency-based parsing principles must still incorporate a grammar in one way or another. We will see examples of probabilistic grammars in §2. For now, let us simply note that the issues pertaining to RT remain the same regardless of whether one opts for the resource-based or the frequency-based approach to ambiguity resolution. If the HSPM contains an explicit representation of a grammar, then RT is true. If, on the other hand, the grammar is merely *embodied* in the HSPM, then a weaker claim is all that can be sustained.

Weinberg (1999) offers a helpful taxonomy of the various approaches to modeling human parsing preferences. She contrasts the two types of “oracle” that we have discussed thus far with a third, which ties the principles of ambiguity resolution much closer to a competence grammar.

---

<sup>8</sup> MacDonald, Pearlmutter, and Seidenberg (1994); Filip, Tanenhaus, Carson, Allopena, and Blatt (2002); Rohde (2002); Jurafsky (2003); Jurafsky and Martin (2007: ch. 14).

<sup>9</sup> See Rohde (2002) and Jurafsky (2003) for discussion of the daunting problems associated with frequency-based approaches. While it’s clear that frequency information plays some (perhaps significant) role in comprehension, it’s not at all clear *which* frequencies matter—i.e., which factors the HSPM is keeping track of. The space of logical possibilities is vast. Frequencies can be defined over quite coarse-grained properties, like having an N-V-N structure. On the other end of the spectrum, it has been suggested that the HSPM keeps track of lexical frequencies defined over extremely fine-grained, *recherché* categories, hobbled together from semantic and pragmatic properties. To take a typical case, Filip, Tanenhaus, and Carson (2002) propose that frequency information about the distribution of “proto-Patient and proto-Agent properties” plays a role in language comprehension. They define *Proto-Patient* and *Proto-Agent* as follows: “*Proto-Agent*: Volitional, sentient causers of the event described, who move and exist independently of the verb described (*sic*). *Proto-Patient*: Undergo a change of state, serve as incremental themes, are causally affected by another event participant, are stationary relative to another event participant, and do not exist independently of the event described” (*ibid.*) The reader will have no trouble noticing that these are some very fine-grained categories.

Research in the theory of human sentence processing can be characterized by three styles of explanation. Researchers taking the first track have tried to motivate principles of structural preference from extralinguistic considerations like storage capacity in working memory, or bounds on complexity of incremental analysis. Frazier and Rayner's (1982) Minimal Attachment and Right Association [i.e., Late Closure] principles, and Gorrell's simplicity metric, are examples of this type of theory. The second track eschews "parsing strategies," replacing them with a fairly complex tuning by speaker/hearers to frequency in the hearer's linguistic environment. The difficulty of recovering an analysis of a construction in a particular case is a function of how often similar structures or thematic role arrays appear in the language as a whole. The work of Trueswell et al (1994), Jurafsky (1996) and MacDonald et al (1994) are examples of frequency or probability based constraint satisfaction theories. The third track takes a more representational view and ties processing principles to independently needed restrictions derived from competence and language learning. This approach claims that the natural language faculty is extremely well designed in the sense that the same set of principles that govern language learning also contribute to a theory of sentence processing. This track is represented by the work of Gibson (1981), Gorrell (1995) Pritchett (1992), Philips (1995, 1996) and Weinberg (1992), who argue that processing can be seen as the rapid incremental satisfaction of grammatical constraints such as the Theta Criterion, which are needed independently to explain language learning or language variation. (p. 283)

Weinberg goes on to develop a theory of the third type, arguing that

a particular version of the Minimalist Program (Chomsky, 1993; Uriagereka, 1998) provides principles needed to explain both initial human preferences for ambiguous structures and provides a theory of reanalysis, explaining when initial preferences can be revised given subsequent disconfirming data, and when they lead to unrevisable garden paths.

We will explore Weinberg's proposal in §6. The interest of the grammar-based approach that she favors consists in its potential to unify our understanding of the structure-rules of natural language with both the parsing algorithm and the ambiguity-resolution principles of the HSPM (i.e., the "oracle"). In addition, her proposal has a strong bearing on the psychological reality issue, suggesting that the grammar of a language is psychologically real in some very interesting sense. However, we must take care not to overstate the consequences of her proposal for RT.

Devitt (2006) lays out several positions on the psychological reality issue, of which the following three will be most relevant to our discussion in this chapter. (We will return to these frequently.)

- (i) The structure rules of a language—i.e., the syntactic rules or principles that comprise a grammar—are also processing rules that are *represented* in the speaker and *applied* in language use.
- (ii) The structure rules of a language are represented and used *as data* by the processing rules of language use.
- (iii) The structure rules of a language are also processing rules for language use but the processing rules are *merely embodied* and *not* represented.

Devitt casts doubt on positions (i) and (iii), arguing that the rules of a grammar cannot plausibly serve as the processing rules that govern the HSPM. Weinberg's model—and the other models that fall into the grammar-based approach—provides some reason to believe that the structure-rules of a language may, after all, be of the right sort to govern language processing. Her proposal, as well as those of Pritchett (1992) and Philips (1996), illustrates the advantages of the psycholinguists' shift in focus toward the Principles and Parameters formalism (§5). In this sense, their contributions make real progress in the psychological reality debate, by putting positions (i) and (iii) back in the running, at least for the time being.

Unfortunately, however, this alone does not settle the question of RT. For, while the theorists who adopt the grammar-based approach to ambiguity resolution have provided evidence in support of *either* position (i) *or* position (iii), nothing in their accounts suggests that position (i) is the superior of the two. That is nothing in their proposals suggests that the structure-rules of a language are explicitly *represented* in the minds of competent speakers. For all that such theorists have shown, these rules may very well be only *embodied* in the HSPM,



constituting procedural pattern-action dispositions (fn. 1). Thus, while these theorists have bolstered the case for positions **(i)** and **(iii)**, it cannot be said that they have established RT.<sup>10</sup>

We began this discussion with the aim of clarifying the notion of a processing rule. We have now considered three competing proposals regarding one kind of processing rule—viz., the kind that constitutes what Steedman (2000) calls the “oracle” in a language processing system. Whether they appeal to extralinguistic computational constraints, frequency information, or independently-motivated grammatical principles, the function of the processing rules (or “parsing principles”) that constitute an oracle is to govern a parser’s behavior in the face of ambiguity. Let us now turn our attention to a second type of processing rule, which constitutes what Steedman (2000) calls “the algorithm” of the parser.

### §1.3 *The Algorithm*

Computational linguists sometimes complain that psycholinguists’ proposals concerning the operations of the HSPM are imprecise, in that they don’t make clear the underlying computational architecture. Consider, for instance, the following remarks from Harkema (2001).

The psycholinguist selects a set of sentences with grammatical constructions that are predicted to engage the human sentence processor in a way that will reveal its inner workings, designs an experiment with these sentences and runs it with a number of human subjects, interprets the results and proposes certain properties that the human sentence processor should have in order to account for the experimental results. The results of this approach greatly expand our knowledge about the human sentence processor, but it sometimes fails to satisfy linguists of a more formal bent, because a mere description of its properties does not constitute a complete specification of a

---

<sup>10</sup> The reader may be curious about position **(ii)**, which is also committed to RT. This position is best represented by Lyn Frazier and Janet Dean Fodor—two of the main proponents of what I earlier called the “resource-based” approach to ambiguity resolution. In a telling passage, they write: “when making its subsequent decisions, the executive unit of the parser refers to the geometric arrangement of nodes in the partial phrase marker that it has already constructed. It then seems unavoidable that *the well-formedness conditions on phrase markers are stored independently of the executive unit*, and are accessed by it as needed. That is, the range of syntactically legitimate attachments at each point in a sentence must be determined by a survey of the syntactic rules for the language, rather than being incorporated into a fixed ranking of the moves the parser should make at that particular point...” (Frazier and Fodor, 1978: 322n, emphasis added). We will examine their argument for this very attractive view in §4.

parser. In most psycholinguistic proposals, the operations of the parser are only specified insofar as they bear on the natural language constructions being explored. Moreover, this specification is usually very informal. While the parser works for the set of examples provided, absence of a description of the overall architecture of the parser leaves open the question whether a parser with the desired properties covering the entire language actually exists. Another drawback of some of the work in psycholinguistics is that it is sometimes based on simplistic and outdated conceptions of syntactic structure. (pp. 2-3)

To avoid falling prey to such concerns, in what follows we will have to go beyond ambiguity resolution principles, and consider processing rules that are pitched at a more basic level—what one might call the “algorithmic level,” following Marr (1980). In doing so, our primary goal will be to examine the ways in which psychologically plausible parsing models make use of the structure rules of various grammars—some “outdated” grammars but also some recent ones.

After a brief overview of the issues that dominate the sentence-processing literature, Devitt (2006) says the following:

The main debate in the literature we have been summarizing is about the *interpretative clues* that the hearer gets and the order in which they are used.<sup>[11]</sup> There is much talk about knowledge and information but little attention to the way in which this knowledge and information is embodied in a hearer; see, for example, Tanenhaus and Trueswell 1995 and Pickering 2003. And, we should note, that there is much about the choice between interpretations that are possible given the language, but little about how linguistic competence makes those interpretations available for choice in the first place. Before there can be a resolution of a lexical or syntactic ambiguity, competence must generate some candidate interpretations. (p. 238, emphasis added)

I am largely in agreement with Devitt’s claim that “there is little attention [in the literature] to the way in which [the grammar] is embodied in a hearer.” Psycholinguists and computational

---

<sup>11</sup> In chs. III and IV, we examined the debate about what Devitt calls “interpretive clues”—specifically, about the order in which they are accessed in processing. I argued that the relevant empirical research has shored up a wealth of behavioral and neurocognitive evidence for so-called “syntax-first” models. Recall that, on these models, only syntactic processes are at play in the earliest stages of comprehension; their output is made available, only after a brief delay, to processes that make use of semantic and pragmatic information. Moreover, *even if* further evidence about the relative timing of these processes turns out to be disconfirming to syntax-first models, the arguments of the previous two chapters nevertheless leave very little room for doubt about the fact that the HSPM constructs representations that encode distinctly syntactic information—what I have called ‘mental phrase markers’ (MPMs).

linguists make heavy use of the idea that syntactic rules or principles are *consulted* from the very outset of the parsing process. It is commonly held that constructing a mental phrase marker, let alone revising one in cases of local error, requires drawing on an “internal representation” of a grammar—that there is simply no way of building accurate mental phrase markers without making use of such information. But, oftentimes, when the focus of a theorist’s discussion is the HSPM, rather than an artificial model implemented on a general-purpose computer, the notions of “consultation,” “drawing on,” and “making use of” are either assumed (without argument) to carry over directly from the computational model, or characterized only informally.

Following Stabler (1983), Devitt makes the following point: Although the algorithms implemented on stored-program computers typically have a separate data structure that contains an explicit representation of the grammar, it is well-known that such algorithms can be “hardwired”—i.e., embodied in a special-purpose circuit, whose internal organization *directly implements* the algorithm, rather than explicitly representing its instructions as data (in either declarative form or as procedural imperatives). Devitt argues that such “hardwired” systems are much more plausible as models of the HSPM, for we have no grounds for thinking that the human brain implements a parsing algorithm in the way that a desktop computer implements programs like *iTunes*, *Photoshop*, and *Excel*. In §7, we will examine the notion of hardwiring and contrast it with two notions of procedural rules.

Before we can explore all of these issues, we must first get a clear understanding of the specific algorithms under discussion. Devitt is right to point out, in the passage quoted above, that “[b]efore there can be a resolution of a lexical or syntactic ambiguity, competence must generate some candidate interpretations” (*ibid*). In the pages that follow, we will explore parsing models that make explicit claims about how such interpretations are generated, and what role the

grammar plays in this process.<sup>12</sup>

Given that any characterization of a parsing model must, in one way or another, make reference to a grammar, we can categorize such models (admittedly somewhat crudely) on the basis of the grammars that they were designed to embody. The grammars that receive extended discussion in this chapter include recursive phrase structure grammar (and probabilistic extensions thereof), the Standard Theory of transformational grammar, ATN grammar, Government and Binding theory, and, finally, grammars in the Minimalist tradition.<sup>13</sup> At the outset of each section, I will provide an introduction to the basic theoretical machinery of each of these formalisms. My discussion of their descriptive and explanatory resources will, perforce, be rather brief. It is not my aim to motivate the inclusion of any particular type of structure rule in a contemporary syntactic theory. I seek only to illustrate several important distinctions between various types of structure rule, so that we may go on to evaluate the prospects of including any of them in a realistic account of human sentence processing.

In connection with context-free grammars, I discuss the top-down, bottom-up, and left-corner parsing algorithms. These form the foundation of a wide range of parsing models—arguably the majority. §2 provides an overview of the algorithms that make use of dynamic programming techniques—specifically, the famous Earley and CYK algorithms. I emphasize the fact that these algorithms, when implemented in general-purpose computers, draw on explicit

---

<sup>12</sup> All of the models discussed in this chapter do, of course, go on to make further claims about the computation of *semantic* representations. As in previous chapters, I restrict my focus (in large part) to syntax.

<sup>13</sup> Constraints of space preclude a fuller discussion of a number of formalisms that are currently popular in parsing theory. These include the lexicalist grammars that rely on feature unification—e.g., Lexical Functional Grammar (Bresnan, 2001) and Head-driven Phrase Structure Grammar (Pollard and Sag, 1994)—as well as Tree-Adjoining Grammar (Schabes, Abeille, and Joshi, 1988) and Combinatory Categorical Grammars (Steedman, 2000). Throughout the discussion, I will occasionally mention these, but I reserve a detailed treatment for future work. I do not think that the philosophical conclusions pertaining to the Representational Thesis (RT) are in any way affected by this omission. Moreover, it must be noted that feature unification strongly resembles the *Merge* operation, which plays a central role in the Minimalist grammars that we discuss in §6.

representations of grammatical rules.<sup>14</sup> I then discuss how the Earley and CYK algorithms have been applied to probabilistic extensions of context-free grammars, with impressive gains in both efficiency and psychological plausibility.

§2 concludes with a discussion of an exciting development in parsing theory, known as the Parsing as Deduction (PAD) approach. Here, syntactic analysis is seen, quite literally, as a species of *natural deduction*, in first-order quantificational logic. Parsing models that implement this approach run through an explicit proof procedure that takes the rules of a grammar as axioms and derives theorems concerning input strings. The PAD framework has been shown to subsume the Earley and CYK algorithms and to be applicable to a wide array of syntactic formalisms. It also constitutes the most concrete implementation of the core idea of position (ii)—viz., that structure-rules are drawn on *as data* in the course of comprehension. Models that adopt the PAD approach treat rules as *truth-evaluable claims*, which the parser uses as *premises* in the course of its deductive procedures. The success of such models substantiates my claim in ch. IV (§3.3.5) that the HSPM can exhibit a subpersonal analogue of inferential coherence.

Context-free rules form the base of early versions of transformational grammar. In the Standard Theory (Chomsky, 1957), the output of these rules is fed into a component of the grammar consisting of *transformational* rules. After introducing transformations in §3, I discuss the experiments that Fodor, Bever, and Garrett (1974) interpreted as shedding doubt on the psychological reality of transformational operations. Following Berwick and Weinberg (1984), I conclude that Fodor, Bever, and Garrett's argument is unsuccessful. Nevertheless, we will see that parsing models based on the Standard Theory were found to be unworkable on purely computational grounds.

---

<sup>14</sup> As noted above, how such algorithms might be implemented in the human brain is a separate question.

In the wake of these findings, the framework of Augmented Transition Networks (ATNs) rose to prominence in the 1970s. The ATN formalism offered a promising alternative for describing the surface syntax of natural language. Moreover, it suggested a natural implementation of the grammar as a parsing model—a top-down search procedure over the space of syntactically possible structures. In §4, we will examine ATN parsers, which serve as a clear example of how grammatical rules can be seen as procedural dispositions. §4 concludes with a discussion of the psychological plausibility of ATN parsers. As we shall see, a strong criticism of the ATN architecture, due to Lyn Frazier and Janet Dean Fodor, relies heavily on the fact that ATNs do not explicitly represent the rules of a grammar in a separate data structure. Frazier and Fodor’s argument illustrates the substantive explanatory payoff that a model might derive from a commitment to RT.<sup>15</sup>

Next, in §5, I turn to a discussion of principle-based parsers—models whose internal operations “transparently implement” the principles of Government and Binding (GB) theory. Principle-based models constitute a particularly exciting application of the Parsing as Deduction (PAD) approach mentioned above. The GB principles are formalized as logical constraints on the syntactic analysis of linguistic input. I show that modern principle-based parsers are quite sophisticated and efficient, offering a number of advantages over traditional rule-based systems.

In recent years, the Principles and Parameters approach has moved away from the GB formalism. Seeking a more elegant and compact account of syntactic structure, Chomsky (1995, 2001) and others have developed a variety of Minimalist grammars, in which the basic operations of Merge and Move (*alt.* external and internal Merge) serve to forge hierarchical

---

<sup>15</sup> Historically, the exploration of ATN parsers gave way to a research program launched by Marcus (1980) into the capabilities of *deterministic* parsing models. I will not, however, discuss deterministic parsers at any length. Despite their intrinsic interest, I do not see that they shed any new light on the Representational Thesis. In addition, there are compelling grounds for rejecting both the original Marcus parser and its D-theoretic offspring as viable models of human parsing. (See Pritchett, 1992: pp. 44-51). I therefore omit a detailed discussion of these models.

relations between lexical items, conceived as sets of features. Unsurprisingly, computational linguists have undertaken the task of building parsing models that incorporate these grammars. In §6, I examine these developments, focusing on the algorithms developed by Harkema (2001). Harkema employs the PAD approach in constructing bottom-up, top-down, and mixed (Earley) algorithms for parsing with Minimalist grammars. I also examine the proposal of Weinberg (1999), which attempts to account for a number of psycholinguistic results by viewing parsing as “the incremental satisfaction of grammatical constraints” imposed by Minimalist grammars.

#### *§1.4 The Grammar, First Pass: Context-Free Phrase Structure Grammars*

Let us now flesh out more carefully what is meant by a *structure rule*. Devitt (2006) characterizes structure rules as follows:

The outputs of a linguistic competence ... are governed by a system of rules, just like the outputs of the chess player, the logic machine, and the bee. Something counts as a sentence only if it has a place in the linguistic structure defined by these structure rules. Something counts as a particular sentence, has its particular syntactic structure, in virtue of the particular structure rules that govern it, in virtue of its particular place in the linguistic structure. Like the theory of the idealized outputs of the chess player, logic machine, and bee, our theory can be used to make distinctions among the nonideal. Strings that are not sentences can differ in their degree of failure. For they can differ in the sort and number of linguistic structure rules that they fail to satisfy. (Devitt, 2006: p. 24)

The parsing and comprehension models that psycholinguists and computational linguists have developed differ along a number of dimensions. One important difference has to do with the types of structure rules that they incorporate. Another is the *way* in which the structure rules are embodied or represented—a matter that is often determined by reference to the distinctions between the structure rules themselves. For these reasons, it’s important to discuss the evolution of the structure rules that have played a prominent role in linguistics and psycholinguistics. In this section, I will begin by introducing context-free grammars. As the chapter continues, I will rehearse the distinctions between this formalism and a variety of others.

Context-free phrase structure rules, sometimes called “production” or “rewrite” rules,

constitute the starting point of many contemporary syntactic theories. Fig. 1 contains an impoverished context-free grammar (CFG) for a fragment of English. Like all CFGs, this grammar is divided into two types of rule: lexical and grammatical.

The lexical rules are used to make claims about the syntactic categories to which the words of a language belong. They do so by mapping *nonterminal* symbols like ‘*Noun*’ and ‘*Verb*’, which denote those categories (sometimes called “parts of speech”) into *terminal* symbols like ‘morning’ and ‘prefer’—i.e., actual words or morphemes. The lexical rules, taken together, can be regarded as the lexicon of the language that the grammar is used to describe.

| Lexical rules                                                                                                                 |                                               |
|-------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------|
| <i>Noun</i> → <i>flight</i>   <i>breeze</i>   <i>trip</i>   <i>morning</i>                                                    |                                               |
| <i>Verb</i> → <i>is</i>   <i>prefer</i>   <i>like</i>   <i>need</i>   <i>want</i>   <i>fly</i>   <i>leave</i>                 |                                               |
| <i>Adjective</i> → <i>cheapest</i>   <i>non-stop</i>   <i>first</i>   <i>latest</i>   <i>other</i>   <i>direct</i>            |                                               |
| <i>Pronoun</i> → <i>me</i>   <i>I</i>   <i>you</i>   <i>it</i>                                                                |                                               |
| <i>Proper-Noun</i> → <i>Alaska</i>   <i>Baltimore</i>   <i>Los Angeles</i>   <i>Chicago</i>   <i>United</i>   <i>American</i> |                                               |
| <i>Determiner</i> → <i>the</i>   <i>a</i>   <i>an</i>   <i>this</i>   <i>these</i>   <i>that</i>                              |                                               |
| <i>Preposition</i> → <i>from</i>   <i>to</i>   <i>on</i>   <i>near</i>                                                        |                                               |
| <i>Conjunction</i> → <i>and</i>   <i>or</i>   <i>but</i>                                                                      |                                               |
| <i>Complementizer</i> → <i>that</i>                                                                                           |                                               |
| Grammatical rules                                                                                                             | Quasi-English Examples                        |
| <i>S</i> → <i>NP VP</i>                                                                                                       | I + want a morning flight                     |
| <i>S</i> → <i>VP</i>                                                                                                          | Leave me in Chicago.                          |
| <i>NP</i> → <i>Pronoun</i>                                                                                                    | I                                             |
| <i>NP</i> → <i>Proper-Noun</i>                                                                                                | Los Angeles                                   |
| <i>NP</i> → <i>Det Nominal</i>                                                                                                | a + flight                                    |
| <i>Nominal</i> → <i>Nominal Noun</i>                                                                                          | morning + flight                              |
| <i>Nominal</i> → <i>Noun</i>                                                                                                  | flight                                        |
| <i>VP</i> → <i>Verb</i>                                                                                                       | do                                            |
| <i>VP</i> → <i>Verb NP</i>                                                                                                    | want + a flight                               |
| <i>VP</i> → <i>Verb NP PP</i>                                                                                                 | leave + Boston + in the morning               |
| <i>VP</i> → <i>Verb PP</i>                                                                                                    | leave + on Thursday                           |
| <i>VP</i> → <i>Verb Complementizer S</i>                                                                                      | prefer + that + the flight leave this morning |

**Fig. 1: A toy context-free grammar. In the lexical rules, the symbol ‘|’ expresses exclusive disjunction. Adapted from Jurafsky and Martin (2007: p. 394)**

The lexicon presented in fig. 1 is rather poor, describing only the basic syntactic category



of any given word. No description of the words' phonological, morphological, or semantic properties is provided. Similarly, this lexicon does not specify the more fine-grained syntactic properties of the words—e.g., the verbs' subcategorization frames or the nouns' agreement features. Nor does it provide any statistics about the frequency with which the words are used, either alone or in combination. Below, we will explore the prospect of adding these features to the grammar—particularly frequency information, which yields a *Probabilistic Context-Free Grammar* (PCFG).

Unlike the lexical rules, the grammatical rules of a CFG are used to make lawlike claims about the relationships *between* syntactic categories. For instance, the rule ' $VP \rightarrow Verb\ NP$ ' amounts to the modal claim that if something *were* to consist of a standalone verb followed by a noun phrase, then it *would* thereby be a verb phrase.<sup>16</sup>

The grammar is recursive, hence capable of describing an infinite number of sentences (if there happen to be that many; see ch. I). The recursive nature of the grammar is due to the following two rules:

$$\begin{aligned} S &\rightarrow NP\ VP \\ VP &\rightarrow Verb\ Complementizer\ S \end{aligned}$$

Jointly, these two rules entail that the language contains structures in which a sentence is embedded within another sentence, as in the following examples:

- (1) I prefer that the flight leave this morning.
  - (2) I prefer that I prefer that the flight leave this morning.
  - (3) I prefer that I prefer that I prefer that the flight leave this morning.
- etc.

The ' $\rightarrow$ ' symbol must be treated with care. A common practice is to read a rule like

---

<sup>16</sup> We can trade in the awkward subjunctive locutions for ordinary material conditionals, so long as we keep firmly in mind that the latter are *lawlike*, in the sense of Goodman (1954/1983). See ch. I for more in-depth discussion of the bearing this point has on the ontology of linguistic theory.

‘ $S \rightarrow NP VP$ ’ as “S *goes to* NP VP.” But the import of ‘goes to’ is rather opaque. And the same is true of the popular “rewrite” and “produce” locutions. The claims that “S *can be rewritten as* NP VP” or that “S *produces* NP VP” are equally obscure. (What modality attaches to ‘can’? Who is doing the rewriting or the “producing”? What does it even mean to “produce” here?) For some purposes, of course, this nonspecificity is desirable; and for many purposes it’s either harmless or irrelevant. But for *our* purposes such formulations are dangerously sloppy, for they slur over the distinction between structure rules and processing rules.<sup>17</sup> Let us, then, be pedantically explicit: If we wish to treat the above statements as structure rules of English, then we should read ‘ $S \rightarrow NP VP$ ’ as claiming that something *would be* a sentence *if it were* to consist of a concatenation between a noun phrase and a verb phrase (see fn. 16). The same goes, *mutatis mutandis*, for all of the grammatical rules listed above.

Context-free rewrite rules impose a binary distinction between the strings that are and the strings that are not sentences of some language.<sup>18</sup> The question arises: Which language? If we take a language to be a set of sentences, there are several options. Consider first the language that consists of precisely those strings that are sentences according to the claims of the grammar. Trivially, the grammar is true of *this* language. But now contrast this with your own idiolect—i.e., the variant of English that you currently speak (assuming, for the moment, that there really is just one; see ch. II). The grammar listed above will not be true of *it*, so long as you would, in some relatively ordinary circumstance, express yourself by uttering the string in (4) and so long as you are strongly disposed to *not* express yourself by uttering the string in (5).

(4) Linguistics is difficult.

---

<sup>17</sup> See Karttunen and Zwicky (1985: pp. 3-5) for a discussion of closely related issues.

<sup>18</sup> Strictly speaking, in order to secure this result, we would have to include an explicit rule to the effect that *nothing else* is a sentence of the language. Following convention, I have omitted this in the main text.

(5) Fly morning on breeze.

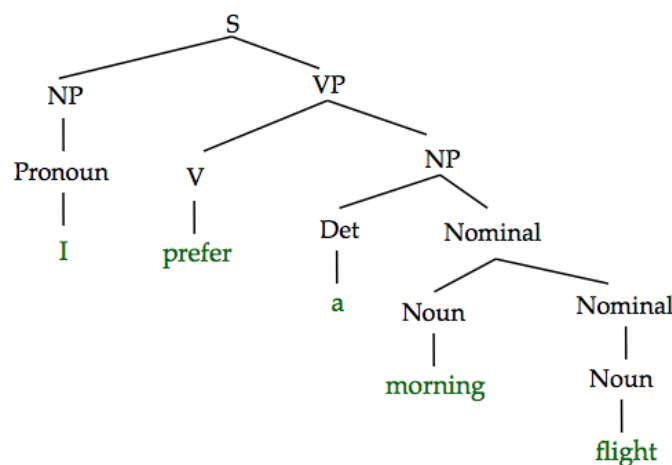
The reason that the grammar presented above is not true of *your* idiolect is that it both undergenerates and overgenerates, in the following sense: (4) uses lexical items that do not appear in that grammar, while the aberrant (5) *can* be generated by the rules of that grammar. Similar remarks can be made for a public *dialect* of English (ch. II).

A context-free grammar can be used to do more than just demarcate sentences from nonsentences. It can also be used to describe the *hierarchical structure* of the sentences in a language. For instance, the grammar presented above describes (6) as having a particular structure, which can be represented with brackets (7), trees (8), or in ordinary English (9).

(6) I prefer a morning flight.

(7) [S [NP [Pronoun I]] [VP [Verb prefer] [NP [Determiner a] [Nominal [N morning] [Nominal [N flight]]]]]]]

(8)



(9) The sentence ‘I prefer a morning flight’ consists of the concatenation of a noun phrase and a verb phrase. The noun phrase consists of the pronoun ‘I’. The verb phrase consists of a concatenation of the verb ‘prefer’ and another noun phrase. This (second) noun phrase itself consists of a concatenation of the determiner ‘the’ and a nominal. The nominal consists of a concatenation of the noun ‘morning’ and another nominal. This remaining nominal consists of the noun ‘flight’.

As (7)-(9) illustrate, it is trivial to express in ordinary English any claims that one can make in

either of the other notations. But the bracket and tree notations can be very convenient for a number of purposes—brackets render any such claim shorter; trees render it easier to read. The important point to keep in mind, though, is that these are merely notational variants. Whatever the notation, (7)–(9) all serve to make claims about the structure of a sentence. Without further argument, nothing at all follows about the psychology of a language user, nor about the internal operations of a computational system designed to parse sentences of a language.

In the next section, I discuss the role that CFGs play in a number of parsing algorithms, including the (mostly) top-down Earley algorithm and the bottom-up CYK algorithm, both of which draw on explicitly represented context-free rules, which are stored in a separate data structure. We will also consider left-corner parsers, and models that employ probabilistic CFGs. Finally, we will see how the Parsing as Deduction framework uses first-order logic and other familiar formalisms to render CFGs suitable for use in deductive proof procedures.

## §2. Foundational Parsing Algorithms

### §2.1 *Basic Terms*

The discussion in this section draws heavily on Jurafsky and Martin (2007), a canonical text in computational linguistics. The authors of this text open their chapter on parsing algorithms by drawing a distinction between structure rules and processing rules.

We defined parsing ... as a combination of recognizing an input string and assigning a structure to it. Syntactic parsing, then, is the task of recognizing a sentence and assigning a syntactic structure to it.<sup>[19]</sup> This chapter focuses on the

---

<sup>19</sup> There is a distinction between “recognizing” a sentence and assigning a syntactic structure to it. A system’s recognizing a string, or “accepting” it, amounts to no more than that system’s issuing a judgment to the effect that the string in question is grammatical, relative to the grammar with which the system is operating. Parsing, by contrast, involves constructing one or more representations of the string’s syntactic structure and, in the case of ambiguity, selecting one of these as the *privileged* representation—the one that constitutes the system’s “ultimate decision” about how the input should be interpreted. For a discussion of this distinction, see Berwick and Weinberg (1984), p. 252. In fn. 12, Berwick and Weinberg claim (but do not argue) that parsing, over and above recognition,

kind of structures assigned by context-free grammars ... [S]ince they are based on a purely declarative formalism, context-free grammars don't specify *how* the parse tree for a given sentence should be computed. We'll therefore need to specify algorithms that employ these grammars to produce trees. (p. 431)

Jurafsky and Martin view the construction of syntactic representations as a necessary step toward the final goal of semantic interpretation.

[P]arse trees serve as an important intermediate stage of representation for semantic analysis ... and thus play an important role in applications like question answering and information extraction. For example, to answer the question *What books were written by British women authors before 1800?* we'll need to know that the subject of the sentence was *what books* and that the *by* adjunct was *British women authors* to help us figure out that the user wants a list of books (and not a list of authors). (p. 431)

The authors go on to discuss two of the most basic methods available for constructing parse trees. These are known as *top-down* and *bottom-up* parsing.<sup>20</sup> The distinction between these two approaches reflects a more basic distinction, with which philosophers are surely well acquainted.

Regardless of the search algorithm we choose, there are two kinds of constraints that should help guide the search. One set of constraints comes from the data, that is, the input sentence itself. ... The second kind of constraint comes from the grammar. ... These two constraints ... give rise to the two search strategies underlying most parsers: top-down or goal-directed search, and bottom-up or data-directed search. These constraints are more than just search strategies. They reflect two important insights in the western philosophical tradition: the rationalist tradition, which emphasizes the use of prior knowledge, and the empiricist tradition, which emphasizes the data in front of us. (p. 433)

In what follows, I explore the algorithms that have been devised on the basis of these two strategies. I begin by presenting the top-down and bottom-up parsing algorithms in their most austere forms, and then discuss refinements that address issues of computational efficiency.

---

is necessary for effecting a mapping from phonological representations to semantic interpretations. I argued for this in chs. III and IV.

<sup>20</sup> Fodor, Bever, and Garrett (1974) refer to top-down and bottom-up techniques as *analysis-by-analysis* and *analysis-by-synthesis*, respectively. Computer scientists sometimes use the terms *recursive-descent* and *shift-reduce*, for reasons that will become apparent below. Note that the terms 'top-down' and 'bottom-up', as used here, do *not* describe the relations between the surface-structures and deep-structures that are posited by some grammars.

## §2.2 Top-down Parsing

In a top-down parser, the process of assigning syntactic structure to a linguistic input proceeds in phases, sometimes called *plies*. At each ply, the parser uses the internally represented grammar to issue predictions, in something like the way that a scientist uses her background theory to issue predictions prior to performing an experiment. The parser begins by positing an initial tree-node—traditionally an S.<sup>21</sup> This node, alone, constitutes the first ply. In the second ply, the S node is expanded in accordance with the internally represented grammar. For purposes of illustration, let us assume that the parser is working with a pitifully simple CFG, which we label ‘G’ in fig. 2 below. Fig. 3 depicts the parser’s outputs at two successive plies.

| Grammatical rules                  | Lexical rules                                                         |
|------------------------------------|-----------------------------------------------------------------------|
| $S \rightarrow NP VP$              | $Det \rightarrow that \mid this \mid a$                               |
| $S \rightarrow Aux NP VP$          | $Noun \rightarrow book \mid flight \mid meal \mid money$              |
| $S \rightarrow VP$                 | $Verb \rightarrow book \mid include \mid prefer$                      |
| $NP \rightarrow Pronoun$           | $Pronoun \rightarrow I \mid she \mid me$                              |
| $NP \rightarrow Proper-Noun$       | $Proper-Noun \rightarrow Houston \mid NWA$                            |
| $NP \rightarrow Det$               | $Aux \rightarrow does$                                                |
| $Nominal \rightarrow Noun$         | $Preposition \rightarrow from \mid to \mid on \mid near \mid through$ |
| $Nominal \rightarrow Nominal Noun$ |                                                                       |
| $Nominal \rightarrow Nominal PP$   |                                                                       |
| $VP \rightarrow Verb$              |                                                                       |
| $VP \rightarrow Verb NP$           |                                                                       |
| $VP \rightarrow Verb NP PP$        |                                                                       |
| $VP \rightarrow Verb PP$           |                                                                       |
| $VP \rightarrow VP PP$             |                                                                       |
| $PP \rightarrow Preposition NP$    |                                                                       |

**Fig. 2: The context-free grammar ‘G’.** Adapted from Jurafsky and Martin (2007: p. 394)

---

<sup>21</sup> In contemporary syntactic theories, the phrasal category S has been replaced by other phrasal types—different ones in different theories. The list includes, *inter alia*, inflectional phrase (IP), complementizer phrase (CP), and tense phrase (TP). For ease of exposition, I ignore this and related complications.

- (10) Book that flight.  
 (11) [S [VP [V Book]]] [NP [Det that]][Nom [N flight]]].

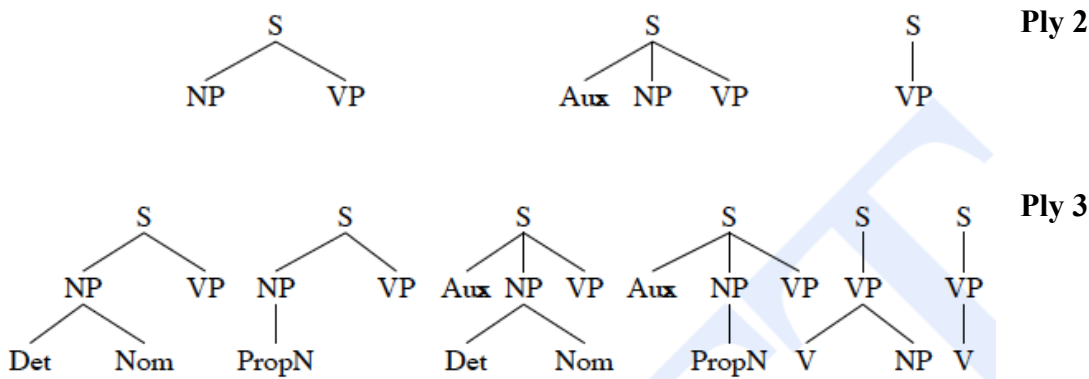


**Figure 3:** Plies 2 and 3 in a simple top-down parse of the input *Book that flight*, assuming grammar G (fig. 2).

In the second ply, the S node is expanded three separate times into (i) NP + VP, (ii) Aux + NP + VP, and (iii) a standalone VP. As figs. 2 and 3 illustrate, these are the three expansion options that G allows. They constitute the parser's disjunctive prediction to the effect that the input will consist of either (i) noun phrase and verb phrase, (ii) an auxiliary phrase, a noun phrase and a verb phrase, or (iii) just a verb phrase. In the third ply, these nodes are themselves expanded in accordance with G.

In theory, this process of expansion would proceed until all of the nodes are parts of speech, which cannot be expanded in accordance with G. When the parser reaches a ply in which all of the leaves (i.e., terminal nodes) of all of the parse trees are parts of speech, it finally looks at the input string to see whether it matches any of the items in the lexicon.<sup>22</sup> This, in turn,

<sup>22</sup> In principle, a top-down parser could make predictions even about which *specific lexical items* will appear in the input stream. But such predictions, even if made on the basis of frequency information and pragmatic/contextual clues, would still be rather risky. I assume, then, that lexical retrieval is in significant measure a data-driven process. This raises a question about whether the “matching” involved in lexical recognition is brute-causal, in the sense introduced by Devitt (2006). I suspect, though cannot adequately establish here, that it is not. One consideration that bears on the issue is whether the HSPM constructs *phonological* representations prior to, and in the service of, lexical retrieval. In ch. IV, I argued that syntactic processing is not a brute-causal process. Moreover, I claimed there that my arguments for the psychological reality of mental phrase markers (MPMs) carry

would allow the parser to see which of its predictions turned out to be correct. For instance if the input is (10), whose structure is given in (11), then the parser could immediately rule out any parse tree whose first leaf is *Determiner*. It could then systematically rule out all of the other parse trees, except the correct one. The parsing process would then be complete.

We have thus far been ignoring a major complication. Notice that, if grammar *G* were recursive, the parser would never reach a ply in which all of the leaves of the parse trees are terminal nodes—i.e., parts of speech. Consider the combined effect of the following two rules.

- (12) a.  $S \rightarrow NP VP$   
b.  $VP \rightarrow Verb Complementizer S$

These rules will conspire to continuously create more and more candidate parse trees by expanding the *S* node into parse trees that contain a *VP* node, and then expanding that *VP* node into parse trees that contain an *S* node. The process would repeat without terminating.

This problem can be overcome by making use of the classic generate-and-test strategy. The internally represented phrase-structure rules can generate some candidate parse trees (say, two plies' worth) and then stop, allowing the system to access some of the input data prior to generating more parse trees. The strategy is no longer *purely* top-down, as it is allowed to “look at” the input prior to generating more predictions. But this adulteration is clearly motivated; the system must avoid issuing infinitely many predictions.

Though it is no longer purely top-down, the parser is still quite simplistic and inefficient.

---

over to the logically independent case of phonological processing. For, as with syntactic processing, the activation of a phonological representation is highly context-sensitive and dependent on factors that are not present in the immediate stimulus, but scattered over discontinuous chunks of time (Fernandez and Cairns, 2011: ch. 6). Another consideration has to do with the fact that prior higher-level syntactic decisions exert a downward influence on lexical retrieval and actively guide the correction of errors in the retrieval process (ch. IV, §3.3). Still, the matter plainly deserves more discussion than I can devote to it here. For present purposes, I note only that my conclusions regarding the rational-causal nature of syntactic processing—i.e., the construction of structured MPMs for whole phrases and clauses—would not be wholly undermined (though it may be threatened) if it turned out that lexical retrieval is a brute-causal process.



The top-down process still generates an unacceptable number of predictions that ultimately fail to accord with the input. Indeed, many of its predictions might be consistent with the input *up until a point*, but then disconfirmed thereafter. And given that it takes time and computational resources for a physical system to generate such predictions, each prediction that is ultimately inconsistent with the input is a waste of limited time and resources. With a realistic broad-coverage grammar, the resulting inefficiency is devastating.

Another possibility is to *pre-compute* the search space allowed by the grammar and store it in memory so that the parser does not have to generate all possible parse trees each and every time before checking the input. As Marcus (1980: p. 29) points out, the parsers known as STRING (by Naomi Sanger), LUNAR (by Bill Woods), and SHRDLU (by Terry Winograd) are all instances of the top-down approach. Focusing his attention on the LUNAR parser, Marcus notes that it is based on an ATN grammar, which is essentially a pre-defined search space that models the syntactic options of an interesting fragment of English.<sup>23</sup> The job of the parser, then, is to navigate the search space. Parsing thus becomes a search task. Marcus goes on to argue that this, too, is highly inefficient, because at each choice point in the search, the parser will have to make a decision that is potentially incorrect—an error that might be discovered arbitrarily late in the sentence. Consider, for instance, the following sentences.

- (13) Is the woman in the office?
- (14) Is the woman in the office on the third floor?
- (15) Is the woman in the office on the third floor of the new building?
  
- (16) Is the woman in the office helpful?
- (17) Is the woman in the office on the third floor helpful?
- (18) Is the woman in the office on the third floor of the new building helpful?

---

<sup>23</sup> The infinitude of the space is handled by a simple formal device that we shall introduce in our discussion of ATN parsers in §4. For a visual aid, the reader can skip down to figs. 14-16 in §4, which illustrate the type of search space I have in mind here.

If the parser assumes that “the woman in the office...” is a reduced relative clause, then when it encounters the question mark in (13)-(15), it will have to engage in a process of “backtracking”—i.e., retracting its commitments earlier in the sentence, one by one, and trying all of the available options at each prior choice point, until it reaches the site of the initial error. If the parser instead assumes that the “the woman in the office...” is a noun phrase, it will encounter the same problem when it reaches the word ‘helpful’ in (16)-(18). Either assumption, when found to be incorrect, leads to an unacceptable amount of backtracking and revision of commitments. (See Marcus, 1980: chs. 1-2 for more in-depth discussion.)

### §2.3 *The Earley Algorithm and Probabilistic Context-Free Grammars*

A solution to some these problems was developed by Jay Earley, who uses a popular technique known as “dynamic programming.”<sup>24</sup> The Earley algorithm, named after its inventor, avoids generating the same partially successful predictions again and again at each successive ply, or backtracking through prior choice points one by one and reduplicating work already done. This is accomplished by storing some number of successful *partial* parse trees in a data-structure known as a “chart” or a “well-formed substring table.” The successful partial parses are then called on when needed, instead of having to be reconstructed anew.

The Earley algorithm is widely used in top-down parsing systems and has been applied to probabilistic extensions of context-free grammars (Hale, 2001; 2003), as well as to Minimalist and other mildly context-sensitive grammars (Harkema, 2001) (§6). Let us take a closer look at its operations. The algorithm (fig. 5) has three main functions: *Predict*, *Scan*, and *Complete*.

---

<sup>24</sup> Earley, J. (1970). “An efficient context-free parsing algorithm.” *Communications of the Association for Computing Machinery*, 13(2). The Earley algorithm is discussed, in various levels of detail, in Hale (2001), Jurafsky and Martin (2007), Kilbury (1985), Harkema (2001), Aho and Ullman (1972), Morawietz (2000), Pereira and Warren (1983), and Shieber, Schabes, and Pereira (1995).

The *Predict* function adds to the chart all expansions of a node that are consistent with the grammar, which is internally represented in a declarative data structure. The *Scan* function looks at the input, prior to issuing further predictions. The *Complete* function stores all of the confirmed predictions, and propagates them to the next stage of the process, while removing from the chart all of the disconfirmed predictions. The rules of the internally represented grammar are annotated, by convention, with a “dot”—i.e., the symbol ‘•’. The dot indicates how much of a rule has already been completed. Fig. 4 illustrates the operations of the Earley parser on the input sentence ‘Book that flight’.

|          |     |                                            |       |                   |
|----------|-----|--------------------------------------------|-------|-------------------|
| Chart[0] | S0  | $\gamma \rightarrow \bullet S$             | [0,0] | Dummy start state |
|          | S1  | $S \rightarrow \bullet NP VP$              | [0,0] | Predictor         |
|          | S2  | $S \rightarrow \bullet Aux NP VP$          | [0,0] | Predictor         |
|          | S3  | $S \rightarrow \bullet VP$                 | [0,0] | Predictor         |
|          | S4  | $NP \rightarrow \bullet Pronoun$           | [0,0] | Predictor         |
|          | S5  | $NP \rightarrow \bullet Proper-Noun$       | [0,0] | Predictor         |
|          | S6  | $NP \rightarrow \bullet Det Nominal$       | [0,0] | Predictor         |
|          | S7  | $VP \rightarrow \bullet Verb$              | [0,0] | Predictor         |
|          | S8  | $VP \rightarrow \bullet Verb NP$           | [0,0] | Predictor         |
|          | S9  | $VP \rightarrow \bullet Verb NP PP$        | [0,0] | Predictor         |
|          | S10 | $VP \rightarrow \bullet Verb PP$           | [0,0] | Predictor         |
|          | S11 | $VP \rightarrow \bullet VP PP$             | [0,0] | Predictor         |
| Chart[1] | S12 | $Verb \rightarrow book \bullet$            | [0,1] | Scanner           |
|          | S13 | $VP \rightarrow Verb \bullet$              | [0,1] | Completer         |
|          | S14 | $VP \rightarrow Verb \bullet NP$           | [0,1] | Completer         |
|          | S15 | $VP \rightarrow Verb \bullet NP PP$        | [0,1] | Completer         |
|          | S16 | $VP \rightarrow Verb \bullet PP$           | [0,1] | Completer         |
|          | S17 | $S \rightarrow VP \bullet$                 | [0,1] | Completer         |
|          | S18 | $VP \rightarrow VP \bullet PP$             | [0,1] | Completer         |
|          | S19 | $NP \rightarrow \bullet Pronoun$           | [1,1] | Predictor         |
|          | S20 | $NP \rightarrow \bullet Proper-Noun$       | [1,1] | Predictor         |
|          | S21 | $NP \rightarrow \bullet Det Nominal$       | [1,1] | Predictor         |
|          | S22 | $PP \rightarrow \bullet Prep NP$           | [1,1] | Predictor         |
| Chart[2] | S23 | $Det \rightarrow that \bullet$             | [1,2] | Scanner           |
|          | S24 | $NP \rightarrow Det \bullet Nominal$       | [1,2] | Completer         |
|          | S25 | $Nominal \rightarrow \bullet Noun$         | [2,2] | Predictor         |
|          | S26 | $Nominal \rightarrow \bullet Nominal Noun$ | [2,2] | Predictor         |
|          | S27 | $Nominal \rightarrow \bullet Nominal PP$   | [2,2] | Predictor         |
| Chart[3] | S28 | $Noun \rightarrow flight \bullet$          | [2,3] | Scanner           |
|          | S29 | $Nominal \rightarrow Noun \bullet$         | [2,3] | Completer         |
|          | S30 | $NP \rightarrow Det Nominal \bullet$       | [1,3] | Completer         |
|          | S31 | $Nominal \rightarrow Nominal \bullet Noun$ | [2,3] | Completer         |
|          | S32 | $Nominal \rightarrow Nominal \bullet PP$   | [2,3] | Completer         |
|          | S33 | $VP \rightarrow Verb NP \bullet$           | [0,3] | Completer         |
|          | S34 | $VP \rightarrow Verb NP \bullet PP$        | [0,3] | Completer         |
|          | S35 | $PP \rightarrow \bullet Prep NP$           | [3,3] | Predictor         |
|          | S36 | $S \rightarrow VP \bullet$                 | [0,3] | Completer         |
|          | S37 | $VP \rightarrow VP \bullet PP$             | [0,3] | Completer         |

**Figure 5:** The first column contains labels of the successive charts that are created in the course of parsing the sentence ‘<sub>0</sub> Book <sub>1</sub> that <sub>2</sub> flight <sub>3</sub>’ (where the numbers denote positions in the sentence). The second column contains labels of the steps that constitute the process. The third column displays the “dotted” rules of the internally represented grammar. Before the process is complete, the right-hand side of each rule has some nodes to the right of the dot (‘•’), indicating that these nodes have not yet been expanded. The nodes to the left of the dot have been successfully expanded. For instance, state S14 indicates that the *Verb* node has been expanded and completed, but the *NP* node has not. The fourth column indicates how much of the material from the input has been successfully integrated. For instance, ‘[1,3]’ in line S30 indicates that

this state has managed to integrate the words ‘that flight’ in ‘<sub>0</sub> Book <sub>1</sub> that <sub>2</sub> flight <sub>3</sub>’. The fifth column displays the function responsible for entering each state into the chart. A parse is complete when all of the input material has been integrated and the dot is on the rightmost edge of a rule that begins with the symbol ‘S’—in this case, state S36. *Source*: Jurafsky and Martin (2007), p. 451.

A cursory glance at the pseudocode provided by Jurafsky and Martin (2007) reveals that each of the three main functions of the Earley algorithm draw on an internally represented grammar (fig. 4). This is particularly obvious in the case of the *Predictor* function, which makes explicit reference to a data structure that the authors have aptly labeled GRAMMAR-RULES.

|                                                                                                                                                                                                                                                                                                             |       |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
| <pre> <b>procedure</b> PREDICTOR(<math>(A \rightarrow a \bullet B b, [i, j])</math>)   <b>for each</b> <math>(B \rightarrow g)</math> <b>in</b> GRAMMAR-RULES-FOR(<math>B, grammar</math>) <b>do</b>     ENQUEUE(<math>(B \rightarrow \bullet g, [j, j])</math>, <math>chart[j]</math>)   <b>end</b> </pre> | ← !!! |
| <pre> <b>procedure</b> SCANNER(<math>(A \rightarrow a \bullet B b, [i, j])</math>)   <b>if</b> <math>B \in \text{PARTS-OF-SPEECH}(\text{word}[j])</math> <b>then</b>     ENQUEUE(<math>(B \rightarrow \text{word}[j], [j, j+1])</math>, <math>chart[j+1]</math>)   <b>end</b> </pre>                        | ← !!! |
| <pre> <b>procedure</b> COMPLETER(<math>(B \rightarrow g \bullet, [j, k])</math>)   <b>for each</b> <math>(A \rightarrow a \bullet B b, [i, j])</math> <b>in</b> <math>chart[j]</math> <b>do</b>     ENQUEUE(<math>(A \rightarrow a B \bullet b, [i, k])</math>, <math>chart[k]</math>)   <b>end</b> </pre>  | ← !!! |

**Fig. 4: The pseudocode for the Earley algorithm. I have marked explicit references to grammatical rules with the symbol ‘!!!’.** *Source*: Jurafsky and Martin (2007), p. 448.

It is a significant feature of the parser described here that the rules of the grammar are stored in a separate data structure. This allows the programmer to swap out one grammar and replace it with another. For better or worse, other parsers do not share this feature. For instance, the famous Marcus parser (Marcus, 1980), as well as the ATN models that we discuss in §4, use the rules *procedurally*. That is, the rules are “implicit” in the parser’s operations, not explicitly represented and accessed as data. We will return to this contrast in §4, and again in §7.

A number of refinements can be introduced to make the Earley algorithm more efficient. For instance, instead of blindly making predictions on the basis of all of the rules in its internally

represented grammar, the parser can be designed to choose intelligently which rules to apply at each point in the parse. Moreover, the order in which the *Predict*, *Scan*, and *Complete* functions are called can be controlled as well. This is accomplished by allowing the process to be guided by what's known as an *Agenda*.<sup>25</sup> This seems to be a natural site for the implementation of the least-effort parsing principles MA, LC, and MCP (§1). For instance, consider the principle of Late Closure, which dictates that the HSPM will incorporate newly encountered material into the most recent phrase or clause of the parse tree that it has already constructed. The *Agenda* can implement this principle as follows: After the *Scan* function determines a word's part-of-speech category, the *Agenda* can dictate that the predictor insert into the chart only those rules of the grammar that expand nodes in such a way that this part-of-speech category appears to left of the dot. More complex strategies on the part of the *Agenda* might serve to implement the principle of Minimal Attachment.<sup>26</sup>

Another method that has been used to combat the problem of inefficiency makes use of probabilistic information. As noted above, the Earley algorithm has been applied to Probabilistic Context-Free Grammars (PCFGs), which include information about the frequency with which each rule of the grammar is applied, as well as the frequencies of particular lexical items. Fig. 6 illustrates this with a toy grammar. More sophisticated grammars have been culled from

---

<sup>25</sup> For details, see Jurafsky and Martin (2007: pp. 452-454). Kaplan (1973) contains an early but prescient discussion of various chart-parsing techniques.

<sup>26</sup> Though I cannot say with certainty whether such strategies would be useful for a parallel parser, it appears (*prima facie*) to be a promising way of cutting down on the burdensome memory requirements associated with parallel parsing. There may, however, be a difficulty in the specific case of Minimal Attachment. Fodor and Frazier (1980) develop a number of arguments against the psychological plausibility of ATN parsers (to be examined in §4). In the course of doing so, they point out (fn. 8) that, while Late Closure can be implemented in the agenda of an ATN parser (albeit in an unprincipled fashion), Minimal Attachment cannot. It is not clear that their argument transfers over to parsers other than the ATN.

annotated corpuses.<sup>27</sup> Extensions of the Earley algorithm in this direction provide another powerful way of cutting down on inefficiency. In a serial parser, statistical information contained in PCFGs would be used to direct the decisions as to which rule to apply or which lexical item to select in cases of ambiguity.<sup>28</sup> In a parallel parser, like the one that runs the Earley algorithm, statistical information is used to *rank* the parse trees that are constructed at every successive ply and to determine which of these is the most likely to eventuate in an accurate parse.

In light of the data now available, it is well-nigh indisputable that the HSPM is sensitive to frequency information at various levels of analysis.<sup>29</sup> For this reason, parsing models that make use of probabilistic grammars not only have practical applications in commercial settings, but also have a strong claim to psychological reality. To take just one example, Hale (2001, 2003) presents evidence for the psychological plausibility of an Earley parser that draws on an

---

<sup>27</sup> For realistic PCFGs, see Hale (2001, 2003).

<sup>28</sup> Jurafsky and Martin (2007: ch. 14) contains an in-depth discussion of the details. See also Hale (2001, 2003).

<sup>29</sup> See Jurafsky (2003) for a comprehensive review of the relevant psycholinguistic studies. See also fn. 9.

internally represented PCFG.

| Grammar                            |       | Lexicon                           |                                  |
|------------------------------------|-------|-----------------------------------|----------------------------------|
| $S \rightarrow NP VP$              | [.80] | $Det \rightarrow that$            | [.10]   $a$ [.30]   $the$ [.60]  |
| $S \rightarrow Aux NP VP$          | [.15] | $Noun \rightarrow book$           | [.10]   $flight$ [.30]           |
| $S \rightarrow VP$                 | [.05] |                                   | $meal$ [.15]   $money$ [.05]     |
| $NP \rightarrow Pronoun$           | [.35] |                                   | $flights$ [.40]   $dinner$ [.10] |
| $NP \rightarrow Proper-Noun$       | [.30] | $Verb \rightarrow book$           | [.30]   $include$ [.30]          |
| $NP \rightarrow Det Nominal$       | [.20] |                                   | $prefer$ [.40]                   |
| $NP \rightarrow Nominal$           | [.15] | $Pronoun \rightarrow I$           | [.40]   $she$ [.05]              |
| $Nominal \rightarrow Noun$         | [.75] |                                   | $me$ [.15]   $you$ [.40]         |
| $Nominal \rightarrow Nominal Noun$ | [.20] | $Proper-Noun \rightarrow Houston$ | [.60]                            |
| $Nominal \rightarrow Nominal PP$   | [.05] |                                   | $NWA$ [.40]                      |
| $VP \rightarrow Verb$              | [.35] | $Aux \rightarrow does$            | [.60]   $can$ [.40]              |
| $VP \rightarrow Verb NP$           | [.20] | $Preposition \rightarrow from$    | [.30]   $to$ [.30]               |
| $VP \rightarrow Verb NP PP$        | [.10] |                                   | $on$ [.20]   $near$ [.15]        |
| $VP \rightarrow Verb PP$           | [.15] |                                   | $through$ [.05]                  |
| $VP \rightarrow Verb NP NP$        | [.05] |                                   |                                  |
| $VP \rightarrow VP PP$             | [.15] |                                   |                                  |
| $PP \rightarrow Preposition NP$    | [1.0] |                                   |                                  |

**Fig. 6:** A toy probabilistic context-free grammar (PCFG). The numbers in brackets denote the probability of applying a production rule—i.e., the frequency with which that rule has been applied in an annotated corpus, like the Penn Treebank. All probabilities for rules with the same right-hand side should sum to 1. For instance, the three rules for expanding the S node are .85, .10, and .05. *Source:* Jurafsky and Martin (2007), pp. 465

Hale begins by noting his commitment to the Strong Competence Hypothesis, which we mentioned in §1. (More on this below.)

What is the relation between a person's knowledge of grammar and that same person's application of that knowledge in perceiving syntactic structure? The answer to be proposed here observes three principles. **Principle 1** *The relation between the parser and grammar is one of strong competence. Strong competence holds that the human sentence processing mechanism directly uses rules of grammar in its operation, and that a bare minimum of extragrammatical machinery is necessary.* This hypothesis, originally proposed by Chomsky (Chomsky, 1965, page 9) has been pursued by many researchers (Bresnan, 1982) (Stabler, 1991) (Steedman, 1992) (Shieber and Johnson, 1993), and stands in contrast with an approach directed towards the discovery of autonomous principles unique to the processing mechanism.

He then goes on to show how a model that makes use of the Earley algorithm and a PCFG can predict two kinds of processing difficulties that the HSPM is known to exhibit. The first has to



do with the main-clause/relative-clause ambiguity discussed in ch. III. The examples from that discussion are repeated here:

- (19) The horse raced past the barn fell.
- (20) The ship floated down the river sank.
- (21) The dealer sold the forgeries complained.
- (22) The man sent the letter cried.

The second processing difficulty has to do with the asymmetry between subjects and objects in unreduced relative clauses. The examples below are taken from Gibson (1998).<sup>30</sup>

- (23) [*S* The reporter [*S'* who [*S* the senator attacked ] ] admitted the error ].
- (24) [*S* The reporter [*S'* who [*S* attacked the senator ] ] admitted the error ].

Notice the difference between the account of garden-path effects that stems from the least-effort parsing principles, MA, LC, and MCP (§1) and the account that Hale (2001, 2003) provides. The former is committed to a *serial* processing model, on which the HSPM incorporates new input incrementally by constructing only one parse tree—i.e., one mental phrase marker—at each successive step. The processing difficulties associated with garden-path sentences are then accounted for by the following principle: When new input doesn't fit into the parse tree that the HSPM has constructed, an operation of *reanalysis*—sometimes computationally costly—must be undertaken.

The model in Hale (2001, 2003) provides a different account. Hale assumes that the HSPM is a *parallel* parser, which constructs *multiple* mental phrase markers at each successive step—one for each possibility consistent with the grammar (above a given probabilistic

---

<sup>30</sup> Gibson cites the results of a number of psycholinguistic experiments that establish the reality of this processing difficulty: “The object extraction is more complex by a number of measures including phoneme monitoring, on-line lexical decision, reading times, and response-accuracy to probe questions (Holmes, 1973; Hakes et al., 1976; Wanner and Maratsos, 1978; Holmes and O'Regan, 1981; Ford, 1983; Waters et al., 1987; King and Just, 1991). In addition, the volume of blood flow in the brain is greater in language areas for object-extractions than for subject-extractions (Just et al., 1996a,b; Stromswold et al., 1996), and aphasic stroke patients cannot reliably answer comprehension questions about object-extracted RCs, although they perform well on subject-extracted RCs (Caramazza and Zurif, 1976; Caplan and Futter, 1986; Grodzinsky, 1989; Hickok et al., 1993)” (p. 2).

threshold). These are ranked in terms of their probability of success. Processing difficulties are explained, on a parallel model, by appeal to a computationally costly operation of *re-ranking* these analyses. The lower the probability of an analysis at step  $n$ , the more costly it is to rank that analysis higher upon encountering new input at step  $n+1$ .<sup>31</sup> For instance, when the word ‘fell’ comes in after the first several words of sentence (19), what initially seemed like a highly probable analysis must be demoted, while what initially seemed like highly improbable analysis must be ranked highest. Hale shows that it is precisely at this point that an Earley parser operating with a PCFG encounters serious difficulty, thereby simulating human performance.

Resolving the dispute between advocates of serial and parallel models would clearly yield a deeper understanding of the HSPM. However, our goal here is simply to show that some psychologically plausible models of the HSPM carry a commitment to the strong competence hypothesis and perhaps also the Representational Thesis (RT). And for *that* purpose, it is unnecessary to settle the serial-vs.-parallel debate, since *both* types of model are committed to the internal representation of a grammar.

In Hale’s model, the rules are explicitly represented—i.e., stored in a separate data structure and consulted as data, in accordance with the Earley algorithm. As noted in §1, however, such an algorithm can be “hardwired,” in which case the grammar is merely embodied, without being represented. This is consistent with the strong competence hypothesis, which Hale formulates as follows: “Strong competence holds that the human sentence processing mechanism *directly uses* rules of grammar in its operation” (p. 1, emphasis mine). As noted in §1, the Strong Competence Hypothesis is not necessarily equivalent to RT. Hale’s formulation, for instance, leaves it open whether the HSPM’s “direct use” of the rules of a grammar is a

---

<sup>31</sup> For further discussion, see Crocker, Pickering and Clifton (2000) and references therein.

matter of explicitly representing those rules or embodying them in a procedural fashion. Thus, Hale is neutral between position **(ii)** on the psychological reality issue, which is committed to RT, and position **(iii)**, which is not. (See §1 for precise statements of these positions.) Still, I think that the success of Hale's model provides reason to think that position **(iii)** is the *weakest* position that is consistent with what we know about human parsing performance. The discussion in this chapter can be seen as a sustained defense of that very claim.

This concludes our discussion of the top-down approach to parsing. We now turn to bottom-up parsers. Like their top-down counterparts, these have also been subjected to a dynamical programming treatment, and applied to both PCFGs and mildly context-sensitive grammars, including the Minimalist grammars that we discuss in §6.<sup>32</sup>

#### §2.4 *Bottom-up parsing*

Bottom-up parsing was the earliest approach to have been explored, with proposals dating back to 1955. The pioneering research was conducted by computer scientists, largely in connection with artificial programming languages.<sup>33</sup> Karttunen and Zwicky (1985) discuss the relation between parsing algorithms designed for these purposes and ones that explicitly seek to achieve psychological plausibility. As they point out, the concerns of psycholinguists and computer scientists overlap in a number of respects, though the differences between natural languages and programming languages sometimes obscure this. Whereas artificial languages tend to be simple, largely unambiguous, and have a known grammar, natural language is

---

<sup>32</sup> See also Schabes, Abeille, and Joshi (1988), who provide an instructive application of the Earley algorithm to lexicalized versions of context-free grammars, as well as to the mildly context-sensitive tree-adjoining grammar (TAG). The authors discuss the considerable gains in efficiency stemming from the lexicalization of these grammars. We discuss the lexically-driven approach to syntax in connection with Minimalism in §6

<sup>33</sup> See also Aho and Ullman (1972).

complex, rife with ambiguity, and not yet fully understood from a formal perspective.<sup>34</sup>

Moreover, the notions of communication, discourse, and semantic interpretation are significantly different when applied to programming languages, as against natural languages. Nevertheless, with respect to syntax and parsing, there is much that the computer scientist and the psycholinguist can learn from one another. The fact that bottom-up parsing is used in both computational and psycholinguistic applications is an instance of this fruitful exchange of ideas.

In contrast with the top-down parsers examined above, the simplest bottom-up parser begins by immediately looking at the input, rather than making predictions about it. (Aristotle, Locke, and Hume would have been proud.) Hence, the first step in a bottom-up parsing process is to search the lexicon in order to determine all of the possible grammatical categories to which the words in the input can belong. Once this is ascertained, the parser begins to build up all of the trees compatible with those categories and the grammar of the language. At each step of the process, the parser “looks for places in the parse in progress where the right-hand side of some [context-free] rule might fit” (Jurafsky and Martin, 2007: p. 434). As we will see below (fig. 10), a popular kind of bottom-up parser draws on explicit representations of the rules and uses them as templates to determine which parses are grammatically licensed at each ply.

Bottom-up parsers, better known to computer scientists as *shift-reduce* parsers, operate

---

<sup>34</sup> The terms “simple” and “complex” can be given a formal interpretation. Whereas programming languages tend to belong to the class of context-free languages, it has been known for some time that natural languages are slightly stronger than context-free. This was demonstrated by Shieber (1985), who discussed cases of cross-serial dependencies in Dutch and in Scandinavian languages. Syntacticians are in the process of constructing formalisms that fit this specification while avoiding overgeneration—i.e., without allowing the formulation of grammars that are not attested by any known natural language. The Minimalist grammars currently being explored in one branch of syntactic theory are committed to the existence of discontinuous constituents—syntactic chains in which antecedents do not c-command the traces or copies that they leave behind after movement. A grammar that generates discontinuous constituents is stronger than context-free. Minimalist grammars thus belong to the class of mildly context-sensitive grammars. (See Stabler, 2001.) Shieber, Schabes, and Pereira (1995) discuss several other mildly context-sensitive grammars, including the popular tree-adjoining grammars (TAGs). They provide a schema for constructing efficient parsers for these grammars, using the Parsing as Deduction (PAD) paradigm discussed in §2.7 below. Harkema (2001) and Hale (2003) have developed parsers for Minimalist grammars in accordance with PAD.

by attempting to integrate already recognized lexical items into phrasal groups.<sup>35</sup> The parsing process is completed when the partial structures that were built out of the items in the input are integrated into a full sentence—a structure with an S node at its root. In general, it will not be possible to incorporate *all* of the partial structures into a tree rooted in an S node.<sup>36</sup> Many of these will turn out to be “duds” and will have to be discarded. But if the input is a sentence of the language that is adequately captured by the parser’s internally represented grammar, then at least one of the parser’s initial guesses about the structure of the input will turn out to be correct.

Fig. 7 illustrates the first five plies of a bottom-up parse of (25), by reference to the grammar G, which we introduced in our discussion of top-down parsers (§2.2., fig. 2).

(25) Book that flight.

---

<sup>35</sup> Another term for these parsers that may be familiar to computer scientists is ‘LR(*k*)’. The symbol ‘LR’ encodes the fact that the parser works from left to right. The symbol ‘(*k*)’ is a variable that defines the size of the parser’s “look-ahead window”—i.e., how much of the input it is allowed to take in before initiating some operation. An LR(2) parser, for instance, can look at two items of the input before deciding what to do next.

<sup>36</sup> There is a way in which this sort of remark can be misleading. Devitt (2006: pp. 69-71) notes an important distinction between the expressions of a language, on the one hand, and structural *descriptions* of those expressions on the other. The latter can be derived from a theory of a language, but the former cannot. “[W]hat is derived from a grammar is not an expression of the language but a description of an expression, just as what is derived from an astronomical theory is not, say, a star, but a description of a star” (p. 69). If we are careful not to run afoul of this use/mention distinction, we must say that the parser uses its internally represented grammar—conceived now as (a subpersonal analogue of) a *theory* of a language—to generate *structural descriptions* of incoming linguistic stimuli. Hence, when we say that the parser produces “a structure that has an S node at its root,” we do *not* mean that it produces a sentence; rather the parser produces a description of the incoming stimulus, thus characterizing the stimulus *as* a structure that has an S node at its root. The question then arises: How do we get from such a descriptive characterization to the final product of language comprehension? On one view of that final product, this would be a bit of a mystery. For, on that view, the final representation *uses* words, rather than *mentioning* them. And if that is the case, then what accounts for the transition between a description (which merely mentions the words) to the final product (which uses them)? However, there is an alternative view of the final product, according to which it, too, merely mentions the words, rather than using them. This is the view that I suggested in ch. IV (§3.2).

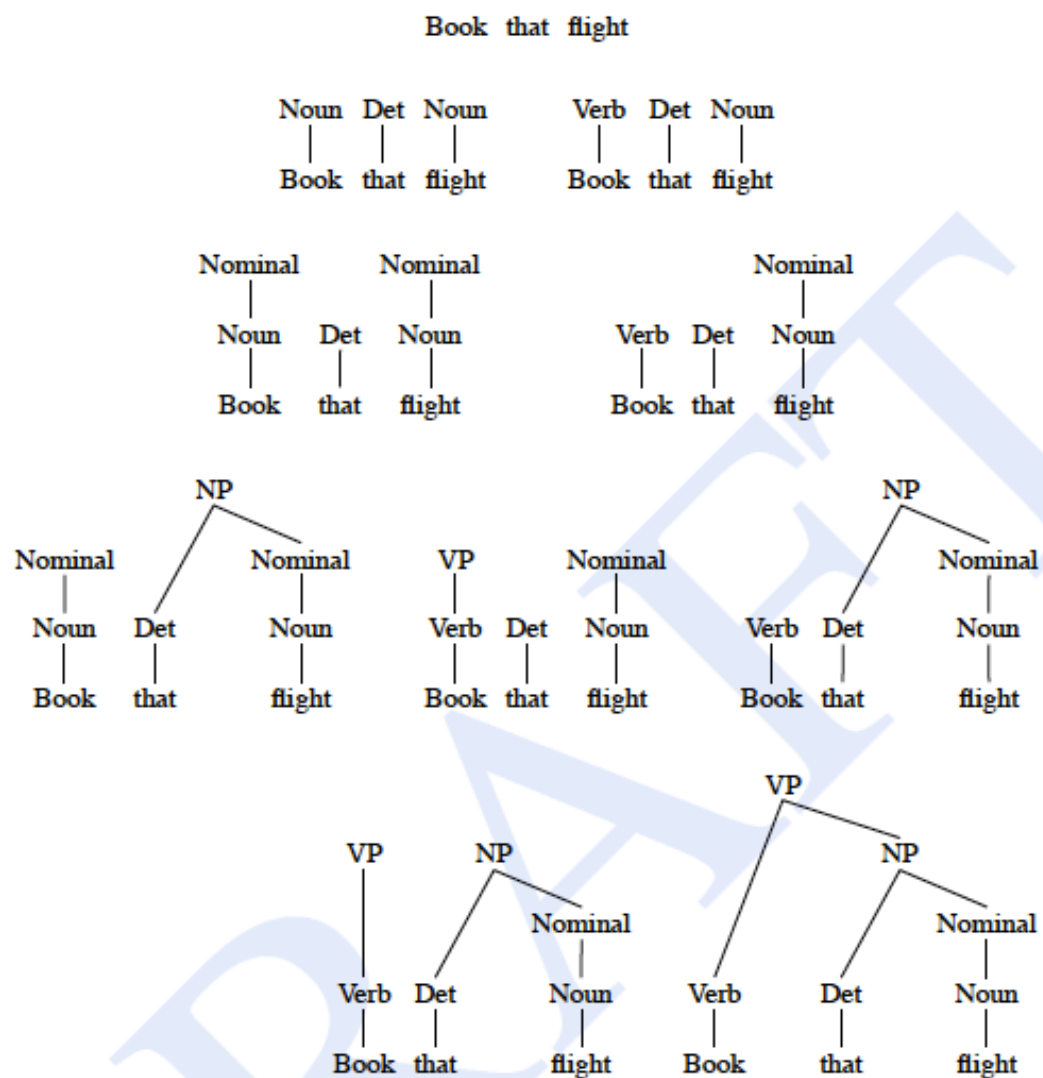


**Figure 7:** The first five plies of a bottom-up parse of *Book that flight*. Source: Jurafsky and Martin (2007), p. 435

Fig. 7 seems to suggest that all of the lexical items are recognized at the same time and that all of the structures are built (simultaneously) only after the full input sentence has been scanned. This need not be the case.<sup>37</sup> To see this point, let's examine the bottom-up parsers that computer

<sup>37</sup> Indeed, it had better *not* be the case, if psychological plausibility is our goal. As noted in chs. III and IV, the HSPM is an "eager" mechanism—the assignment of syntactic structure is never delayed; syntactic analysis begins at the very first hint of linguistic input and continues incrementally, morpheme by morpheme.

scientists call *shift-reduce* parsers, which operate on essentially the principles described above. These have two basic operations: *Shift* and *Reduce*. The *Shift* operation puts new material from the input into a memory stack. The *Reduce* operation unifies distinct items into a single unit. It does so by taking sequences of items from the stack and, if they match the right-hand side of an internally represented grammar rule, replacing them by the left-hand side. For instance, if the stack contains an NP and a VP, and the grammar contains the rule  $S \rightarrow NP VP$ , then the NP and VP will be “reduced” to an S—i.e., the system will posit an S node that immediately dominates both the NP and the VP. This process is illustrated in fig.8.<sup>38</sup>

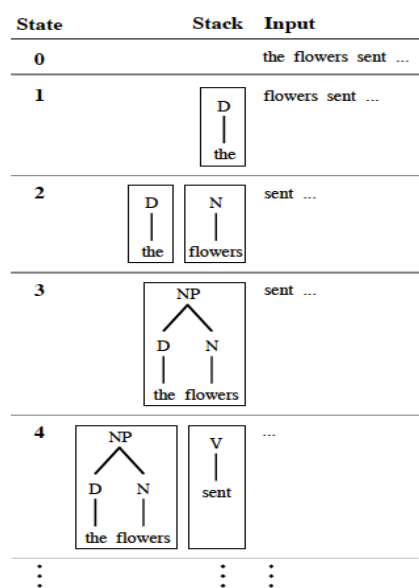


**Figure 8:** An incremental bottom-up parse of *The flowers sent....*  
Source: Shieber and Johnson (1993)

Jurafsky and Martin (2007) discuss the relative efficiency of bottom-up and top-down parsers. The inefficiency of top-down parsers, we saw, stemmed from the fact that they make predictions without first looking at the input, thus wasting resources on predictions that are ultimately found to be inconsistent with the input. Although it turns out that bottom-up parsing

<sup>38</sup> Similar visual aids can be found in Abney and Johnson (1991) and Crocker (1999).

is, in practice, more efficient than top-down parsing (O'Donnell, 1995: ch. 9), there are nevertheless several problems with bottom-up parsers as well.

Looking at fig. 7 above, we notice that many of the structures that the parser considers at the first ply do not—indeed, cannot—eventuate in an S. Building these structures is therefore a waste of time and computational resources. Moreover, in the case of long, highly-ambiguous inputs, one and the same mistake might be made several times over. A top-down parser would never make such mistakes, because it starts out predicting only legal trees—i.e., those that have an S node at their root. By contrast, a bottom-up parser has no way of knowing in advance which structures will be successful as higher and higher nonterminal nodes are “reduced.” As Marcus (1980) argues, a purely bottom-up parser must have some mechanism for either ignoring the mistakes it makes or fixing them when necessary. The former option is inherent in the parallel parsing approach, which simply sets aside the unsuccessful parse trees that are constructed on the way to the final, successful parse. However, parallel architectures require large amounts of memory or storage capacity. The other option—fixing mistakes—necessitates backtracking and revising earlier commitments. As noted above, backtracking requires a great deal of time and resources in realistic applications.

### §2.5 *The Cocke–Younger–Kasami (CYK) Algorithm*

We saw earlier that some of the inefficiency associated with a simplistic top-down approach can be dealt with by the “dynamical programming” technique of storing successful partial parses in a data structure known as a “chart,” and calling on them in order to construct a solution to a larger problem. A similar technique is available for the bottom-up approach. A particularly elegant parsing algorithm that makes use of dynamical programming was discovered



in the late 1960s by three separate researchers: John Cocke, Daniel Younger, and Tadao Kasami. The algorithm is frequently labeled ‘CYK’, in honor of its discoverers. Drawing on the discussion and the visual aids in Jurafsky and Martin (2007), I describe the basic operations of the CYK algorithm, using as an example the input sentence (26).<sup>39</sup>

(26) Book the flight through Houston

First, a chart is constructed—a matrix of cells with individual cells corresponding to the words in the input. This is shown in fig. 9a. The cells marked [0,1], [1,2], [2,3], [3,4], and [4,5] correspond to the words ‘Book’, ‘the’, ‘flight’, ‘through’, and ‘Houston’, respectively. The algorithm begins by using its lexical rules to identify the parts of speech to which these words belong. It then runs through the remaining cells from left to right, bottom to top, as illustrated in fig. 9b. In accordance with the rules of an internally represented context-free grammar,<sup>40</sup> cells that occupy higher positions in each column are filled in with higher-level phrasal categories, on the basis of the contents of the cells that have already been viewed. For instance, in accordance with the rule  $NP \rightarrow Det\ Noun$ , an NP is placed in cell [1,3] on the basis of there being a determiner (*Det*) in cell [1,2] and an noun in cell [2,3]. Similarly, the rule  $VP \rightarrow Verb\ NP$  is used to fill cell [0,3] with a VP, on the basis of there being a *Verb* in cell [0,1] and an NP in cell [2,3]. This process is repeated, left-to-right, bottom-to-top until all of the cells have been filled. The success of the parse is indicated by the presence of an S node in the very last cell, marked [0,5].

---

<sup>39</sup> An animated demonstration of the CYK parser is available at the URL below. Note, however, that the arrangement of the cells in the chart in the animation differs from the one depicted in fig. 9. To be sure, the difference is merely notational. <http://homepages.uni-tuebingen.de/student/martin.lazarov/demos/cky.html>

<sup>40</sup> Note that the CFG must first be transformed into a binary-branching format known as Chomsky Normal Form (CNF). This transformation is well-defined and computationally trivial. See Jurafsky and Martin (2007: pp. 441-2).

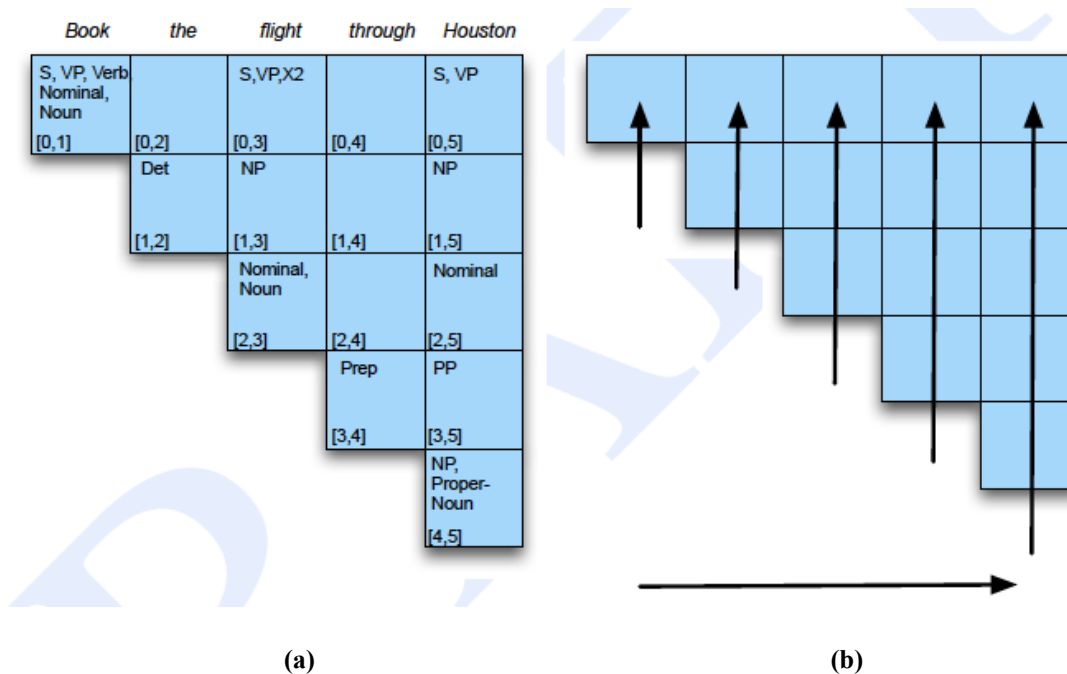


**Figure 9:** The CYK algorithm in action. The figure in (a) shows a completed chart for sentence (26). The cells at the bottom of each column correspond to the words in the sentence and encode their parts of speech. The algorithm proceeds from left to right, bottom to top, as shown in figure (b). *Source:* Jurafsky and Martin (2007), p. 443

Just as we did with the Earley algorithm, we can survey the pseudocode of the CYK algorithm and see that explicit reference is made to a data-structure in which the rules of the grammar are explicitly represented.

```

function CKY-PARSE(words, grammar) returns table ← !!!
  for j ← from 1 to LENGTH(words) do
    table[j-1, j] ← {A | A → words[j] ∈ grammar } ← !!!
    for i ← from j-2 downto 0 do
      for k ← i+1 to j-1 do
        table[i, j] ← table[i, j] ∪
          {A | A → BC ∈ grammar, ← !!!
            B ∈ table[i, k],
            C ∈ table[k, j] }

```

**Figure 10:** The pseudocode for the CYK algorithm. I have marked explicit references to grammatical rules with the symbol ‘!!!’. *Source:* Jurafsky and Martin (2007), p. 444.

Similarly, the CYK algorithm has an elegant extension to PCFGs. The probabilistic information contained in such a grammar can be computed for each cell in the manner illustrated in fig. 11.

|                   |                                     |
|-------------------|-------------------------------------|
| Det: .40<br>[0,1] | NP: .30*.40*.02<br>= .0024<br>[0,2] |
|                   | N: .02<br>[1,2]                     |

the                      dog

$S \rightarrow NP VP$  .80  
 $NP \rightarrow Det N$  .30  
 $VP \rightarrow V NP$  .20  
 $V \rightarrow looks$  .05  
 $Det \rightarrow the$  .40  
 $Det \rightarrow a$  .50  
 $N \rightarrow meal$  .01  
 $N \rightarrow dog$  .02

**Figure 11:** The CYK algorithm applied to probabilistic CFGs. The input is the phrase ‘the dog’. Here we see the parser computing a probability estimate of there being an NP in cell [0,2], on the basis of the presence of a determiner (Det) in cell [0,1], a noun in cell [1,2], and the rule  $NP \rightarrow Det N$ . The probability of the determiner in cell [0,1] is .4 and the probability of the noun in cell [1,2] is .02. This is then multiplied by the probability of applying the  $NP \rightarrow Det N$  rule, which is .3. The calculation  $.3 \times .4 \times .02$  yields the value .0024 in cell [0,2].

Experiments with bottom-up parsers that draw on probabilistic grammars have lent support to the claim that such schemes are psychologically plausible. Brants and Crocker (2000a,b) review a number of experiments designed to “counter... the possible criticism that probabilistic parsers are too powerful and resource intensive to be considered as the basis of a cognitively plausible model” (2000b: p. 666). They conclude that

wide-coverage, probabilistic parsers do not suffer impaired accuracy when subject to strict cognitive memory limitations and incremental processing. Furthermore, parse times are substantially reduced. This suggests that it may be fruitful to pursue the use of these models within computational psycholinguistics, where it is necessary to explain not only the relatively rare ‘pathologies’ of the human parser, but also its more frequently observed accuracy and robustness. (2000a: p. 117)

The reader may also be interested to learn that the CYK algorithm has been successfully implemented in connectionist networks. This requires converting a context-free grammar into what Smolensky and Legendre (2006) call a *Harmonic Grammar*. Entering into the details would, unfortunately, take us too far afield. A full description of this implementation can be found in Hale (1999) and Hale and Smolensky (2006).

## §2.6 *Left-corner Parsing*

Purely top-down and bottom-up parsers have problems with memory capacity, some of which we mentioned above. Harkema (2001) summarizes one of the problems as follows:

Pure bottom-up and top-down parsers are generally rejected as adequate models for human sentence processing because they require unbounded memory for right-branching and left-branching structures respectively, while the human sentence processor, which is equipped with finite memory, seems to have no particular problems with structures of this kind (e.g., Crocker, 1999).<sup>41</sup>

There is, however, a third approach, which combines the insights of both top-down and bottom-up parsers. Known as *left-corner parsing*, this technique uses bottom-up methods to look at the input, thus avoiding the pitfalls associated with blind guessing, and then top-down methods to make informed predictions about further input. The strategy is to build each constituent bottom-up and then posit a plausible parent node as well as whatever further structure is required by that node. Fig. 12 illustrates this idea.<sup>42</sup> Note again that which predictions are made on the basis of the input is a function of what grammar the algorithm represents, either explicitly or implicitly.

As many authors have pointed out, left-corner parsing is psychologically plausible on a

---

<sup>41</sup> It should be noted, however, that Sahoo (2011) casts doubt on Harkema's claims: "although it has been believed that simple implementations of top-down parsing cannot accommodate direct and indirect left-recursion and may require exponential time and space complexity while parsing ambiguous context-free grammars, more sophisticated algorithms for top-down parsing have been created, e.g. by Frost, Hafiz, and Callaghan (2007), which accommodate ambiguity and left recursion. Their algorithm is able to produce both left-most and right-most derivations of an input with regard to a given context-free grammar." The reference is to Frost, R., Hafiz, R. and Callaghan, P. (2007). "Modular and Efficient Top-Down Parsing for Ambiguous Left-Recursive Grammars," 10th International Workshop on Parsing Technologies (IWPT), ACL-SIGPARSE, pp. 109-120, June 2007, Prague.

<sup>42</sup> Similar visual aids can be found in Abney and Johnson (1991) and Crocker (1999).

number of measures.<sup>43</sup> First, it is incremental, which we know the HSPM to be (ch. III.)

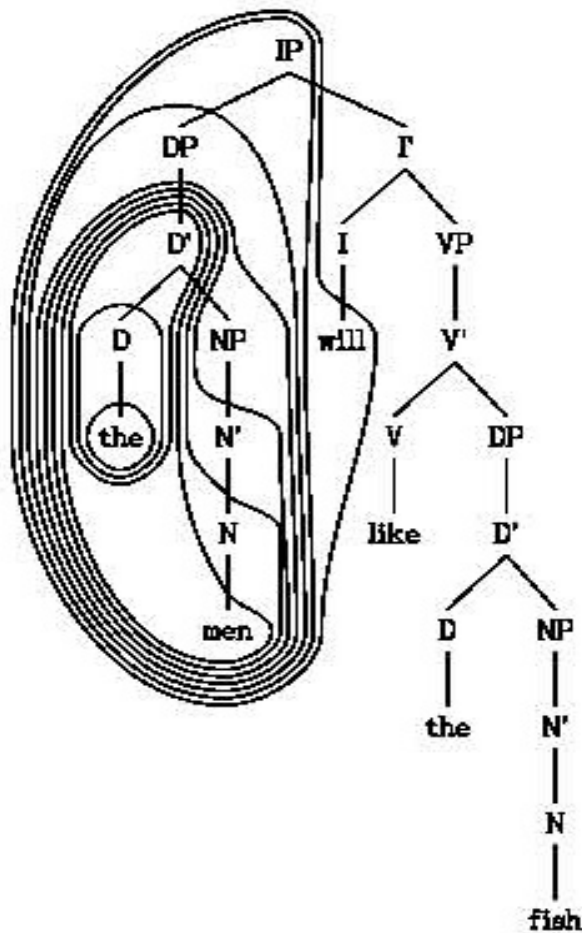


**Figure 12:** A left-corner parse of *The men will like the fish*. The figure is from Stabler (1994), who describes the process as follows: “We can diagram an LC parse of a sentence by circling the nodes that have been completed at each step. We have done this for the first 10 steps of a left corner parse [here]. The steps indicated there are the following, beginning with the smallest circle:

- i. The word ‘the’ is heard.
- ii. The parent D is built, since the, its left corner, has been completed.
- iii. The parent D is built, since its left corner D is complete, and the sibling NP is predicted.

<sup>43</sup> See Abney and Jonson (1993), Stabler (1994), Crocker (1999), Harkema (2001). Abney and Jonson (1993) distinguish between what they call arc-eager and arc-standard left-corner parsers. The latter, they argue, are less efficient than the former. I omit the details here. Note also that left-corner parsing bears a close resemblance to what Frazier and Fodor (1980) term ‘information-paced parsing’. Frazier and Fodor argue that strictly top-down and bottom-up routines are too rigid to deal effectively with locally ambiguous inputs.

Notice that completed categories are indicated by upward arcs across each circle, and predicted categories are indicated by downward arcs. The arcs which connect the completed structure to the rest of the structure are what must be kept on the stack in standard stack-based implementation strategies—these are the points in the structure that the parser still needs to be working on, so they must be held in memory.

- iv. The word *men* is heard.
- v. The parent *N* is built, since its left corner is complete.
- vi. The parent *N* is built, since its left corner is complete.
- vii. The parent *NP* is built, since its left corner is complete, and this constituent is attached as the predicted *NP*.
- viii. The parent *DP* is built, since its left corner *D* is complete.
- ix. The parent *IP* is built, since its left corner *DP* is complete.
- x. The word ‘*will*’ is heard.”

Second, unlike the bottom-up and top-down parsers, which have problems with right- and left-branching structures, the left-corner parser requires unbounded memory only for center-embedded structures. (Fig. 13 provides schemas for the three types of structure.) These are known to be problematic for human readers and listeners, as the example in (27) illustrates.<sup>44</sup>

**(27) The guitar the woman the man loved played is broken.**

For these reasons, many researchers now subscribe to the view that the HSPM is itself a left-corner parser.



**Figure 13:** Possible tree-geometries. *Source:* Abney and Jonson (1993)

It is important to note that dynamical chart-parsing techniques like the Earley and CYK algorithms discussed above have been applied to the left-corner parsing approach. Moore (2000) presents a comparison between the performance of a left-corner chart-parsing system and that of

<sup>44</sup> For an in-depth discussion of center-embedding, see Thomas (1995). Stabler (1994) discusses a range of center-embedded structures and related constructions from languages other than English.

previous parsers. His analysis concludes that left-corner chart-parsers enjoy an overall lead in efficiency. Finally, left-corner *probabilistic* approaches have also been explored, with promising results. Roark and Johnson (1998) conclude a review of their preliminary studies with the following remarks.

We have examined several probabilistic predictive parser variations, and have shown the approach in general to be a viable one, both in terms of the quality of the parses, and the efficiency with which they are found. We have shown that the improvement of the grammars with non-local information not only results in better parses, but guides the parser to them much more efficiently, in contrast to dynamic programming methods. Finally, we have shown that the accuracy improvement that has been demonstrated with left-corner approaches can be attributed to the non-local information utilized by the method. This is relevant to the study of the human sentence processing mechanism insofar as it demonstrates that it is possible to have a model which makes explicit the syntactic relationships between items in the input incrementally, while still scaling up to broad-coverage. (p. 426)

## §2.7 *Parsing as Deduction*

An exciting development in parsing theory, dating back to Pereira and Warren (1983), is the interpretation of the parsing process as a natural deduction procedure in first-order logic. This has been dubbed the *Parsing as Deduction* approach (henceforth, PAD). As the label suggests, a theorist who adopts this approach sees parsing as a sort of deductive process—a species of theorem-proving—in which the grammar is formulated as a set of axioms. The deductive procedure begins with these axioms and derives a theorem to the effect that the input string is a sentence of the language generated by the grammar. Harkema (2001) summarizes the approach in the following passage:

Taking a logical perspective on parsing as presented in [PW83], [SSP95] and [Sik97], the definition of the recognizer includes a specification of a grammatical deductive system and a specification of a deduction procedure. The formulae of the deductive system, which are commonly called items, express claims about grammatical properties of strings. Under a given interpretation, these claims are either true or false. For a given grammar and input string, there is a set of items that, without proof, are taken to represent true grammatical claims. These are the axioms of the deductive

system. Goal items represent the claim that the input string is in the language defined by the grammar. Since our objective is to recognize a string, the truth of the goal items is of particular interest. The deductive system is completed with a set of inference rules, for deriving new items from ones derived earlier. The other component of the definition of the recognizer is the specification of a deduction procedure. This is a procedure for finding all items that are true for a given grammar and input string. (p. 96)

PAD is the most vivid and efficient implementation of the familiar idea that grammatical rules are explicitly represented and used “as data.” Here, we introduce the approach by reference to CFGs. In §§5-6, we discuss its application to GB and Minimalist grammars.

Notice that any CFG rule can be formulated in the language of first-order quantified predicate logic (with identity). For instance, the rule  $S \rightarrow NP VP$  would be translated into (28).

$$(28) \quad \forall x [\exists yz [(NP(y) \ \& \ VP(z)) \ \& \ x = yz] \Rightarrow S(x)]$$

In their classic discussion of PAD, Pereira and Warren (1983) opt for a slightly different notation, known as Definite Clause Grammar (DCG). In DCG, the rule  $S \rightarrow NP VP$  would be translated into the formula in (29). (All variables are implicitly quantified.)

$$(29) \quad S(s_0, s_2) \Leftarrow NP(s_0, s_1) \ \& \ VP(s_1, s_2)$$

Pereira and Warren use the ‘s’-variables to represent positions in the input string. Hence, (29) can be paraphrased as follows: The material between positions 0 and 2 in the input string constitutes a sentence if between positions 0 and 1 there is a noun phrase and between positions 1 and 2 there is a verb phrase. Pereira and Warren go on to show how the Earley algorithm discussed above can be reinterpreted in the PAD framework. The basic functions of the parser—*Scan*, *Predict*, and *Complete*—are interpreted as inference rules (on a par with, say, *modus ponens* and *existential instantiation* in first-order logic).

The example below, adapted from Pereira and Warren (1983) shows how the Earley algorithm would appear when rendered in the deductive formalism. In the justifications table



(which appears on the next page) *instantiation* corresponds to the *Predict* function, while *reduction* corresponds to the *Scan* and *Complete* functions. The sentence in (30) is assumed to have been divided into distinct morphemes. The outputs of a background lexical decision process are treated as axioms for the purpose of the deduction.

(30) <sub>0</sub> Agatha <sub>1</sub> 's <sub>2</sub> husband <sub>3</sub> hit <sub>4</sub> Ulrich <sub>5</sub>

| Lexical Rules | DCG notation | English translation                                     |
|---------------|--------------|---------------------------------------------------------|
| Axiom 1       | N(0,1)       | The item between positions 0 and 1 is a noun            |
| Axiom 2       | Gen(1,2)     | The item between positions 1 and 2 is a genitive marker |
| Axiom 3       | N(2,3)       | The item between positions 2 and 3 is a noun            |
| Axiom 4       | V(3,4)       | The item between positions 3 and 4 is a verb            |
| Axiom 5       | N(4,5)       | The item between positions 4 and 5 is a noun            |

| Grammar Rules | DCG notation                                          | Context-free notation       |
|---------------|-------------------------------------------------------|-----------------------------|
| Axiom 6       | $S(s_0,s) \Leftarrow NP(s_0,s_1) \ \& \ VP(s_1,s)$    | $S \rightarrow NP \ VP$     |
| Axiom 7       | $NP(s_0,s) \Leftarrow Det(s_0,s_1) \ \& \ N(s_1,s)$   | $NP \rightarrow Det \ N$    |
| Axiom 8       | $Det(s_0,s) \Leftarrow NP(s_0,s_1) \ \& \ Gen(s_1,s)$ | $Det \rightarrow NP \ Gen$  |
| Axiom 9       | $Det(s_0,s) \Leftarrow Art(s_0,s)$                    | $Det \rightarrow Art$       |
| Axiom 10      | $Det(s,s)$                                            | $Det \rightarrow \emptyset$ |
| Axiom 11      | $S(s_0,s) \Leftarrow NP(s_0,s_1) \ \& \ VP(s_1,s)$    | $VP \rightarrow V \ NP$     |

| Deductive procedure                                       | Justifications                    |
|-----------------------------------------------------------|-----------------------------------|
| Line 1 $ans \Leftarrow S(0,5)$                            | Goal statement (i.e., assumption) |
| Line 2 $S(0,5) \Leftarrow NP(0,s_1) \ \& \ VP(s_1,5)$     | Line 1 instantiates Axiom 6       |
| Line 3 $NP(0,s) \Leftarrow Det(0,s_1) \ \& \ N(s_1,s)$    | Line 2 instantiates Axiom 7       |
| Line 4 $Det(0,s) \Leftarrow NP(0,s_1) \ \& \ Gen(s_1,s)$  | Line 3 instantiates Axiom 8       |
| Line 5 $Det(s_0,s) \Leftarrow Art(0,s)$                   | Line 3 instantiates Axiom 9       |
| Line 6 $NP(0,s) \Leftarrow N(0,s)$                        | Axiom 10 reduces Line 3           |
| Line 7 $NP(0,1)$                                          | Axiom 1 reduces Line 6            |
| Line 8 $S(0,5) \Leftarrow VP(1,5)$                        | Line 7 reduces Line 2             |
| Line 9 $VP(1,5) \Leftarrow V(1,s_1) \ \& \ NP(s_1,5)$     | Line 8 instantiates Axiom 11      |
| Line 10 $Det(0,s) \Leftarrow Gen(1,s)$                    | Line 7 reduces Line 4             |
| Line 11 $Det(0,2)$                                        | Axiom 2 reduces Line 10           |
| Line 12 $NP(0,s) \Leftarrow N(2,s)$                       | Line 11 reduces Line 7            |
| Line 13 $NP(0,3)$                                         | Axiom 3 reduces Line 12           |
| Line 14 $S(0,5) \Leftarrow VP(3,5)$                       | Line 13 reduces Line 2            |
| Line 15 $Det(0,s) \Leftarrow Gen(3,s)$                    | Line 13 reduces Line 4            |
| Line 16 $VP(3,5) \Leftarrow V(3,s_1) \ \& \ NP(s_1,5)$    | Line 14 instantiates Axiom 11     |
| Line 17 $VP(3,5) \Leftarrow NP(4,5)$                      | Axiom 4 reduces Line 16           |
| Line 18 $NP(4,5) \Leftarrow Det(4,s_1) \ \& \ N(s_1,5)$   | Line 17 instantiates Axiom 7      |
| Line 19 $Det(4,s) \Leftarrow NP(4,s_1) \ \& \ Gen(s_1,s)$ | Line 18 instantiates Axiom 8      |

|         |                                                                      |                              |
|---------|----------------------------------------------------------------------|------------------------------|
| Line 20 | $\text{Det}(4,s) \Leftarrow \text{Art}(4,s)$                         | Line 18 instantiates Axiom 9 |
| Line 21 | $\text{NP}(4,s) \Leftarrow \text{Det}(4,s_1) \ \& \ \text{N}(s_1,5)$ | Line 19 instantiates Axiom 7 |
| Line 22 | $\text{NP}(4,5) \Leftarrow \text{N}(4,5)$                            | Axiom 10 reduces Line 18     |
| Line 23 | $\text{NP}(4,s) \Leftarrow \text{N}(4,s)$                            | Axiom 10 reduces Line 21     |
| Line 24 | $\text{NP}(4,5)$                                                     | Axiom 5 reduces Line 22      |
| Line 25 | $\text{VP}(3,5)$                                                     | Line 24 reduces Line 17      |
| Line 26 | $\text{Det}(4,s) \Leftarrow \text{Gen}(5,s)$                         | Line 24 reduces Line 19      |
| Line 27 | $\text{S}(0,5)$                                                      | Line 25 reduces Line 14      |
| Line 28 | ans                                                                  | Line 27 reduces Line 1       |

Shieber, Schabes, and Pereira (1995) extend the work of Pereira and Warren (1983) by subjecting the CYK algorithm and a variety of other parsing algorithms to the same treatment.

[W]e will in this paper investigate ... how to synthesize parsing algorithms by combining specific logics of grammaticality claims with a fixed search procedure. In this way, deduction can provide a metaphor for parsing that encompasses a wide range of parsing algorithms for an assortment of grammatical formalisms. We flesh out this metaphor by presenting a series of parsing algorithms literally as inference rules, and by providing a uniform deduction engine, parameterized by such rules, that can be used to parse according to any of the associated algorithms. The inference rules for each logic will be represented as unit clauses and the fixed deduction procedure, which we provide a Prolog implementation of, will be a version of the usual bottom-up consequence closure operator for definite clauses. As we will show, this method directly yields dynamic-programming versions of standard top-down, bottom-up, and mixed-direction (Earley) parsing procedures. (p. 4)

They also go on to apply the PAD approach to syntactic formalisms other than the context-free grammars that we've been considering thus far, including combinatory categorial grammars (CCGs) and tree adjoining grammars (TAGs). Harkema (2001) and Hale (2003) apply the approach to Minimalist grammars (§6). These richer formalisms (slightly stronger than context-free) are widely taken to be capable of providing a more descriptively adequate treatment of natural language than CFGs. We shall see in §5 that principle-based grammars, such as GB theory, likewise receive a natural interpretation in the PAD framework. For this reason, the PAD approach has been crucial in the development of principle-based parsers.<sup>45</sup>

---

<sup>45</sup> See Berwick (1991a,b) and Johnson (1989; 1991).

From the point of view of a computational linguist, the PAD approach has an immediate payoff in that it makes possible the application of well-known logic programming techniques to the task of parsing natural language. The programming languages *LISP* and *Prolog* have been adapted to natural language processing with great success. For *philosophical* purposes, the payoff is quite different, though no less intriguing.

It is difficult to find up-and-running examples of psychological processes *convincingly* modeled as deductive operations defined over truth-evaluable statements. The PAD approach shows that such an example is available in the case of at least one aspect of natural language comprehension. We should not, however, overstate the claim. It is not as though neurocognitive research has revealed that the language inscribed in the brain is first-order logic! Rather, what has been shown is that *if* the CYK or Earley algorithms play a role in a psychologically plausible model of the HSPM (as the results from Hale and others indicate), *then* there is a perfectly workable interpretation of the HSPM according to which the neural mechanisms that underpin it can be said to carry out deductive procedures. This, incidentally, substantiates my claim in ch. IV (§3.3.5) that the HSPM can exhibit a subpersonal analogue of inferential coherence.

The availability of an interpretation on which the HSPM is a deduction engine can be brought to bear on the recent debate in the philosophy of linguistics, concerning whether knowledge of language is, strictly-speaking, *propositional*.<sup>46</sup> The fact that a rich array of successful parsing models can be represented in a familiar propositional format shows, at the very least, that it is *not incoherent* to claim that knowledge of language should be propositional. Nevertheless, I will argue in §7 that it is better to regard the HSPM *subpersonal* system, which operates on representations that are not usefully assimilated to propositional attitudes.

---

<sup>46</sup> See Knowles (2000) and Rattan (2002).

## §2.8 Efficiency Considerations

As discussed above, both top-down and bottom-up parsers are, *in their purest and simplest forms*, wildly inefficient—so inefficient, indeed, as to be practically unusable for realistic natural language processing. Devitt (2006) draws on this point in the following argument against position **(ii)**—the idea that syntactic rules are represented and treated as data.

How can the represented rules be used as data in language use? Consider language comprehension. Suppose that, somehow or other, the processing rules come up with a preliminary hypothesis about the structure of the input string. In principle, the represented rules might then play a role by determining whether this hypothesis could be correct (assuming that the input is indeed a sentence of the language). The problem in practice is that to play this role the input would have to be tested against the structural descriptions generated by the rules and *there are just too many descriptions*. The “search space” is just too vast for it to be plausible that this testing is really going on in language use. (209)

In the preceding discussion we saw computational linguists have devised an array of strategies for dealing with this difficulty. Some of these are implementations of the least-effort parsing principles, MA, LC, and MCP (§1). Others are based on clever computational tricks, like the storage and reuse of partial solutions.<sup>47</sup> Still others have a different flavor, making heavy use of statistical methods to reduce error and avoid ambiguity. For present purposes, we can remain neutral on the issue of which of these approaches is optimal. The important point is various refinements can be made to models that are committed to position **(ii)**, though it remains to be seen whether these refinements can fully eliminate the inefficiency that Devitt points to.

---

<sup>47</sup> In principle-based parsers, “interleaving” the principles of a modular grammar like GB leads to impressive gains in efficiency. See Berwick (1991a, 1991b) and Merlo (1995).

### §3. The Psychological Reality of Transformational Rules

#### §3.1 *A Primer on Transformational Rules*

In an effort to capture a variety of generalizations about the relations between sentences of a natural language, Chomsky (1957) introduced the notion of a transformational rule. Such rules serve to map entire syntactic structures—typically for whole clauses—into *other* syntactic structures. The following examples, adapted from Fodor, Bever, and Garrett (1974: pp. 97-98) illustrate a number of transformations, as applied to (31)-(38).

|      |                                      |                           |
|------|--------------------------------------|---------------------------|
| (31) | Morris solved the problem.           | Active                    |
| (32) | The problem was solved by Morris.    | Passive                   |
| (33) | Morris didn't solve the problem.     | Negative                  |
| (34) | Did Morris solve the problem?        | Question                  |
| (35) | Was the problem solved by Morris?    | Passive question          |
| (36) | The problem wasn't solved by Morris. | Negative passive          |
| (37) | Didn't Morris solve the problem?     | Negative question         |
| (38) | Wasn't the problem solved by Morris? | Negative passive question |

The relations between (31)-(38) are encoded in the following transformations.<sup>48</sup>

|      |                                                                             |                                   |
|------|-----------------------------------------------------------------------------|-----------------------------------|
| (39) | $NP_1 V+ed NP_2 \leftrightarrow NP_2 \text{ was } V+ed \text{ by } NP_1$    | Passivization                     |
| (40) | $NP_1 V+ed NP_2 \leftrightarrow NP_1 \text{ didn't } V NP_2$                | Negation                          |
| (41) | $NP_1 V+ed NP_2 \leftrightarrow \text{Did } NP_1 V NP_2$                    | Question-formation                |
| (42) | $NP_1 V+ed NP_2 \leftrightarrow \text{was } NP_2 V+ed \text{ by } NP_1$     | Passive question-formation        |
| (43) | $NP_1 V+ed NP_2 \leftrightarrow NP_2 \text{ wasn't } V+ed \text{ by } NP_1$ | Passive negation                  |
| (44) | $NP_1 V+ed NP_2 \leftrightarrow \text{Didn't } NP_1 V NP_2$                 | Negative question-formation       |
| (45) | $NP_1 V+ed NP_2 \leftrightarrow \text{wasn't } NP_2 V+ed \text{ by } NP_1$  | Negative passive quest.-formation |

The transformations in (39)-(45) map active declarative structures into a variety of other sentence structures, and vice versa, thereby capturing the systematic relations between them.

The symbol ' $\leftrightarrow$ ', and the attendant notion of a "mapping," should be treated with care.

The transformation rules, *qua* structure rules, embody claims about the relations between

---

<sup>48</sup> My discussion here is oversimplified in an important respect. Chomsky did not posit separate transformations for each linguistic phenomenon. Some were dealt with by the ordered, serial application of several transformational rules. For instance, negative questions would have undergone two transformations: Negation and Question-formation. We return to this point in §3.2 below.

sentences. The rule in (39), for instance, is a claim to the effect that if a sentence has the syntactic structure  $NP_1 V+ed NP_2$ —as dictated by the context-free rewrite rules that we examined above—then there is another sentence in the language that has the syntactic structure  $NP_2 was V+ed by NP_1$ .<sup>49</sup> As before, without additional argument, no part of this explanatory framework entails anything about psychological or computational operations.

Transformations *presuppose* the prior application of rules that generate the initial structure—e.g., the context-free rewrite rules discussed in §2.<sup>50</sup> The grammar proposed by Chomsky (1957) is thus a theory pitched at (at least) *two* levels of description. The descriptions at the first level assign a structure to a string of words in accordance with, e.g., a context-free grammar. The descriptions at the second level assign a structure to a different, though related, string of words in accordance with the transformational rules. Using transformations, we can even eliminate from our grammar all of the context-free rewrite rules that generate the kinds of structures that appear on the right-hand sides of (39)-(45). Once the context-free rules assign structure to the active declarative sentence in (31), the remaining structures can be generated *directly* by the transformation rules, without appeal to the rules that form the “base” of the grammar. This eliminates redundancy, making for a more compact formalism.

From the point of view of a syntactician, compactness is a desirable property for a grammar to have. However, pending an argument based on verified information about the HSPM, compactness is *not* automatically a desideratum for a grammar designed for parsing (see Stabler, 1984). Hence, the mere fact that transformations render a descriptive grammar compact

---

<sup>49</sup> In the examples above, the sentences share many of their lexical items, but differ in meaning. Fodor, Bever, and Garrett (1974: ch. 3) discuss restrictions on the relationship between transformationally related sentences. I pass over these complications here.

<sup>50</sup> Other possibilities are certainly imaginable. See Chomsky (1965) for a proposal to enrich the base of the grammar.

does not *by itself* constitute a reason for asserting that transformations are psychologically real operations. There has been some confusion on this point. For instance, Maratsos and Wanner (1978) interpret Chomsky as denying this claim.

Chomsky has consistently assumed that the psychologically real representation of linguistic knowledge will be just that grammar that captures linguistically significant generalizations; according to Chomsky, ‘we have [such] a generalization where a set of rules about distinct items can be replaced by a single rule about the whole set...’ (Chomsky, 1965: p. 42). This assumption is fundamental to Chomsky’s claim for the psychological reality of transformational rules; it is by accumulating clear cases in which the transformational notation permits a set of nontransformational rules to be replaced by a single transformation that Chomsky builds a psychological case for transformations. (p. 158)

In defense of these authors, it must be noted that the corpus of Chomsky’s work in the 1960s and 70s provides somewhat contradictory clues about his commitments on the psychological reality issue. We examined the ensuing interpretive difficulties in chs. I and II.<sup>51</sup>

Following standard nomenclature, I will use the label “D-structure” to denote the level at which the base rules—context-free productions or some enrichment thereof (fn. 34)—yield a set of syntactic descriptions. Transformational rules are then brought to bear, mapping D-structure descriptions into others, which we will now call “S-structure” descriptions. (Informally, ‘D’ is a mnemonic for ‘deep’ and ‘S’ for ‘surface’.)

### §3.2 *On the Psychological Reality of Transformations*

In a highly influential discussion, Fodor, Bever, and Garrett (1974) reviewed a number of psycholinguistic studies that they interpreted as evidence against the psychological reality of transformational operations. The model that they contemplated construed the process of syntactic analysis as involving two phases. In the first phase, the HSPM computes an S-structure

---

<sup>51</sup> See also Devitt (2006), ch. 4.

representation of the linguistic input. In the second phase, this representation is transformed into a D-structure representation, from which semantic information would be recovered. The computation of the D-structure representation was to be accomplished by the *reverse* application of transformational rules. (We will sometimes refer to these as “backwards transformations.”)

The studies reviewed by Fodor, Bever, and Garrett (henceforth, FBG) were premised on the assumption, inherent in the *Standard Theory* of transformational grammar (Chomsky, 1957; 1965), that transformations must apply in a particular order. If this is correct, then the reverse process should proceed in a certain order as well. For instance, negative passive questions would have to go through more backwards transformations than negative declarative sentences in the course of comprehension. Similarly, passives would have to go through one more backwards transformation than actives. (See FBG, pp. 228 for details.)

FBG’s discussion takes aim at what they called the *Derivational Theory of Complexity* (DTC). According to the DTC, the transformations of the Standard Theory are psychologically real procedural operations. That is, the HSPM is seen as literally transforming syntactic representations in the ways that the rules specify.<sup>52</sup> Given that each step in the comprehension

---

<sup>52</sup> The DTC is often attributed to Miller and Chomsky (1963). Berwick and Weinberg (1984) claim that Miller and Chomsky “identified rules of the grammar with computational operations of the parser in a one-to-one fashion. This identification led to specific behavioral predictions, collapsing grammatical with processing complexity” (pp. 38-9). They quote the following a telling passage from Miller and Chomsky (1963): “The psychological plausibility of a transformational model of the language user would be strengthened, of course, if it could be shown that our performance on tasks requiring the appreciation of the structure of transformed sentences is some function of the nature, number, and complexity of the grammatical transformations involved” (p. 481). The core idea can, however, be stated in a way that avoids commitment to Chomsky’s Standard Theory and, more generally, to any specific grammar, transformational or otherwise. Berwick and Weinberg (1984) note that “Miller and Chomsky’s original (1963) suggestion is really that grammars be realized more or less directly as parsing algorithms. We might take this as a methodological principle. In this case, we impose the condition that the logical organization of rules and structures incorporated in a grammar be mirrored rather exactly in the organization of the parsing mechanism. We will call this *type transparency*” (p. 39). This idea corresponds directly to the strong competence hypothesis (Bresnan and Kaplan, 1982; Steedman, 2000; Hale, 2003) and positions (i) and (iii) from Devitt (2006). As many authors have noted (e.g., Philips, 1996), the troubles with the DTC do not automatically translate into an argument against type transparency or the strong competence hypothesis. Opting for a different grammar or a different parsing architecture can make all the difference.



process—i.e., each transformation—would have to take a determinate amount of time, it seems to follow that reaction times on comprehension tasks should track the number of backwards transformations that each sentence would have to undergo before being successfully comprehended. That, anyway, is what FBG took the theory to predict.

Shortly after the formulation of the Standard Theory, experimental results emerged which appeared to disconfirm these predictions. Let us consider just one of the studies that FBG draw on. The experiment reported in Slobin (1966) used active and passive sentences, of which half were semantically reversible and half nonreversible, as illustrated in (47)-(50).<sup>53</sup>

- |      |                                                              |                           |
|------|--------------------------------------------------------------|---------------------------|
| (47) | John saw Bill. / Bill saw John.                              | active, reversible        |
| (48) | John was seen by Bill. / Bill was seen by John.              | passive, reversible       |
| (49) | John ate the apple. / #The apple ate John.                   | active, nonreversible     |
| (50) | The apple was eaten by John. / #John was eaten by the apple. | passive,<br>nonreversible |

Slobin asked his subjects to match a picture to each such sentence, and measured their reaction times on this task. He found that in the nonreversible case, processing the active form (49) and the passive form (50) took an equal amount of time, contrary to the prediction of the DTC. On the strength of an array of experimental results such as this, FBG concluded that the DTC was probably false and that “there exist no suggestions about how a generative grammar might be concretely employed as a sentence recognizer in a psychologically plausible model” (p. 368).

FBG’s argument against the DTC prompted a number of reactions. Berwick and Weinberg (1982, 1983, 1984) point out, very helpfully, that four separate claims are in play:

- [i] The correct theory of syntax is directly realized as a model of sentence processing.
- [ii] The Standard Theory of transformational grammar is the correct theory of syntax.

---

<sup>53</sup> The sentences marked with a ‘#’ exhibit a semantic defect or pragmatic infelicity. Following the policy of ch. III, the present discussion does not presuppose any particular way of drawing the semantics/pragmatics distinction.

[iii] Active and passive sentences are processed in the same amount of time.

[iv] The time it takes to process a sentence is an accurate measure of processing complexity.

FBG's argument consists in pointing out that these claims comprise an inconsistent set—at least one member of which has to be given up. FBG suggest that the weakest link is claim [i]. But Berwick and Weinberg point out that we may, instead, hold on to claim [i] and reject any of the others. For instance, Bresnan (1978) gives up claim [ii], opting for an alternative theory of syntax.<sup>54</sup> Bresnan famously argued that the experimental results show that the Standard Theory of transformational grammar must be replaced by a different grammar—one that *can* be put to use in a psychologically plausible processing model. On these and other grounds, she proposed a new formalism, known as Lexical Functional Grammar (LFG), a framework in which the effects of some transformations are offloaded to the lexicon. For instance, whereas the Standard Theory had a transformation for turning actives into passives, LFG posits two distinct lexical entries for the active and the passive forms of a verb.<sup>55</sup> This appears to have consequences for psycholinguistics. If LFG were implemented in a real-time processing model, such information would be *precomputed* and stored in lexical memory. And because accessing one lexical item is just as quick as accessing another, active and passive sentences are predicted take an equal amount of time to process, in accordance with Slobin's experimental result.<sup>56</sup>

Forster and Olbrei (1973) challenge claim [iii], presenting data that conflicts with Slobin's results. Using an experimental paradigm known as rapid serial visual presentation

---

<sup>54</sup> Berwick and Weinberg (1984) note that the revisions need not be as extensive as Bresnan envisions. Minor revisions—e.g., eschewing the so called “whiz-deletion” transformation—might do the trick.

<sup>55</sup> We saw a version of this strategy in ch. IV, in our discussion of the models developed by Schank and his colleagues.

<sup>56</sup> This prediction only holds when the active and passive sentences are matched in all other relevant respects, e.g., sentence length and the frequency of lexical items.

(RSVP), Forster and Olbrei found that nonreversible passives took significantly *longer* to process than nonreversible actives. They also argued that Slobin's experimental paradigm—matching sentences to pictures—does not reveal anything about sentence processing *per se*. In support of this, they pointed out that subjects were slower at matching passive sentences to pictures even when there was a 3 second delay between the presentation of the sentences and the pictures. It's known that the HSPM takes less than 3 seconds to process a sentence, regardless of whether the sentence is an active or a passive. A reasonable conclusion, then, is that the slower response times that Slobin (1966) observed are better explained by appeal to processing effects *outside* of the HSPM, perhaps in the visual system or the language-vision interface.

A different sort of criticism of FBG's argument is due to Berwick and Weinberg, (henceforth, BW). They concentrate on claim [iv], which concerns the relationship between reaction-time data and processing complexity. They point out that FBG's argument assumes the HSPM conducts its operations *serially*, first computing a *full* S-structure representation and only later computing the D-structure. BW argue that this assumption is groundless. In support of this, they sketch a parsing model (closely resembling that of Marcus, 1980), on which S-structure and D-structure representations are computed simultaneously, *in parallel*.<sup>57</sup> On such a model, actives and passives would be processed in an equal amount of time (*ceteris paribus*, fn. 56)

BW conclude that chronometric behavioral data can have a bearing on claims about the psychological plausibility of specific grammars *only when the parsing architecture and complexity measure have been fixed*. Given that we do *not* know what these parameters are in the case of the HSPM, it is unwarranted to conclude that transformational grammar is not at the

---

<sup>57</sup> NB: This is just one kind of parallelism; it is *not* the same as the kind discussed above, in connection with the Earley parser. Nor does 'parallel' mean the same here as it does in the label 'Parallel Distributed Processing'. There are, then, at least three distinct kinds of parallelism. See Vasishth and Lewis (2006: p. 419).

core of a psychologically plausible parsing model. In light of such considerations, many psycholinguists came to see the DTC untestable, hence irrelevant. As one critic put it:

Psycholinguists are entitled to ignore the DTC not because they showed it to be untrue, but because the experiments that have been performed are irrelevant to its truth, and it is difficult to see how to devise experiments that would test it without having to make arbitrary assumptions about the other parts of the language understanding system.”<sup>58</sup>

BW also challenged a pair of assumptions made by Bresnan (1978). Bresnan’s argument for the psychological plausibility of a parsing model that draws on LFG rested in part on the claim that computing a structure on-line is more costly than accessing precomputed information. BW point out that, in a parallel processing architecture, computation is cheap. And, lacking a detailed knowledge of the specific neural mechanisms underlying language use, we cannot tell whether the HSPM has devoted its resources to memory access or to online computation.

Bresnan also assumes that *only* a parser that draws on LFG can make use of precomputed structural information. In response, BW note that a parser can draw on a *modified* version of transformational grammar. For instance, in such a parser, “bounded movement”—i.e., the kind of NP-movement associated with passive and raising constructions—has been precomputed and stored in mental lexicon. “Transformational grammar ... provides the necessary principles to distinguish among rule types so as to permit some rules to be handled by memory retrieval and others by active computation, all within a machine architecture like that assumed in the DTC or in Bresnan 1978” (BW 1984: p. 75).

If this strategy were adopted, then the grammar used for parsing would not be *transparently* related to the transformational grammar advocated by Chomsky and others. But the resemblance between them would still be substantive. And the grammar used for parsing

---

<sup>58</sup> Garnham, A. (1983). “Why Psycholinguists Don’t Care About the DTC: A Reply to Berwick and Weinberg,” *Cognition*, 15, pp. 263-269.

would differ in important ways from Bresnan's LFG.

[Bresnan assumes] that the transformations postulated by EST [the *Extended Standard Theory* of transformational grammar] should be thought of as “active,” time-consuming computations. However, the necessity of this assumption is unclear unless there is also insistence upon some strong version of transparency and we state the Move NP rule and its parsing correlate in exactly the same form. If transparency is relaxed, it is possible to embed an ST [Standard Theory] or EST based parser in a serial computational model. We could do this by “precomputing” the effects of the transformational component and storing the results in the lexicon. Note that this in no way disconfirms transformational grammar either as a grammar or a central component of a parsing model; it says merely that the way in which a grammar may be embedded as a part of the model of language use is less than straightforward... Given the assumptions of transformational grammar ... one should note that the precomputed lexical templates associated with NP movement would necessarily be governed by purely structural principles, unlike the templates of extended lexical grammar. (BW, 1984: pp. 73-4).<sup>59</sup>

This yields a novel position on the psychological reality issue, which I'll call *precomputed realization* (PR).

(PR) The principles of a grammar, *G*, are realized in a parser in virtue of the fact that the lexical entries on which parser draws encode precomputed structures that are in accord with those principles.

Many current parsing models are based on a *lexicalist* grammar, of which LFG is an example. Such grammars attempt to capture *all* syntactic phenomena by enriching the information stored in the lexicon. Head-driven Phrase Structure Grammar (HPSG) and other grammars whose primary formal operation is *feature unification* all fall under this umbrella.<sup>60</sup>

---

<sup>59</sup> Berwick and Weinberg do not discuss the issue of whether this approach can be extended to unbounded movement—the kind exhibited by *wh*-questions and relative clauses. Head-driven Phrase Structure Grammar (HPSG) and other lexicalist formalisms seem to do precisely that.

<sup>60</sup> Constructing grammars whose primary formal operation is feature unification has become a dominant approach in syntactic theory. (For a helpful introduction, see Jurafsky and Martin, 2007: ch. 15.) Parsers that incorporate such grammars have been developed by Jurafsky (1993, 1996) and Vosse and Kempen (2000, 2009). While both models were constructed for the explicit purpose of accounting for a wide range of psycholinguistic data, Vosse and Kempen's model has the additional virtue of explaining the processing difficulties characteristic of aphasia. As they point out, this is a significant feature of their approach, given that most models aim to account only for the processing difficulties that statistically normal speakers encounter on garden-path sentences and other local ambiguities. A notable exception to this trend is Grodzinsky (2000), who argues that Broca's aphasics lack the operations that theorists in the GB tradition treat as “A-movement” (see Haegeman, 1994).

It's not exactly clear whether the relation between the parser and the grammar in these models is correctly described by PR. Transformational grammars were initially proposed in a non-lexicalist formalism. But, as BW point out, it can be modified for use in a parser that precomputes the effects of some of its rules and stores those effects in the lexicon. By contrast, LFG, HPSG and similar grammars were developed in a lexicalist formalism from the very beginning. No modification is necessary in order to implement them in a unification-based parsing model. Hence, their relation to the parser seems more direct than PR suggests.

Suppose a theorist were bent on showing that the grammar constructed by syntacticians is *precisely* the one that's used for language comprehension and production. Such a theorist would have to decide whether PR is strong enough to sustain his or her position, as applied to some specific grammar. My own view is that we should not undertake this task. Stabler (1984) argues convincingly that syntacticians place constraints on descriptive grammars that may not be applicable in psycholinguistics. The compact and simple grammar that's useful for the explanatory practices of a syntactician may not have the right formal properties for efficient use in a parsing system.<sup>61</sup>

Still, if *some* descriptively adequate grammar is psychologically real, then questions arise

---

<sup>61</sup> Steedman (2000) offers argument to the contrary: "[C]onsiderations of parsimony in the theory of language evolution and language development ... might also lead us to expect that, as a matter of fact, a close relation is likely to hold between the competence grammar and the structures dealt with by the psychological processor, and that it will in fact incorporate the competence grammar in a modular fashion. One reason that has been frequently invoked is that language development in children is extremely fast and gives the appearance of proceeding via the piecemeal addition, substitution, and modification of individual rules and categories of competence grammar. Any addition of, or change to, a rule of competence grammar will not in general correspond to a similarly modular change in a covering grammar. Instead, the entire ensemble of competence rules will typically have to be recompiled into a new covering grammar. Even if we assume that the transformation of one grammar into another is determined by a language-independent algorithm and can be computed each time at negligible cost, we have still sacrificed parsimony in the theory and increased the burden of explanation on the theory of evolution. In particular, it is quite unclear why the development of either of the principal components of the theory in isolation should confer any selective advantage. The competence grammar is by assumption unprocessable, and the covering grammar is by assumption uninterpretable. It looks as though they can only evolve as a unified system, together with the translation process. This is likely to be harder than to evolve a strictly competence-based system" (pp. 227-228).

about the degree to which it resembles any other descriptively adequate grammar. Only further inquiry can shed light on this issue. It may well turn out that a lexicalist theory provides the most elegant and successful approach to syntactic issues *and* that the HSPM will be shown to implement a unification-based algorithm. In that case, the grammar used for parsing and the grammar preferred by the syntactician would be largely identical. But other possibilities are certainly imaginable.

Whatever the case about its fit with psychological data, the Standard Theory of transformational grammar was taken up by computational linguists who attempted to use it in writing parsing algorithms.<sup>62</sup> According to Harkema (2001), these typically worked in 4 steps.

First, the parser will generate a set of possible surface trees for the sentence, using a so-called covering grammar. Next, the parser will reconstruct a set of possible base trees from the set of possible surface trees by applying the transformational rules of the grammar in reverse. After this step, the parser will discard those base trees that cannot actually be generated by the base component of the grammar. Finally, the parser will feed the remaining base trees into the transformational component to establish which base trees can in fact produce the sentence to be parsed. The set of base trees surviving this last test will be returned by the parser as the result of the parsing process. (Harkema, 2001; p. 6)

Such models ran into two kinds of problems. The first type is familiar from our earlier discussion: inefficiency. Early versions of transformational grammar were not formally restricted. It was clear from the outset that they were stronger than context-free, but later research (Peters and Ritchie, 1973) showed that they were actually capable of generating recursively enumerable languages.<sup>63</sup> The search space generated by reverse transformations was correspondingly vast.

---

<sup>62</sup> See, e.g., King (1983).

<sup>63</sup> Recursively enumerable languages are type-0 languages on the Chomsky hierarchy. A grammar that generates them is essentially unrestricted, hence stronger than even a context-sensitive grammar. As discussed in fn. 34, a grammar strong enough to generate natural languages is widely believed to be mildly context-sensitive.

If inefficiency were the only issue, theorists might be justified in holding out hope that clever search strategies and suitable heuristics would eventually solve the problem (§2.8). There was, however, a more principled difficulty. Some transformations in the Standard Theory were not reversible, on account of the fact that they deleted material from trees.

Applying these rules in reverse can be very problematic, because the tree to which the reverse transformational rule is to apply may not always contain enough information to reconstruct the content and position of the deleted material. Because of these computational problems, early transformational parsers were not considered very successful. (Harkema, 2001: pp. 6-7; cf. Fodor, Bever, and Garrett, 1974: pp. 315-316)

Hence, by late 1970s, many were convinced that transformational grammars were not only at odds with psycholinguistic data, but also unusable as computational models of language comprehension. Following Berwick and Weinberg (1984), I take it to be false that transformational grammars were disconfirmed by psycholinguistic data.<sup>64</sup> However, the failure of the computational parsing models based on the Standard Theory was quite real. Moreover, the Standard Theory itself suffered from a number of problems. Its unrestricted expressive power (fn. 63) and its proliferation of language- and construction-specific rules came to be seen as its most serious defects. This led to a number of revisions, thus giving rise to the *Extended Standard Theory* and the *Revised Extended Standard Theory*. At last, the appeal to construction-specific ordered transformations was abandoned altogether, in favor of what is now widely known as the *Principles and Parameters* (P&P) approach, within which the Government and Binding theory (GB) and the Minimalist grammars were formulated.<sup>65</sup> In the course of these

---

<sup>64</sup> A similar position is advocated by Phillips (1994, 1996, and forthcoming).

<sup>65</sup> The P&P framework gave rise to a novel and exciting approach in parsing theory, known as Principle-Based Parsing (§5). More recently, Chomsky (1995) initiated the Minimalist Program in linguistics—an effort to simplify descriptive grammars while retaining the insights of the P&P framework (§6).



developments, the distinction between D-structure and S-structure was maintained, though in a modified form. We will return to the central notions of GB in §5.

The failure of parsing models based directly on the Standard Theory gave rise to a host of proposals regarding how the HSPM assigns syntactic structure to linguistic input. Fodor, Bever and Garrett (1974) suggested that it does so by using a battery of construction-specific heuristic strategies, such as the “canonical sentoid strategy,” which dictates that N-V-N structure be interpreted as having an S-O-V form. This proposal has been roundly rejected. Pritchett (1992) provides a range of data that is at odds with this proposal. Moreover, from the standpoint of a psycholinguist, the heuristics-based approach is unsatisfactory, because it posits different strategies for what are essentially the same linguistic phenomena. This complicates both the parsing theory and the theory of language acquisition. Wanner and Maratsos (1978) express this point in the following important passage.

With the empirical failure of psychological models that incorporate transformational rules, however, many psychologists have abandoned not only transformational grammar but also the assumption about the psychological reality of the generalizations on which transformational grammar is based. The heuristic strategies schemes advanced by Bever, Fodor, and others make no claim to capture linguistically significant generalizations. For example, Fodor and Garrett (1967; also Fodor, Garrett, and Bever 1968) propose one heuristic strategy to handle object relative clauses and an entirely different heuristic strategy to handle subject relative clauses (Fodor, Bever, and Garrett 1974). Presumably, their system of strategies can only be expanded to handle the full range of relative clauses by adding more heuristic strategies. Such a system does not class relative clauses together in any way, nor does it represent their relation to declarative clauses. If a child acquired this type of grammatical representation of relative clauses, he would face the formidable task of having to learn each relative clause strategy independently of all the others, and none of his knowledge of declarative clauses would transfer to relative clauses.

Wanner and Maratsos take these considerations to motivate an approach based on what they call Augmented Transition Network grammars, for which a number of processing algorithms were developed in the 1970s. They write:

In our view, ... the process of comprehension places psychological constraints on the form in which grammatical generalizations can be maintained. Thus, whatever the listener knows about the structure of his language, he must represent it in a form that can be used in a comprehension process that extracts semantically relevant grammatical information in a single, beginning-to-end pass over the surface structure of the sentence. The virtue of ATN notation is that it gives us a precise way of representing grammatical generalizations in just this form, thus permitting us to make strong tests of our assumptions about comprehension. (Wanner and Maratsos, 1978: pp. 158-159)<sup>66</sup>

In the next section, we continue our historical tour of the parsing literature, examining the both the details of the ATN approach and the objections have been leveled against it.

## **§4. Augmented Transition Networks**

### *§4.1 Parsing with an ATN*

The general framework of Augmented Transition Networks (ATNs) can be conceived of as a set of notational devices for constructing recursive phrase structure grammars that differ in some ways from the context-free grammars described above. ATN grammars consist of a set of networks, each of which characterizes a kind of phrase. There is a network for full sentences, another for noun phrases, a third for verb phrases, and so on. Fig. 14 illustrates two very simple networks. (We will see more complex examples below.)

---

<sup>66</sup> It is important to recognize that the ATN framework is by no means the only conceivable solution to the problems that Wanner and Maratsos identify in Fodor, Bever and Garrett's heuristics-based approach. J. D. Fodor (1998a) writes: "Though the failure of the derivational theory of complexity caused a stir, before long a way was found to cope with derivational operations algorithmically and on-line, by folding them into the phrase structure assignment operations: Establish a transformational dependency between two sentence positions just in case it is needed in order to reconcile surface derivations from base phrase structure rules. (See the HOLD hypothesis of Wanner and Maratsos, 1978; the Superstrategy of Fodor, 1980.) As a result, parsing even with a transformational grammar could be seen as "left to right," systematic and incremental, effected by a precise program that faithfully applies the mental grammar. There was no need any longer for heuristics..." (p. 289). In this passage, Fodor identifies the HOLD hypothesis—an important feature of ATN parsers, to be discussed below—as one of two competing ideas for avoiding Fodor, Bever and Garrett's heuristics; the other is her own Superstrategy.

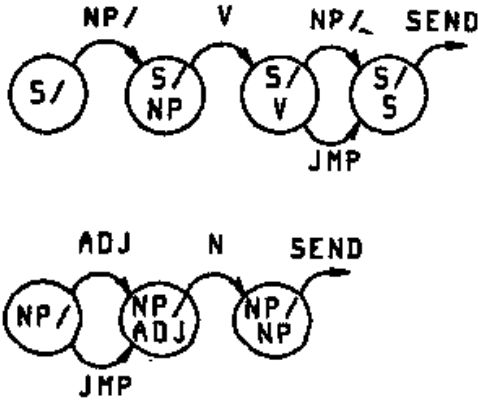


**Figure 14:** Top: A network for constructing simple sentences. Bottom: A network for constructing simple noun phrases. *Source:* Bates (1978)

The task of the network in fig. 14a, at the top of the diagram, is to construct a full sentence. It does so by first looking for a noun phrase and then for a verb. After the verb is found, it checks for another noun phrase. If it finds one, then it ends its analysis of the sentence. If it doesn't, then it ends the analysis anyway. The network in fig. 14b, at the bottom of the diagram, is designed to build noun phrases. It begins by checking for an adjective. If it doesn't find one, it looks for a noun. When a noun is found, it ends its analysis of the noun phrase. At this point, whatever process initially called for a noun phrase receives the go-ahead to resume.

The circles in the diagram represent states, linked to one another by arcs. The arcs are labeled with conditions that must be satisfied before a transition can be made. State **S/** in the network at the top of the diagram initiates the search for a full sentence. This state activates the **NP/** arc, which corresponds to an imperative to find a noun phrase; unless a noun phrase can be found, processing cannot continue. Building noun phrases is the job of the network in fig. 14b, so the activation of the **NP/** arc is an instruction for the activation of that network. By activating the **NP/** state, the sentence network has in effect predicted that the first phrase in the sentence

will be a noun phrase.<sup>67</sup> Once this prediction is made, the noun phrase network sets to work.

Let's imagine that the sentence it receives as input is (51).

(51) Accomplished artists command respect.

The **NP/** state has two arcs linking it with the **NP/ADJ** state. The arc at the top is labeled **ADJ** and the one at the bottom **JMP**. By convention, the action associated with the higher of two arcs will take priority. In this case, then, the **ADJ** arc is attempted first, initiating a search of the lexicon to determine whether the first word of the input is an adjective. The noun phrase network determines that 'Accomplished' is indeed an adjective. If it had not, the **JMP** arc would have been activated, which would allow the network to move to the next state without having satisfied any condition—a sort of “free pass.” Once the **ADJ** condition is satisfied, the network transitions into the **NP/ADJ** state, which has only one arc to the **NP/NP** state. This arc, labeled **N**, imposes the condition that a noun must be found before the process can continue. The network thus initiates another search of the lexicon, and determines that 'artists' is indeed a noun. This allows for a transition into the **NP/NP** state, from which the only arc is **SEND**. The **SEND** arc instructs the network to complete the noun phrase by storing in memory the structure it has built thus far—viz., [<sub>NP</sub> [<sub>ADJ</sub> Accomplished] [<sub>N</sub> artists]]. Control is then ceded back to the process that initiated the noun phrase network.

Returning, then, to the sentence network in fig. 14a, we see that the condition of the **NP/** arc has been satisfied, allowing for a transition from the **S/** state to the **S/NP** state. The only arc leaving the **S/NP** state imposes the condition that a verb must be found. A lexical search reveals that 'command' is a verb. The network transitions into the **S/V** state, from which there are two

---

<sup>67</sup> Recall from §2.2 that making such predictions prior to receiving any input is a characteristic of top-down parsers. The ATN system described here is therefore an instance of the top-down approach, though one that relies on backtracking rather than parallel processing in order to deal with local ambiguity resolution.

arcs, hence two possibilities. One is to find another NP, which involves reactivating the noun phrase network. Because the **NP/** arc is the higher of the two, this action is tried first.<sup>68</sup> The input is the word ‘respect’, so the lexical search concludes that the next item in the input is not an adjective. Hence, the noun phrase network uses the **JMP** arc to get to the **NP/ADJ** state. The **N** arc initiates another lexical search, this time successful—‘respect’ is a noun. The **SEND** arc then “pops” the processing back up to the sentence network, which is now in the **S/S** state—its final state. The **SEND** arc from the **S/S** state completes the sentence, outputting the parse in (52).

(52) [s [NP [ADJ Accomplished] [N artists]] [VP [V command] [N respect]]].

As noted above, Wanner and Maratsos (1978) argue that sophisticated ATN parsing systems are both psychologically plausible and superior to the heuristics-based strategies advanced by Fodor, Bever, and Garrett (1974). In support of these claims, they consider how an ATN would parse relative clauses—a process that involves finding positions occupied by empty categories, specifically *wh*-trace.<sup>69</sup> The following three examples, adapted from Wanner and Maratsos, demonstrate the role of *wh*-traces in relative clause constructions.

(53) ... the girl who *wh*-trace talked to the teacher about the problem ...

(54) ... the teacher whom the girl talked to *wh*-trace about the problem ...

(55) ... the problem that the girl talked to the teacher about *wh*-trace ...

As these examples show, the *wh*-trace can appear in subject, object, and adjunct positions. The antecedent of each *wh*-trace is always the NP that appears before a relative pronoun—i.e., before the complementizers ‘who’, ‘whom’, and ‘that’, in the examples above. The parser’s task is to locate the gap and co-index it with the correct NP.

---

<sup>68</sup> Notice that imposing an uniform order on the operations of the ATN can be a way of implementing parsing principles like Late Closure (Wanner, 1980). See the discussion of Frazier and Fodor (1980) for a decisive critique of this approach.

<sup>69</sup> See ch. III for a discussion of *wh*-traces and their effects on sentence processing. See also J. D. Fodor (1989).

In order to accomplish this, the ATN parser that Wanner and Maratsos develop introduces a feature that we did not discuss in connection with the simple ATN that we considered at the outset of this section. They call this new feature the HOLD list, which is a memory bank in which items are stored and retrieved at later points in the parsing process. The addition of this single feature allows the ATN to parse all three types of relative clauses exemplified in (53)-(55). Let us walk through a processing sequence, to illustrate this concept. Following Wanner and Maratsos, we will consider the object-extracted relative clause in (56). (A diagram of an ATN that can be used to parse this sentence appears in fig. 15 below.)

(56) The old train that the boy watched left the station.

As before, the process begins with a prediction to the effect that the first part of the sentence is a noun phrase. The noun phrase network is activated. It runs through the CAT ART, CAT ADJ, and CAT N arcs, finding first the article 'The', then the adjective 'old', and then the noun 'train'. The network then attempts to traverse the CAT REL PRO arc, thus checking for a relative pronoun. If there were no relative pronoun in the input, then the noun phrase would be completed and processing would "pop" back up to the sentence network. In the case we're considering, though, the network finds the relative pronoun 'that'. This automatically places the noun phrase that has so far been constructed (i.e., 'The old train') onto the HOLD list. The system's working assumption is that this noun phrase will be reactivated when the position of the *wh*-trace in the relative clause is discovered. The network now enters state NP<sub>3</sub>, from which the

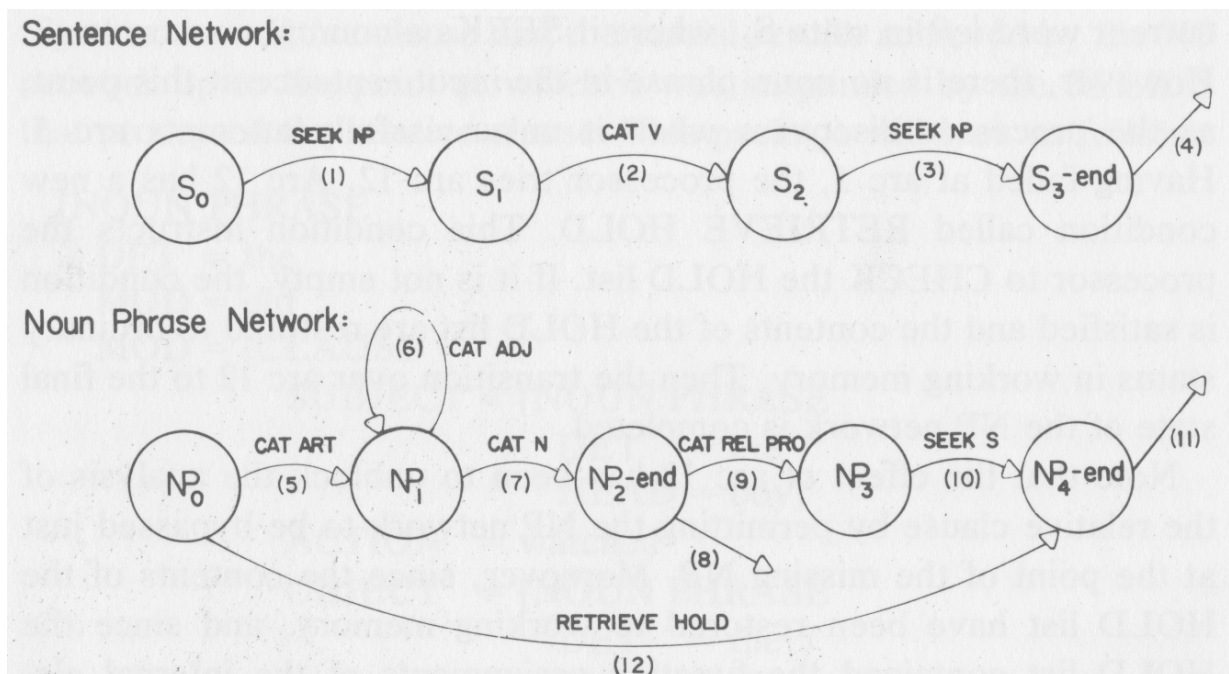


**Figure 15:** The labeling conventions in fig. 14 are slightly different from those employed here. The arcs carry SEEK and CAT labels. The label SEEK NP is equivalent to the NP/ label in fig. 14. It indicates that the noun phrase network should be activated before the transition to the next state can be made. The CAT V label is equivalent to the V label in fig. 14. It indicates that the lexicon should be consulted to determine whether the next word in the input is a verb. The arc labels ART and REL PRO have been added. These call for a lexical decision regarding whether the next word in the input is an article ('a', 'the', etc.) and a relative pronoun ('who', 'which', etc.), respectively. Since adjectives can be repeated indefinitely ("the *awesome shiny new expensive* car"), the ADJ arc can lead back into the state from which it originated, so as to allow for additional activations of that very same arc. Finally, the arcs have been numbered. The numbers do *not* correspond to the order in which the actions are carried out. They serve as an expository device in Wanner and Maratsos's discussion of the thematic role assignments that are carried out in the course of the parsing process.

only arc is a directive to activate the sentence network. The logic behind this move is that relative clauses are basically just declarative sentences with a missing noun phrase. For instance, the relative clause bracketed in (57) is just the declarative sentence (58), except lacking a subject at the position occupied by the *wh*-trace.

- (57) ... the girl who [ *wh*-trace talked to the teacher about the problem ] ...  
 (58) The girl talked to the teacher about the problem.

Hence, in the course of constructing the relative clause (construed by the network as a declarative sentence), the ATN will find that some noun phrase is needed but not present in the input. This will prompt the ATN to check the HOLD list, to see if there are any noun phrases waiting to be incorporated. Here is how this happens.

The sentence network, now activated for the purpose of constructing the relative clause, begins by traversing the SEEK NP arc—i.e., seeking a noun phrase.<sup>70</sup> This re-activates the noun phrase network. The words ‘the’ and ‘boy’ are identified as a noun phrase and processing pops back up to the sentence network. The SEEK V arc is successfully traversed when the word ‘watched’ is identified as a verb. At this point, the sentence network attempts to locate the next NP—i.e., the object of ‘watched’. The lexical search indicates that that ‘left’ is not a noun phrase. So the network checks the HOLD list and finds a noun phrase there. This noun phrase satisfies the condition imposed by the SEEK NP arc between states  $S_2$  and  $S_3$ . The relative clause, treated as a declarative sentence, has been processed. The network goes on to work on the verb phrase ‘left the station’ in the familiar way, which we need not rehearse here.

Wanner and Maratsos note that the ATN just described, if taken to be a psychologically plausible model of the HSPM, makes predictions about the memory costs involved in human sentence processing. In particular, when a relative pronoun is found, the NP that has been constructed (‘the old train’, in the case just discussed) must be placed on the HOLD list, thus exacting a toll on short-term memory resources. They go on to present the results of two experiments that they conducted to show that human listeners do indeed exhibit processing effects indicative of increased memory load after the relative pronoun. The details of the

---

<sup>70</sup> Notice that the ATN described here, unlike the simple one discussed at the outset of this section, makes it possible for the sentence network to be activated in the course of an operation initiated by that very network. As fig. 15 illustrates, the sentence network can activate the noun phrase network, which can subsequently activate the sentence network. This allows for the embedding of sentences within sentences. This and similar formal devices render the ATN recursive, with a power matching (and perhaps superseding) that of a recursive context-free grammar.



experiments need not concern us here. Suffice it to say that the results bode well for the psychological plausibility of the ATN model.

Wanner and Maratsos go on to compare their model with the heuristics-based approach advanced by Fodor, Bever, and Garrett (1974). On that approach, three separate heuristics would need to be consulted and applied in the course of processing the three types of relative clause exemplified in (53)-(55).

In an ATN notation such a system of independent strategies would appear as a large set of distinct arc sequences following the head noun arc in the noun phrase network. But recall that this proliferation of arc sequences is exactly what is avoided in the HOLD model, and avoiding it is possible only because the HOLD model preserves a generalization about the class of all relative clauses—namely, that all such clauses appear to be declarative clauses with a phrase missing. Our evidence for the HOLD hypothesis, then, is evidence that the language user's comprehension system does incorporate at least this one generalization about the relation between clause types. As everyone knows, this is precisely the kind of generalization that can be captured by transformations. Thus, our evidence indicates that the comprehension system may employ the kinds of generalizations captured by transformational grammar, although not in transformational form. (p. 159)

#### *§4.2 The bearing of ATNs on the psychological reality issue*

The networks just described can be thought of as simple grammars. Although the presentation above casts each of them as a parsing model, there is no obligation to think of them in that way. An ATN may be seen simply as an alternative way of describing the surface syntax of a language—i.e., for stating the optional and obligatory relations (both linear and hierarchical) between syntactic categories. Instead of stating syntactic relations in a declarative format, as context-free grammars do, an ATN grammar encodes these relations in *procedural* instructions—rules for what to *do* in a certain context. Still, although an ATN can be seen as nothing more than a grammar, it is nevertheless quite natural to interpret it as a processing model. The only declarative data structures that need to be appended in order to make it operate

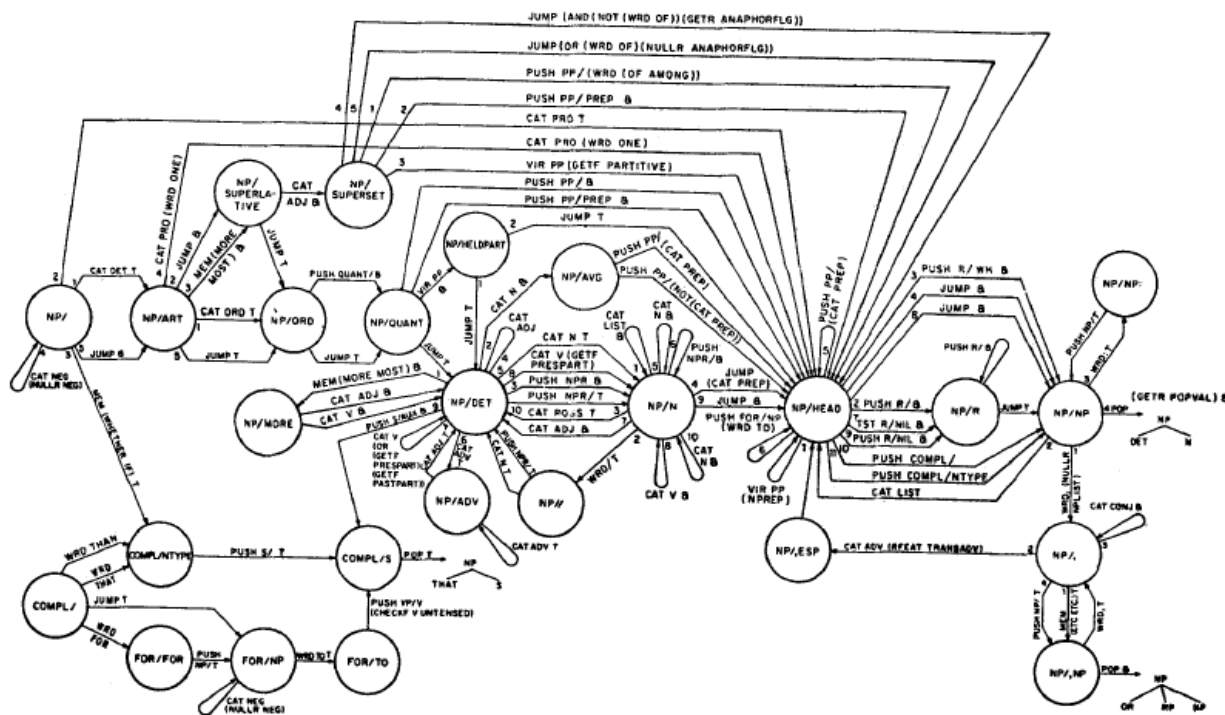
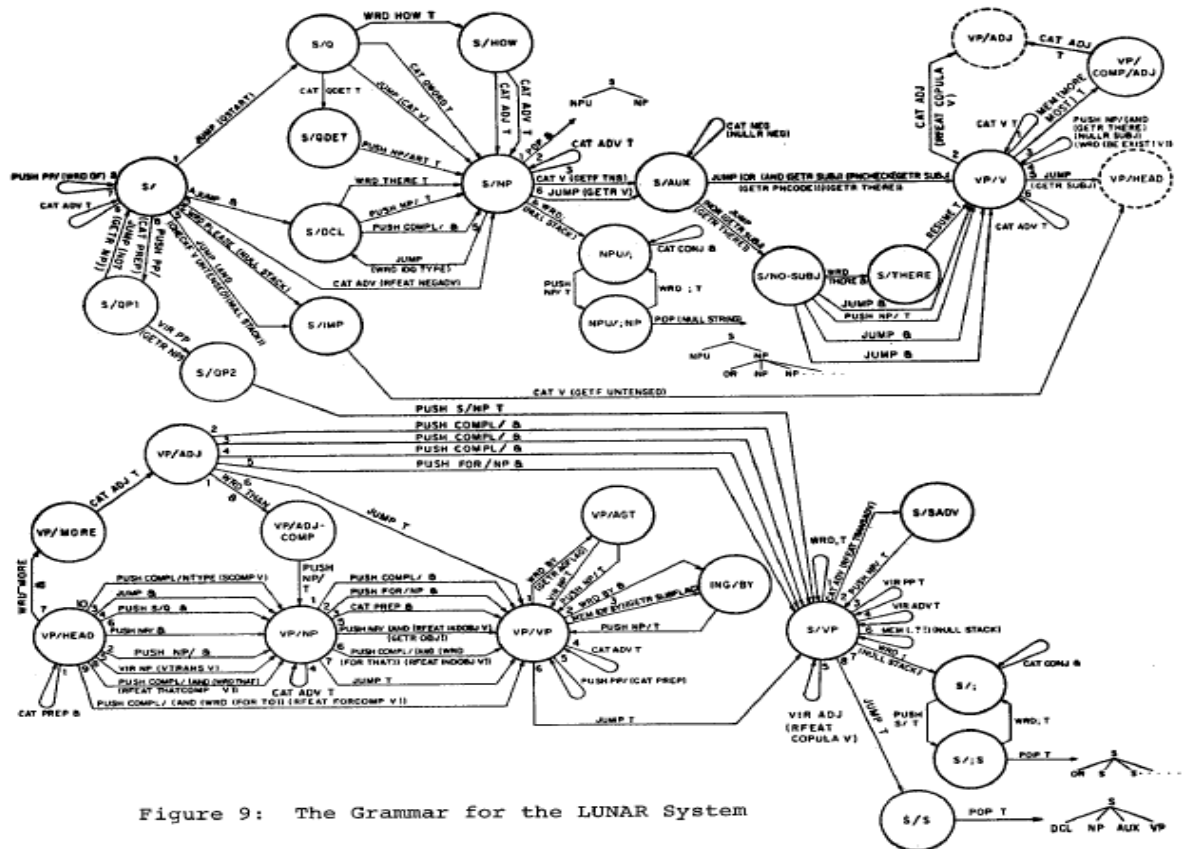
as a parser are: (1) a lexicon and (2) a memory bank for storing the partial structures that are assembled at intermediate points in the process—i.e., mental phrase markers.

This duality of the ATN formalism has significant implications for our view of what Devitt (2006) calls positions **(i)** and **(iii)** on the psychological reality issue. According to position **(i)**, “the structure rules are represented and applied in processing” (p. 207). According to position **(iii)**, “the syntactic rules of a language are embodied somehow in competent speakers without being represented, and [they] govern processing” (p. 117). The ATN formalism provides a vivid example of the way in which the rules of a grammar (i.e., the structure-rules) can *govern* processing, in the sense that comports with position **(iii)**.

The very simple ATNs discussed above are perhaps not the best aid to our thinking about this issue. Consider, then, a more robust ATN grammar—the one used in the LUNAR parser (Woods, 1973), depicted in fig. 16 below.<sup>71</sup> Imagine a hardwired electrical circuit whose states correspond, one-by-one, to the states of this grammar (denoted by the circles in the diagrams in fig. 16). The circuit can be embedded in an object that resembles an ordinary pocket calculator, except that the rubber buttons each stand for a word, rather than a number, and the output on the screen is a phrase marker rather than a numerical solution to a calculation query. When linguistic input is entered (i.e., when the rubber buttons are pressed), the flow of current through the circuit constitutes state-to-state transitions—precisely those encoded in the arcs of the ATN grammar. The result is a parse of the input sentence.

---

<sup>71</sup> I owe these beautiful diagrams to Bates (1978), whose comprehensive discussion of ATNs is second to none. Bates’ treatment of the various ways in which ATNs can be compiled and implemented demonstrates that the hardwired circuit described in the main text is by no means the only possible implementation of ATN parsing routines. Indeed, to my knowledge, such a circuit has never been constructed. ATN grammars are typically encoded as data, just like the context-free grammars that we considered earlier. Despite its being merely hypothetical, the hardwired syntax-processing circuit is useful for the purposes of discussing positions (i) and (iii).



The ATN circuit that we are now contemplating can be said to *embody* an ATN grammar. But, in connection with the Representational Thesis (RT), we should ask whether, in addition, it *represents* the grammar. The answer is that it does not. For, the grammar is not stored in a declarative format, as a separate data structure awaiting to be read by the system at the opportune time. (This marks a difference between the ATN circuit and the machines that are traditionally used to implement the Earley and CYK algorithms that we discussed in §2.) Nevertheless, we can see some reasons for thinking that the intimate relationship between the ATN circuit and the ATN grammar warrants the claim that the rules of that grammar are, *in this weak sense*, psychologically real. Let us lay out considerations that underlie this contention.

First, the transitions between circuit states—i.e., the actions of the parser—are “in accordance with” the ATN grammar, not merely in the weak sense that they produce the phrase markers that the grammar assigns to strings of the language, but in a much stronger sense: The device could not output those phrase markers if its internal organization did not *transparently implement* the grammar—i.e., if the states of the circuit did not correspond, one by one, to the nodes of the grammar. An alteration in the circuit would be naturally interpreted as an alteration in the grammar that the circuit embodies. If the altered grammar were not *true of* the expressions of English, then the parser *wouldn't work*. It would either halt when presented with perfectly well-formed inputs—producing a SYNTAX ERROR message familiar from the old Texas Instruments graphing calculators—or it would process the input in a way that would lead to systematic misconstruals at the level of semantics.

For this reason, the circuit embodies the grammar of English in a way that makes it both natural and useful to interpret its internal organization as *storing information* about what syntactic structures are allowed or obligatory in English. (Though this sense of ‘storing

information’ must, of course, be distinguished from the sense in which information storage requires representation.) This, in turn, allows us to offer algorithmic descriptions and explanations of the activities of the circuit—pitched at David Marr’s second level (Marr, 1980). Moreover, we can appeal to the grammar in issuing predictions about the observable effects of those activities, *in the absence of detailed knowledge of their physical implementation*. This powerful mode of description, prediction, and explanation typically does not require that we appeal solely to representations.<sup>72</sup>

#### §4.3 Problems for the ATN framework

Given the success of ATN models, both as descriptions of surface syntax and as parsing systems, one might wonder why the approach fell out of favor in subsequent decades. There are several reasons, which we briefly review here. The first couple are rather weak. I leave the more weighty objections for last.

First, ATN parsers are “brittle,” in the sense that they cannot parse ungrammatical input. In contrast to human listeners, who can understand sentences like (59) with relative ease, an unadorned ATN simply grinds to a halt when confronted with such input.

(59) \*The boys is running.

This is by no means fatal. The ATN architecture can be modified in various ways to deal with the problem. For instance, upon encountering (59), a more sophisticated ATN parser can treat its failure in parsing the sentence as a signal to back up and revise its lexical decisions, treating ‘boys’ as ‘boy’ or ‘is’ as ‘are’, depending on the details of the design. Other modifications are

---

<sup>72</sup> Horgan and Tienson (1996) touch on this issue, using the term “explicit representation” for cases in which rules are *not* “hardwired” in a system, but declaratively represented in a separate data structure. Presumably, they would consider hardwiring as a kind of *implicit* representation. But whatever label we ultimately decide to use for such cases, the distinction between declarative representation and procedural embodiment remains an important one.

imaginable. It has been remarked that these strategies are *ad hoc*.<sup>73</sup> But it's conceivable that experimental evidence would reveal that human hearers employ them in dealing with ungrammaticality. Concerning the relevant psychological experiments, Berwick (1991b) writes:

[S]imple agreement violations seem to cause more problems than they ought to. Given this, Crain and Fodor (1985) argue that the real problem has to do with confusability: subjects are looking for a near-miss grammatical examples, and that's possible with agreement violations but not possible with violations such as *Who did you see Picasso's picture of*. Therefore, in the agreement violation case, processing will be longer as the subject searches for a possible alternative; in the second, constrain violating example, the subject gives up rapidly. Freedman and Forster (1985) attempt to lay this matter to rest by trying to show that correctability does not account for all such results, but from my point of view the entire matter still seems unclear. (p. 182)

The second problem for ATNs has to do with familiar efficiency issues. ATN parsers are top-down systems, which are known to be less efficient than their bottom-up counterparts. Marcus (1980: ch. 2) discusses some of the problems that plague ATNs. He notes that ATN parsers are *nondeterministic*, in the sense that they rely on a backtracking strategy in the face of an error in mid-sentence ambiguity resolution (§2.2). Given the fact that the parser may have to backtrack arbitrarily far, checking all of the decisions it made at every choice-point, this strategy can quickly become inefficient.

As we noted above, other nondeterministic systems invoke parallel processing to solve this problem, essentially relying on expanded memory resources to store multiple parses whenever an ambiguity presents itself before the end of a sentence. Again, though, the expansion of memory resources can quickly spiral out of proportion, and in parallel systems that rank candidate parses, computational resources are needed to revise the rankings on the fly.<sup>74</sup> Some theorists suggested using the HOLD feature to delay parsing decisions until the system

---

<sup>73</sup> See, e.g., Berwick (1991b) for discussion.

<sup>74</sup> For a discussion and defense of ranked parallel models, see Gibson (1991).

reaches a point in the input where an ambiguity can be resolved without the risk of error—a “minimal commitment” model. Unfortunately, this is psychologically implausible; there is ample experimental evidence that the HSPM does not delay its decisions in this fashion (ch. III).

Marcus (1980) invoked these considerations in rejecting the nondeterministic models that were available at the time, including ATN parsers, and in motivating his own deterministic parser.<sup>75</sup> However, the problems are not as severe as Marcus’s discussion would lead one to believe. As discussed in §2, a variety of methods have been developed to increase the efficiency of top-down parsers. The use of grammar-based and resource-based processing heuristics, as well as probabilistic techniques, may well be able to resolve these difficulties. Indeed, in an article championing ATN models, Wanner (1980) makes precisely these suggestions.<sup>76</sup>

Let us now turn to the strongest arguments against the ATN framework. Fodor and Frazier (1980) argue that the empirically well-confirmed processing principle, Right Association (a.k.a Late Closure.), is not realizable in an ATN architecture in a principled way.<sup>77</sup> More damning, Fodor and Frazier point out that ATNs cannot implement Minimal Attachment (MA) *at all*. In a moment, we will see how their competing model, the “Sausage Machine,” implements these principles. Presently, let us examine their argument for the claim regarding MA.

Wanner (1980) suggested that MA can be implemented in an ATN by a scheduling routine that enforces the following procedure: Traverse all CAT arcs before attempting to

---

<sup>75</sup> For a critique of both the original Marcus parser and its D-theoretic offspring, see Pritchett, 1992: pp. 44-51.

<sup>76</sup> The majority of Wanner’s discussion is devoted to examining how Minimal Attachment and Late Closure (which he calls Right Association) can be incorporated into an ATN system. The only mention of probabilistic strategies is at the very end of the paper (p. 224, fn. 8).

<sup>77</sup> Right Association (RA) is a principle originally proposed by Kimball (1973). According to this principle, “terminal symbols optimally associate with the lowest nonterminal node” (Kimball, 1973: p. 24). A version of this principle survives today under the label ‘Late Closure’ (Frazier 1978; Fodor and Frazier, 1978, 1980; Fodor, 1998a). Kimball’s commitments about the architecture of the parser differ slightly from Fodor and Frazier’s, making it hard to say that RA and LC are identical claims clothed in different words. But the claims are certainly very similar.

traverse a SEEK arc. However, as Fodor and Frazier (1980) point out, the CAT-before-SEEK strategy is *not* equivalent to MA. The clinching example is the sentence fragment in (60), which can be construed as the beginning of multiply-conjoined NP, as in (61), or as a reduced relative clause, as in (62).

(60) The man the girl...

(61) The man the girl and the dog played in the park.

(62) The man the girl dislikes was walking in the park.

Evidently, the HSPM prefers the conjunction analysis (61) over the relative-clause analysis (62).<sup>78</sup> But, while MA predicts this preference, the CAT-before-SEEK strategy does not, as the reader can verify by scrutinizing the ATN grammar in Wanner (1980).

Fodor and Frazier argue that these problems for the ATN models stem from their not *explicitly* representing the rules of a grammar in a separate data structure. By contrast, Fodor and Frazier's model, the Sausage Machine, stores the grammar in what they call a "rule library."<sup>79</sup>

We have argued (in connection with Minimal Attachment) that the grammatical rules for the language should be stored separately from the specification of the computational operations involved in the application of those rules. They will reside in a special 'rule library', and will be accessed by the executive component of the parser as needed. (p. 451)

This is a significant difference between the ATN and the Sausage Machine. Fodor and Frazier point out that, while both models can implement the principle of Late Closure (LC), only their model can implement MA and give a *principled* explanation of the fact that the parsing preferences of the HSPM are governed by LC. The explanation is principled because it follows

---

<sup>78</sup> Frazier and Fodor do not discuss the lack of commas in (60)-(61). But this is not a problem for their claim that the HSPM prefers the reading in (61) over the one in (62). Indeed, it *bolsters* their claim. For, lacking a comma, the fragment in (60) is more likely to be interpreted as introducing a reduced relative clause than a conjunctive list. The fact that the HSPM nevertheless prefers the conjunctive analysis is therefore even *more* noteworthy.

<sup>79</sup> See Frazier and Fodor (1978) for a detailed description and motivation of the Sausage Machine. McRoy and Hirst (1990) extend the model in interesting ways. Criticisms can be found in Pritchett (1992: pp. 30-40).



directly from the architecture of the Sausage Machine—including, crucially, the existence of a separate rule library—whereas the ATN model would have to append the relevant scheduling routines in an *ad hoc* fashion. Let us briefly examine the Sausage Machine, to see how it incorporates MA and RA/LC.

The Sausage Machine (henceforth, SM) has two components—the preliminary phrase packager (PPP) and the Sentence Structure Supervisor (SSS). The PPP examines the first five or six words of the input and builds a partial mental phrase marker for those words. This material is then erased from its memory and passed on to the SSS, where it is temporarily stored, awaiting integration with further input. When the PPP passes along the next partial phrase marker, the SSS puts it together with the prior one. If necessary, the process repeats several times, eventually yielding a complete mental phrase marker for the entire sentence.

The motivations for this dual-component structure had to do with memory limitations. As is well known, only a small number of “items” can be held in short-term working memory. But this limitation can be transcended, to some extent, by “chunking” several simple representations into a complex structure and treating the result as a single “item”—the equivalent of hyphenation in the formal work of, say, Goodman and Quine. The cost of this strategy is that the system becomes blind to the internal structure of the complex representation, and must expend computational resources to “look within” that representation if it becomes necessary to do so. It follows that if the SSS is unable to integrate the outputs of the PPP, then those representations have to be “unpacked,” so to speak, before structural revision can take place. This gives rise to the processing principle known as Revision as a Last Resort (RALR), thus explaining why many kinds of reanalysis are computationally costly.

In the SM, the input triggers a sort of competition between whatever rules might be applicable. The rules are “accessed in parallel and selected in terms of the outcome of a ‘race’—the first rule or combination of rules that successfully relates the current lexical item to the phrase marker dominates subsequent processing” (Fodor and Frazier, 1980: p. 434). A particular analysis of the input—i.e., a specific MPM—is generated by whatever set of rules wins the race. And since the application of each rule takes time and computational resources, the winning set of rules is always going to be the *minimal* set required by the input. For example, suppose that the model is working with the following simple grammar:

- (63)  $S \rightarrow NP VP$
- (64)  $VP \rightarrow V NP (PP)$
- (65)  $NP \rightarrow Det N$
- (66)  $NP \rightarrow NP PP$

In this case, the input in (67) would be parsed as (68), not as (69).

- (67) David read the letter to Jennifer.
- (68) David [<sub>VP</sub> read [<sub>NP</sub> the letter] [<sub>VP</sub> to Jennifer]].
- (69) David [<sub>VP</sub> read [<sub>NP</sub> [<sub>NP</sub> the letter [<sub>PP</sub> to Jennifer]]]].

The reason is that modifying a verb requires applying only one rule—viz., (64)—whereas modifying a noun requires applying two—viz., (65) and (66). Thus, SM predicts this particular parsing preference. Notice also that the structure in (68) requires attaching the PP directly to the VP, whereas the structure in (69) requires attaching it to the NP, thus necessitating the creation of an extra NP node. The creation of extra nodes violates MA. Thus, as Fodor and Frazier point out, the “race” aspect of the SM not only predicts an isolated parsing preference, but also explains why, quite generally, the preferred attachments conform to MA.<sup>80</sup>

---

<sup>80</sup> Another strength of the model is that it explains why MA sometimes does *not* apply. For instance, consider the sentence *Joe read the newspaper article, the card, the telegram, and the letter to Mary*. MA dictates that ‘to Mary’ should be attached to the verb, analogously to (68) above. But, in this case, most readers prefer that this preposition be attached to ‘the letter’. Fodor and Frazier’s explanation is this: By the time the HSPM encounters the preposition ‘to Mary’, the verb ‘read’ and the structure surrounding will have already been passed from the PPP to

From this, Fodor and Frazier conclude the SM is superior to the competing ATN models, and that the rules of a grammar must be *explicitly represented* by the HSPM.

[W]hen making its subsequent decisions, the executive unit of the parser refers to the geometric arrangement of nodes in the partial phrase marker that it has already constructed. It then seems unavoidable that *the well-formedness conditions on phrase markers are stored independently of the executive unit, and are accessed by it as needed*. That is, the range of syntactically legitimate attachments at each point in a sentence must be determined by *a survey of the syntactic rules for the language*, rather than being incorporated into a fixed ranking of the moves the parser should make at that particular point... (Frazier and Fodor, 1978: 322n, emphasis added).

Fodor and Frazier's argument yields support for position **(ii)** on the psychological reality issue and hence for RT. It constitutes the beginning of an answer to the kinds of concerns that Devitt (2006) voices in the following passage.

Frazier and Fodor take their theory to support the view that a hearer mentally represents the well-formedness conditions of her language ... So they are proposing a version of position (ii): rules of the language are represented and used as data. [fn. 35: Note also Fodor's later commitment to the processor being "transparent" in the following sense: it "makes use of information about linguistic structure in the form in which the mental ("competence") grammar provides it, so that statements from the grammar do not have to be modified, translated, or 'pre-compiled' before they can be applied to sentences" (1989: 177).] But it is unclear why the well-formedness conditions have to be represented and surveyed. Why could not the rules governing the "subsequent decisions" be embodied but unrepresented rules that respect (in my technical sense) the well-formedness conditions? It is hard to see what "payoff" there is in having the conditions represented. Pylyshyn's Razor counts against our supposing that they are. (*Ignorance of Language*, p. 239)

If Fodor and Frazier are right, then the explanatory payoff that we get from positing represented rules is this: The "race" aspect of the SM model accounts for the fact the attachment preferences of the HSPM conform to MA. If the rules were embodied but *unrepresented*, as they are in the ATN, then we would not be able to explain this important finding. Its reliance on the explicit

---

the SSS. Therefore, when the preposition is encountered, it the PPP will have no choice but to attach it to the NP 'the letter'.

representation of rules gives the SM model an advantage over the procedural approach of the competing ATN model.

This argument is, however, by no means conclusive. Devitt can challenge it by pointing out that the hardwired ATN circuit (§4.2) is not the only competitor to the SM model. Indeed, he holds out hope that contemporary rivals will eventually prove themselves superior. Having reviewed some of the sentence processing literature, he writes:

Where does this brief discussion leave our preliminary assessments? Most importantly, it does not undermine them. And I think that it gives further support to our negative view of RT. Clearly constraint-based connectionist approaches support our assessment that the future will be fairly brute-causal, hence support our fourth tentative proposal [that the speedy automatic language processes arising wholly, or at least partly, from linguistic competence are fairly brute-causal associationist processes that do not operate on metalinguistic representations of the syntactic and semantic properties of linguistic expressions]. These approaches look promising but it is too early in the study of language comprehension to be confident that they are right. (*Ignorance of Language*, p. 240)

In chs. III and IV, we reviewed the difficulties facing the brute-causal associationist proposal. In §6, we will shed further doubt on the constraint-based models that Devitt mentions in this passage (see also fn. 9). Indeed, I am aware of no procedural model that can do what the ATN model failed to do—i.e., to explain the HSPM's conformity to Minimal Attachment. Still, it is possible that I have overlooked the existence of such a model, or that its unveiling is just around the corner. Moreover, there are well-known objections to the SM model (Pritchett, 1992: pp. 30-40), which must be addressed before we can confidently declare Fodor and Frazier's argument for RT a success.

One such objection has to do with the assumptions that the SM model makes with regard to natural-language syntax. Like the models that we examined in §2, the SM makes use of the

now-outdated context-free phrase structure rules.<sup>81</sup> In §§5-6, we will examine more recent parsing models, which employ the now-familiar Earley and CYK algorithms (rendered in a Parsing as Deduction framework) to operate on more descriptively adequate grammars—in particular, those emerging from the Principles and Parameters tradition.

## §5. Principles and Parameters in Syntax and Parsing

### §5.1 *The First Grammar in the P&P Framework: Government and Binding Theory*

In §3, we considered the Standard Theory of transformational grammar. This theory posited transformational rules, over and above context-free phrase structure rules. The transformational rules operate on the syntactic structures of strings, which are generated by the “base” rules, perhaps of a context-free grammar (§1.4). Transformational rules map these into the syntactic structures of *other* such strings, transforming, e.g., an active into a passive. (See examples (39)-(45) in §3.1.) These rules were construction-specific, in the sense that each separate linguistic construction—passive, raising, question, etc.—was associated with a distinct transformational rule.<sup>82</sup> In addition, many such rules were language-specific, on account of the fact that the structures on which they operate, as well as those they generate, tend to differ from one language to the next.

The proliferation of transformational rules in the Standard Theory came to be seen as a serious problem. From the perspective of the syntactician, a grammar that posits such rules ranks low in elegance and simplicity. From the perspective of an acquisition theorist, a

---

<sup>81</sup> See McRoy and Hirst (1990) for a new version of the SM that uses a more up-to-date grammar. Even this grammar, however, has been superseded in recent decades.

<sup>82</sup> I use the term ‘construction’ here in the way that some linguists use the term ‘linguistic phenomenon’. My usage is derived from Berwick (1991b). Whatever the case about the labels, I do not mean to suggest that any particular phrase or clause can exhibit at most one linguistic phenomenon or construction.

formalism that requires very different sets of rules for distinct languages is unsatisfactory, for it makes it harder to explain how a child could ascertain the rules of the local language in a relatively short amount of time. Finally, as we saw in §3.2, the “backwards-transformation” models of comprehension based on the Standard Theory were generally unsuccessful.

Subsequent work, eventuating in what is now known as the *Principles and Parameters* approach (P&P, henceforth), heralded a striking revision of transformational grammar throughout the 1980s. The first grammar to emerge from that approach was the Government and Binding theory (Chomsky, 1981; Haegeman, 1994). Though this theory (henceforth ‘GB’) retained the two-level framework, distinguishing between D-structure and S-structure, it did away with the myriad transformational rules of the Standard Theory. Only a single rule remained: *Move- $\alpha$* . This rule says, quite literally, that any constituent appearing at D-structure can be moved anywhere in the phrase-structure tree on the way up to S-structure. Left unconstrained, this rule would generate a great many S-structure descriptions that fail to correspond to anything one finds in natural language. The substantive task, then, is to eliminate these unwanted descriptions. In the P&P model, this is accomplished by invoking a set of syntactic *principles*, which operate in the manner of filters, ruling out ill-formed structures and retaining only the well-formed ones, as shown in fig. 17.

The distinction between Standard Theory *rules* and GB *principles* is illustrated in fig. 18. The principles are few in number and compact in formulation (hence naturally seen as axioms in a deductive system). Furthermore, they are parameterized, typically in a binary fashion.

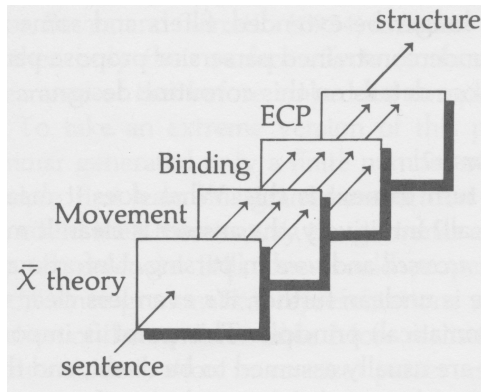


**Figure 17:** Filters and generators.  
*Source:* Berwick (1991b)

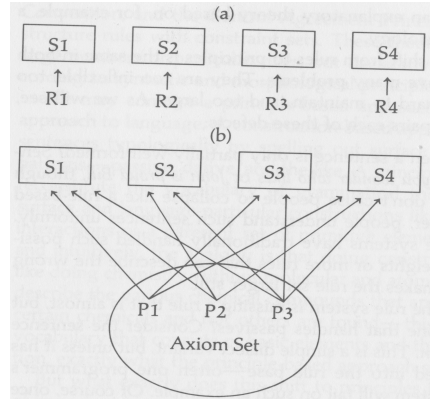


**Figure 18:** (a) Rules vs. (b) Principles  
*Source:* Berwick (1991b)

Each distinct way of setting the parameters allows the grammar to generate the syntactic structures characteristic of one particular language. The only differences between distinct languages, then, are to be found in the lexicon and the parameter settings. Quite apart from its sheer elegance, the resulting formalism provides guidance to the acquisition theorist, who can now view acquisition as a process of *setting parameters* in response to linguistic exemplars.<sup>83</sup>

The conception of D-structure also underwent radical revisions in the transition from the Standard Theory to GB. Leading up to rise of GB, it gradually came to be understood that the syntactic descriptions pitched at D-structure need not describe strings that are actually spoken, written, or signed. That is, the phrase structure rules that form the base of the grammar are free to generate D-structure descriptions that fail to correspond to any string we might find in a corpus.<sup>84</sup> The phrase marker in (70) illustrates this new kind of D-structure description.<sup>85</sup>

<sup>83</sup> This, of course, requires the parameters to be either innate or acquired in some way by the child. Moreover, parameters cannot be set until the child has acquired a relatively rich lexicon, which specifies (at the very least) each word's syntactic category.

<sup>84</sup> We can ignore, for present purposes, those cases where people have made "mistakes" or "performance errors." Note, moreover, that this new conception of D-structure had been seriously entertained decades prior to GB.

<sup>85</sup> This example is adapted from Haegeman (1994), p. 306.

(70) [S [S' was [VP believed [NP the story] by the villagers]]] D-structure

Plainly, (70) is not a sentence of English. Rather, it encodes an *aspect* of the syntactic structure of a sentence—pitched at just one level of description. The full sentence is actually a 4-tuple: {Phonological Form (PF), D-structure (DS), S-structure (SS), and Logical Form (LF)}. As we saw in ch. III, D-structure descriptions reflect, among other things, the argument structures of verbs. For instance, the D-structure in (70) reflects the fact that the object of the verb ‘believes’ is the noun phrase ‘the story’. Hence, D-structure descriptions do not tell us what surface forms we will find in a corpus. Rather, they tell us about the argument structure of the verbs that enter into such sentences.

Once they are generated by the base rules of the grammar, D-structure phrase markers like (70) are mapped, via the transformational rule *Move- $\alpha$* , to S-structure descriptions, like (71).

(71) [S [NP The story]<sub>i</sub> [S' was [VP believed [NP NP-trace<sub>i</sub>] by the villagers]]]. S-structure  
As (71) illustrates, when *Move- $\alpha$*  takes the noun phrase from its canonical position in D-structure to its new position in S-structure, an NP-trace—i.e., an empty category—is left behind. This trace is co-indexed with the noun phrase, thus providing information about what semantic role the noun phrase plays in the sentence.

To summarize, the story is now this: A single sentence has *two distinct aspects of syntactic structure*—D-structure and S-structure. The phrase structure rules that generate D-structures are designed to describe an *aspect* of sentences like (72). The D-structure in (73) is then mapped onto *another* aspect of the syntactic structure of (72), viz., its S-structure (74).

(72) The story was believed by the villagers. English sentence

(73) [S [S' was [VP believed [NP the story] by the villagers]]] The D-structure of (72)

(74) [S [NP The story]<sub>i</sub> [S' was [VP believed [NP NP-trace<sub>i</sub>] by the villagers]]]. S-structure of (72)

The transformation rule *Move- $\alpha$*  defines a relation between these two aspects of a single



sentence. This stands in contrast to the transformational rules that comprised the earliest versions of the Standard Theory, which described relations between two distinct sentences, e.g., an active and a passive.

In older versions of transformational grammar, ordinary context-free phrase structure rules generated the D-structure representations. In the P&P framework, phrase structure is assigned by set of principles, collectively known as X-bar theory. These principles can be rendered in the familiar context-free format, as the rules below illustrate, though there is no pressing need to do so. (Parentheses denote optional elements.)

|                            |                                   |                                         |                                |
|----------------------------|-----------------------------------|-----------------------------------------|--------------------------------|
| $NP \rightarrow (D) N'$    | $VP \rightarrow V'$               | $AdjP \rightarrow Adj'$                 | $PP \rightarrow P'$            |
| $N' \rightarrow (AdjP) N'$ | $V' \rightarrow (AdvP) V'$        | $Adj' \rightarrow (\{AdvP/AdjP\}) Adj'$ | $P' \rightarrow (AdvP)$        |
| $N' \rightarrow N' (PP)$   | $V' \rightarrow V' (\{AdvP/PP\})$ | $Adj' \rightarrow Adj' (PP)$            | $P' \rightarrow P' (PP)$       |
| $N' \rightarrow N (PP)$    | $V' \rightarrow V (NP)$           | $Adj' \rightarrow Adj (PP)$             | $P' \rightarrow P (\{NP/PP\})$ |

Fig. 19 illustrates the X-bar structure of an NP and a VP. Fig. 20 makes use of a somewhat more abstract notation to illustrate the various types of structures generated by X-bar theory.

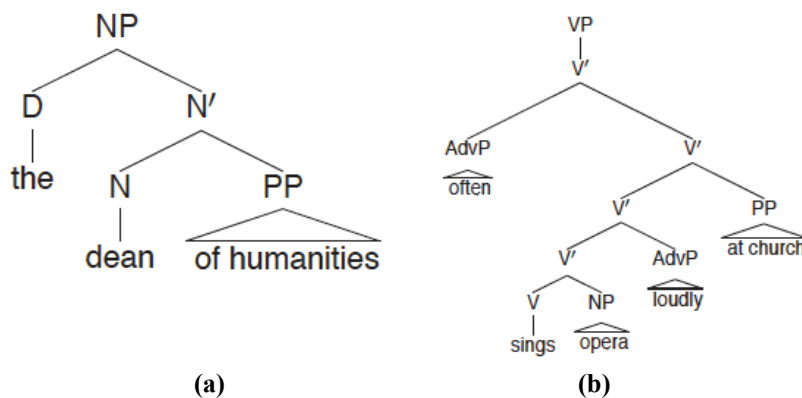


**Figure 19:** Examples of X-bar structures for verbs and nouns. In (a), the determiner 'the' is a specifier, the noun 'dean' is the head of the phrase, and the prepositional phrase 'of humanities' is the complement of the noun. In (b), the adverbial phrase 'often' is the specifier, the verb 'sings' is the head of the phrase, the noun 'opera' is the complement of the verb, and the phrases 'loudly' and 'at church' are adjunct phrases.

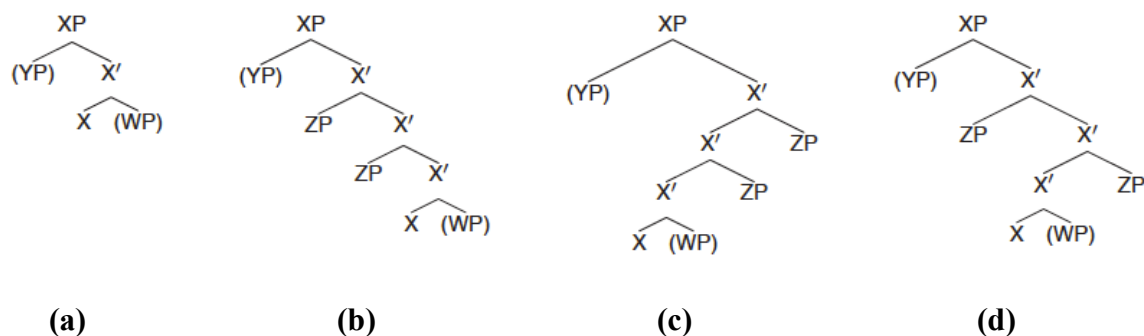


**Figure 20:** The phrase structure type in (a) is the most basic X-bar structure. It has slots for the specifier phrase (YP) and a complement phrase (WP). The structures in (b)–(d) have additional slots for adjunct phrases (ZP). The differences between (b), (c), and (d) all have to do with the location of the adjunct phrases (ZPs).

What is distinctive about X-bar theory is that “bar nodes,” like N' and V' in fig. 19, introduce additional structure into phrases. And the positions defined in terms of these structures can be used to distinguish between *phrasal heads*, *specifier phrases*, *adjunct phrases*, and *complement phrases* (fig. 20). This allows X-bar theory to correctly predict various kinds of constituency relations within phrases. For instance, the structure in fig. 19b allows for a natural explanation of the fact that various parts of the verb phrase in sentence (75) can be replaced by ‘does so’ in sentences (76)–(78).<sup>86</sup>

- |                                                   |                                         |
|---------------------------------------------------|-----------------------------------------|
| (75) Alia often [sings opera loudly at church].   | The VP is bracketed.                    |
| (76) Erika [does so], too.                        | ‘Does so’ replaces the entire VP.       |
| (77) Jennifer often [does so] at sporting events. | ‘Does so’ replaces ‘sings opera loudly’ |
| (78) Oriana [does so] quietly at work.            | ‘Does so’ replaces ‘sings opera’        |

Each phrasal head—i.e., each noun, verb, adverb, adjective, and preposition—projects an X-bar structure like those in figs. 19 and 20. These are then linked together to generate phrase

<sup>86</sup> See Haegeman (1994) for a motivation of X-bar theory and the details of various constituency tests, like the ‘does so’ test that is demonstrated in the main text. Berwick (1991b: pp. 195–203) discusses the virtues of parsing with X-bar theory “directly,” i.e., without first translating X-bar principles into a set of context-free rules. Using this technique, Berwick claims to be able to achieve a psychologically plausible *deterministic* parse, *without lookahead*—plainly a vast improvement over the original deterministic parser presented in Marcus (1980).

structure trees for entire sentences, yielding D-structures descriptions. As discussed above, the transformational rule *Move- $\alpha$*  applies to a phrase marker at D-structure, moving around any element of the structure, thus generating an array of S-structure phrase markers. A variety of principles then serve to filter out some of these and retain others, as illustrated in figs. 17 & 21.

One such principle is the *Empty Category Principle* (ECP), which specifies the structural configurations in which empty categories can appear in a well-formed phrase markers. Another set of principles constitutes the *Binding Theory*, which specifies the permissible relations between pronouns and their antecedents. These serve to rule out structures like (79)-(80).

(79) \*Jeremy<sub>i</sub> believes Dan likes himself<sub>i</sub>.  
(Coindexing 'Jeremy' and 'himself' violates Binding Theory.)

(80) \*Who<sub>i</sub> do you think that *wh*-trace<sub>i</sub> will like me?  
(Coindexing 'What' and '*wh*-trace' violates the ECP.)

Another principle is the *Theta-Criterion*. This principle ensures that the argument structure (i.e., the “ $\theta$ -grid”) of each verb is satisfied at D-structure and beyond. For instance, the verbs ‘sleep’, ‘wear’, and ‘put’ take one, two, and three arguments respectively. In the terminology of  $\theta$ -theory, these verbs have one, two, and three “ $\theta$ -roles” in their respective “ $\theta$ -grids.” This is demonstrated by the following examples.<sup>87</sup>

(81) David is sleeping. / \*David is sleeping the blanket. / \*David is sleeping the blanket on the table.

(82) \*David is wearing. / David is wearing the shirt. / \*David is wearing the shirt on the table.

(83) \*David is putting. / \*David is putting the book. / David is putting the book on the table.

---

<sup>87</sup> The examples given here gloss over a great many of subtleties. In the interest of brevity, I shall not enter into a digression concerning the syntactic arguments for the distinction between arguments and adjuncts, nor the tests for deciding whether a seemingly complete sentence is elliptical for a one that contains additional arguments in its overt form. These are delicate matters and a full discussion would take us well beyond the scope of the present work.

The  $\theta$ -Criterion rules out the ill-formed structures, starred above. Finally, a principle known as the Case Filter ensures that all overt nouns are assigned Case. According to this principle, a noun can only be assigned Case in certain structural configurations. The Case Filter rules out structures like (84), in which the subject is not assigned Case.

(84) \*It is possible Eugene to play soccer.  
(The noun 'Eugene' cannot receive Case in this configuration.)

Fig. 21 summarizes the GB principles that assign syntactic structures to sentences, beginning with the initial structural assignments allowed by X-bar theory, proceeding through the various filters described above, and eventuating in a well-formed phrase marker. The principles are partially ordered, such that the application of some requires the prior application of others. For instance, as fig. 21 illustrates, the Case filter cannot apply until movement operations have applied.<sup>88</sup>

A grammar consisting of X-bar theory, movement operations, and filtering principles can be stated in a compact axiomatic form.<sup>89</sup> This accounts for its elegance and increased explanatory power over older transformational grammars. However, to repeat a point that was stressed in earlier sections, neither the compactness nor the explanatory power of a grammar can, by itself, constitute grounds for ascribing psychological reality to the rules or principles of that grammar (Stabler, 1984). Independent arguments must be supplied even for seeing the *phrase markers* that are generated by a grammar as psychologically real (chs. III-IV). It should be even more evident that the *rules or principles* of a successful grammar are not automatically “in the mind” (chs. I and II).

---

<sup>88</sup> Here, I am ignoring so-called “inherent” case-marking, which was hypothesized to take place at D-structure.

<sup>89</sup> See for instance Johnson (1989; 1991), Berwick (1991a, 1991b), Stabler (1992), and §5.2 below.

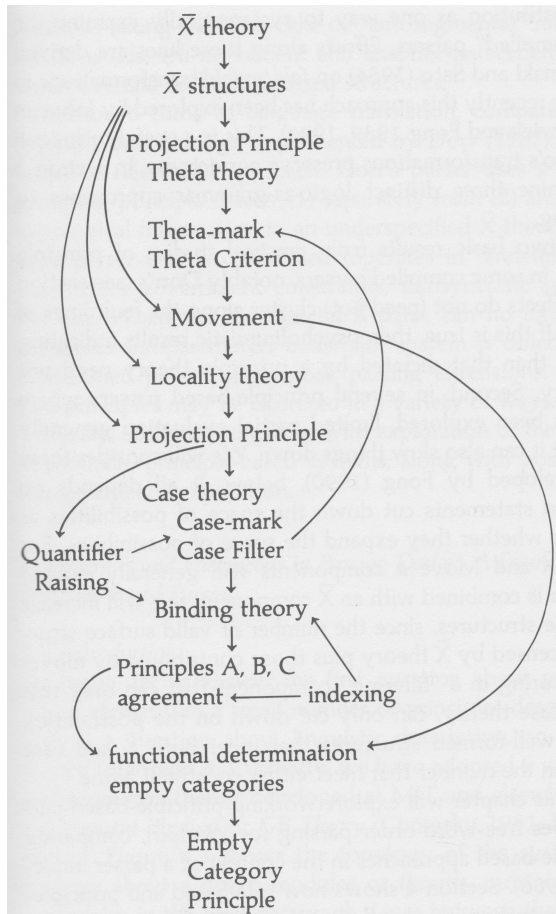


**Figure 21:** The GB principles that assign syntactic structures to sentences, beginning (at the top) with the initial structural assignments allowed by X-bar theory, proceeding through the various filters described above (Binding, ECP, Theta Criterion, etc.), and eventuating in a well-formed phrase marker. The principles are partially ordered, such that the application of some requires the prior application of others. For instance, the Case filter cannot apply until movement operations have applied. *Source:* Berwick (1991b)

In chs. III and IV, we saw reasons to believe that S-structure descriptions should, indeed, be seen as psychologically real representations—what I have been calling mental phrase markers (MPMs). By contrast, D-structure descriptions did not fare so well. Recall the experiment reported in Bock et al. (1992), the results of which cast doubt on the claim that the HSPM constructs D-structure representations for the purposes of comprehension and production. The principle-based parsers described in the following section comport with this conclusion. In describing his model, Johnson (1989) rejects the claim that a “transparent implementation” of

GB theory in a parser requires constructing representations of the D-structure of the input.

Use of knowledge of a level of representation does not entail the explicit construction of that representation, as I demonstrate by exhibiting a parser that uses knowledge of D-structures in the parsing process yet avoids the explicit construction of this level of representation. (p. 105)

Johnson's model does not, of course, provide any comfort to theorists who reject the need for MPMs (ch. IV). In general, "principle-based systems output very rich and accurate structural descriptions, including logical form representations" (Yang and Berwick, 1996: p. 370).

In connection with RT and the psychological reality issue, we must ask whether principles-based parsers incorporate (implicit or explicit) representations of the GB rules and principles. We will be in a position to address this issue only after we've familiarized ourselves with their main features and their algorithmic operations. Let us now turn to those.

### *§5.2 Principles-based parsing*

With the rise of the P&P approach to language acquisition, and the descriptive success of the GB theory, it was only natural that computational linguists would embark on the project of implementing these novel ideas in parsing models. In what follows, I present the main features of a number of efforts in this direction, drawing heavily on the discussions in Berwick (1991a,b).

One striking feature of principle-based parsers is their flexibility with regard to different languages. Yang and Berwick (1996) point out that "[t]raditional parsing technologies utilize language-particular, rule-based formalisms, which usually result in large and inflexible systems (Marcus, 1980)." By contrast, the principle-based parser that Yang and Berwick designed can be used to assign structure to sentences from a variety of typologically distinct languages. To switch from one language to another, all the parser needs to do is swap out one lexicon for another and set the parameters to the values characteristic of the target language. For instance,

an X-bar parameter can be changed from the head-initial setting (as in, e.g., English) to a head-final setting (as in, e.g., Japanese). Similarly, the Case-filter can be set so that only the members of some desired class of phrases are capable of assigning Case (in particular configurations).

It is believed that languages are constrained by a small number of universal principles, with linguistic variations largely specified by parametric settings. The merit of principle-based parsing is two-fold. As a tool for linguists, it is directly rooted in grammatical theories. Therefore, linguistic problems, particularly those that involve complex interactions among linguistic principles, can be cast in a computational framework and extensively studied by drawing directly on an already-substantiated linguistic platform. It is designed from the start to accommodate a wide range of languages — not just ‘Eurocentric’ Romance or Germanic languages. Japanese, Korean, Hindi and Bangla have all been relatively easily modeled in PAPPI (Berwick and Fong 1991, Berwick forthcoming). Differences among languages reduce to distinct dictionaries, required in any case, plus parametric variation in the principles. (Yang and Berwick, 1996)

This kind of flexibility is particularly impressive when the languages are radically different from one another.

Berwick (1991b) discusses at length the advantages of a principle-based parser for the free word-order languages, Warlpiri and Japanese. According to him, attempts to build parsers and translation systems for these languages on the basis of context-free phrase structure rules were discovered to be inefficient and ultimately unsuccessful (pp. 163-170). The problem was that context-free rules fuse together information about two distinct linguistic properties—hierarchical dominance and linear precedence (ch. IV, §3). Though X-bar theory was presented above as a kind of context-free phrase structure grammar, Berwick discusses ways of implementing principle-based parsers without appealing to context-free rules (examples below). He also demonstrates that the principle-based parser designed for Warlpiri performs much better than older models, which were designed to use context-free grammars. The following results should give an idea of the impressive progress that principle-based systems have made in achieving broad coverage within an interesting variety of languages.

We are able to parse sentences with the range of structures including Wh-movement, the Binding Theory, Quantifier Scoping, the BA-construction to complex NP (clausal, possessive, and numeral/ classifier). All testing sentences are correctly analyzed: LF logical form representations are computed for the grammatical sentences and the ungrammatical ones are ruled out ones with linguistic principle violation(s) shown. Each parse takes no more than 2 seconds on a Sparc10 workstation. Overall, excluding a hand-wired dictionary, less than 100 additional lines of Prolog are required.<sup>[90]</sup> Because the PAPPI system implements its model linguistic theory faithfully, adapting new languages is expected to be quite minimal, as our implementation shows. Additionally, it provides a platform on which linguists can experiment theoretical proposals extensively and also cross-linguistically, without having to know much about the internal design of the parser. Furthermore, principle-based systems output very rich and accurate structural descriptions, including logical form representations, that assist in more engineering-oriented NLP tasks that go beyond parsing (Yang and Berwick, 1996: p. 370).

Besides their flexibility with respect to distinct languages, principle-based parsers exhibit tolerance with respect to ungrammatical input. Indeed, a binary notion of grammaticality is difficult to maintain in the context of a principle-based syntactic theory. On its way from X-bar analysis to the ultimate assignment of structure, linguistic input passes through a series of filters, as described above (figs. 17 & 21). Hence, almost every input is assigned *some* interpretation as it runs this gauntlet. Throughout the process, the parser takes note of any failure on the part of the input to satisfy one or another principle of the grammar. In the human case, this would give rise to variable-strength judgments of ungrammaticality—the more principles are violated, the more ungrammatical the input will be judged to be. This stands in contrast with the behavior of parsers based on context-free grammars and ATN grammars, discussed above (§§2, 4).

For our purposes, one of the most interesting facts about principle-based parsers is that they make heavy use of the Parsing as Deduction (PAD) approach, discussed in §2.7. And, as the theorists who develop these models are eager to point out, the combination of GB theory and the PAD yields a satisfying interpretation of Chomsky's talk of "knowledge of language."

---

<sup>90</sup> Both the hardware and the Prolog interpreters have advanced quite a bit since 1996. Nowadays, the parse would take considerably less than 1 second.



the human language processor uses its knowledge of a language to obtain knowledge of the utterances of that language in the “same way” that a theorem-prover “uses” axioms to deduce their consequences. ... My primary goal here is to demonstrate that the Parsing as Deduction hypothesis does in fact provide a viable explanation for how it is that knowledge of a language of the form attributed to human language users by modern GB theory (Chomsky, 1981; Van Riemsdijk & Williams, 1986) can actually be used to obtain knowledge of the utterances of that language. (Johnson, 1989: pp. 105-106)

As Johnson’s paper argues, deductive inference is still perhaps the clearest way to think about how to ‘use’ knowledge of language. In a certain sense, it even seems straightforward. The terms in the definitions like [that of the Case filter] have a suggestive logical ring to them, and even include informal quantifiers like *every*; terms like lexical NP can be predicates, and so forth. In this way, one is led to first-order logic or Horn clause logic implementation (Prolog) as a natural first choice or implementation, and there have been several such principle-based parsers written besides those described by authors of this volume who have build Prolog implementations of principle-based theories (see Sharp, 1985; Kolb and Thiersch, 1988). Parsing amounts to using a theorem prover to search through the space of possible satisfying representations to find a parse... (Berwick, 1991a)

In the latter passage, Berwick refers to the idea of formalizing GB principles in first-order logic. Understanding how this can be done is crucial to making sense of how the PAD approach can be applied to GB theory. Let’s look at a concrete example.

One of the “filters” described in §5.1 is the Empty Category Principle (ECP), which can be stated as follows: *Empty categories must be properly governed*. Formulated in this way, the ECP contains technical terms that stand in need of definition. Passing over some complex issues (Haegeman, 1994; Carnie, 2010), we can make do with the following:

- Proper government: X properly governs Y if and only if X governs Y and X is a lexical category (i.e., N, V, P, Adv, Adj).
- Government: X governs Y if and only if Y is contained in the maximal X-bar projection of X (X-max), and X-max is the smallest maximal projection containing Y, and X c-commands Y.
- C-command: X c-commands Y if and only if the first branching node dominating X also dominates Y, and X does not itself dominate Y.

When made precise, the relations defined above can be expressed in first-order logic, as follows.

(The symbol ' $\leftrightarrow$ ' expresses the biconditional.)

Proper government:  $\text{properly-governs}(X, Y) \leftrightarrow \text{governs}(X, Y) \ \& \ \text{lexical-category}(X)$

Lexical Category:  $\text{lexical-category}(X) \leftrightarrow \text{Noun}(X) \vee \text{Verb}(X) \vee \text{Prep}(X) \vee \text{Adv}(X) \vee \text{Adj}(X)$

Government:  $\text{governs}(X, Y) \leftrightarrow \text{maximal-projection}(X, X_{\text{max}}) \ \& \ \text{dominates}(X_{\text{max}}, Y) \ \& \ \text{least-maximal-projection}(Y, X_{\text{max}}) \ \& \ \text{c-commands}(X, Y)$

C-command:  $\text{c-commands}(X, Y) \leftrightarrow \text{first-branching-node-from}(X, \text{BrNode}) \ \& \ \text{dominates}(\text{BrNode}, Y) \ \& \ \sim \text{dominates}(X, Y) \ \& \ \sim \text{dominates}(Y, X)$

This, in turn, allows us to formulate the Empty Category Principle as follows:

(ECP)  $\forall Y[\text{empty-category}(Y) \rightarrow \exists X(\text{properly-governs}(X, Y))]$

With this definition in hand, developers of principle-based parsing systems can use a Prolog or LISP script to scan set of candidate phrase markers generated by the X-bar module, and to assign partial ungrammaticality to those that fail to satisfy the definition. Berwick provides another example of this strategy, using a definition of the Case filter, stated directly in Prolog.

How then can a theory written in English be mapped into logic? Here for example, following Fong (1991, forthcoming) is a typical statement of the Case filter: “At S-structure, every lexical NP needs Case.” How are we to implement this statement as part of a parsing program? ... We must actually translate the English connectives such as needs or every and the required feature checking (the property of being *lexical* on NPs), into logical formulas. But in fact this can readily be done. Fong’s system ... states principles purely declaratively, as in his version of the Case filter.

```
:- caseFilter in_all_configurations CF where
  lexicalNP(CF) then assignedCase(CF).
lexicalNP(NP) :- cat(NP, np), \+ ec(NP).
assignedCase(X) :- X has_feature case(Case),
  assigned(Case).
```

This says, close in spirit to the original English, that the Case filter is met if, in all configuration (CF) where there is a lexical NP, then make sure that Case has been assigned to that NP configuration. The Prolog has been augmented a bit with more abstract macros such as `in_all_configurations` that render the code closer to English. (Following Fong we read the second clause `lexicalNP(NP)` as stating that a lexical NP is one that has the category feature *np* and is not an empty category. As is conventional `\+` means negation-as-failure—if we can’t prove something to be an `ec` or empty category then it is not one. Finally, the comma is a conjunction.) (Berwick, 1991a: pp. 14-15)

My hope is that these two examples—the ECP and the Case Filter—serve to make clear what theorists like Yang and Berwick (1996) have in mind when they claim that “A principle-based parser transparently reflects the structure of the contemporary linguistic theory, the Principles and Parameters framework” (p. 363).

We are now in a position to examine the computational role of GB principles in the kinds of parsers that have been implemented thus far. Fig. 22 illustrates the flow of information through PRINCIPAR, a parsing system developed by Lin (1993, 1994). Inspecting the diagram, we see that Lin characterizes the grammar as a *static data structure*, which is consulted or accessed by the parser. This conforms to what Devitt (2006) calls position (ii) on the psychological reality issue, discussed in §1 and §4, and repeated here for convenience:

- (ii) The structure rules of a language are represented and used *as data* by the processing rules of language use.

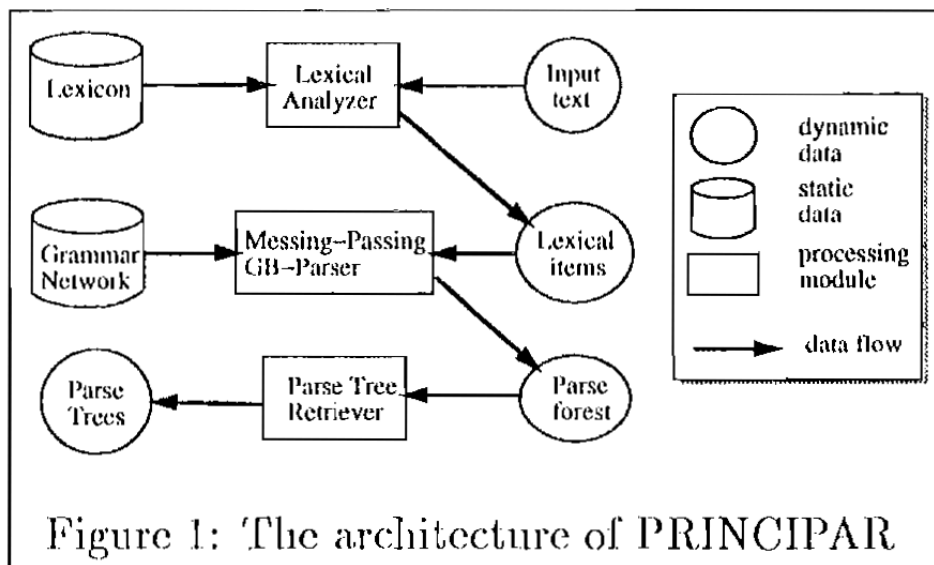


**Figure 22:** An information-flow diagram of the PRINCIPAR system developed by Lin (1993, 1994). Lin annotates it as follows: “Sentence analysis is divided into three steps. The lexical analyser first converts the input sentence into a set of lexical items. Then, a message passing algorithm for GB-parsing is used to construct a shared parse forest. Finally, a parse tree retriever is used to enumerate the parse trees” (1994, p. 482). Note the explicit characterization of the grammar as a static data structure, which is consulted/accessed by the parser. This conforms to

what Devitt (2006) calls position (ii) on the psychological reality issue (see §1 and §4 for details).

Complicating matters somewhat, Berwick (1991b) discusses a number of ways in which the GB grammar can be related to a principles-based parser. He writes:

It seems, then, that there is at least a strong and a weak sense of how a system might “use” a set of principles. In a strong sense, the principles might be literally represented as such, individually, and used directly online as data structures for parsing. For example, this is roughly how the rules of a context-free grammar might be used in, say, Earley’s algorithm... In a weaker sense, the principles might be *interpreted*.

In this passage, Berwick lends support to our claim (§2.3) that the Earley parsing algorithm is strongly suggestive of position (ii) on the psychological reality issue, and hence RT. However, his talk of a “weaker sense” in which the principles are “interpreted” might seem to cast doubt on this idea. Further investigation reveals that it does not. For, when Berwick goes on to discuss what might be meant by “interpreted,” his primary example is a logical procedure known as partial evaluation. A partial evaluation of the predicate ‘*properly-governs*(*X*, *Y*)’ is given in (85).

(85) *properly-governs*(*X*, *Y*)  $\leftrightarrow$  *maximal-projection*(*X*, *Y*) & *dominates*(*X*<sub>max</sub>, *Y*) & *least-maximal-projection*(*Y*, *X*<sub>max</sub>) & *first-branching-node-from*(*X*, *BrNode*) & *dominates*(*BrNode*, *Y*) &  $\sim$ *dominates*(*X*, *Y*) &  $\sim$  *dominates*(*Y*, *X*)

It should be clear that “partial evaluation” is nothing more than the logical combination (in this case, simple conjunction) of various distinct principles and relations defined by the grammar. Hence, Berwick’s distinction between the strong and weak senses in which a parser can use a grammar has no bearing on whether the rules or principles are explicitly represented and used as data. It has to do, instead, with the *logical format* in which they are represented in the parser, and the extent to which this corresponds to how they are represented in linguistic textbooks.

Although Berwick’s distinction does not bear on RT, it does bear on our earlier discussion of efficiency considerations (§2.8). For, it turns out that the logical format of the data structure in which the GB principles are represented can have significant consequences for the

efficiency of the parser. Using a slightly different terminology to express the same ideas, Merlo (1995) discusses in detail her experimental efforts to increase the efficiency of a principle-based system by “precompiling” the grammar, thus “multiplying out” the interactions of the principles. She writes: “By compilation ... I mean off-line computation of some general property of the grammar, for example the off-line computation of the interaction of principles, using partial evaluation or variable substitution” (p. 517, fn. 2). Merlo concludes that (pre)compilation is useful up to a point, but the more of it one does, the more rules one creates. And the more rules the system has to apply, the longer it takes to return a parse. So there is a point at which compiling actually does more harm than good. The PRINCIPAR system discussed above offers a novel approach to the efficiency problem, which incidentally provides support to the conception of mental phrase markers that I advocated in ch. IV. Lin (1994) writes:

Principle-based grammars, such as Government-Binding (GB) theory (Chomsky, 1981; Haegeman, 1991), offer many advantages over rule-based and unification-based grammars, such as the universality of principles and modularity of components in the grammar. Principles are constraints over X-bar structures. Most previous principle-based parsers, e.g., (Dorr, 1991; Fong, 1991; Johnson, 1991), essentially generate all possible X-bar structures of a sentence and then use the principles to filter out the illicit ones. The drawback of this approach is the inefficiency due to the large number of candidate structures to be filtered out. The problem persists even when various techniques such as optimal ordering of principles (Fong, 1991), and co-routining (Dorr, 1991; Johnson, 1991) are used. ... This paper describes an efficient, broad coverage, principle-based parser, called PRINCIPAR. The main innovation in PRINCIPAR is that it applies principles to *descriptions of X-bar structures rather than the structures themselves*. X-bar structures of a sentence are only built when their descriptions have satisfied all the principles. (p. 482, emphasis added)

Unsurprisingly, recent approaches to principle-based parsing seek to increase efficiency by drawing on probabilistic information. Berthouzoz and Merlo (1997) conducted a series of comparative studies, which yielded the following results:

We have presented some experiments in evaluating the performance of a symbolic parser for English (IPS). Due to its massively parallel nature, IPS has good recall

on the set of sentences that we tested, but precision is very low, showing that overgeneration of constructions needs to be reduced. We have tested two methods. In the structural method, both the generating and the filtering device are based on knowledge stored in a linguistically-defined grammar. In the hybrid method, the generating device is structure-based but the filtering is done based on statistical information extracted from a corpus. We observe that the statistical method outperforms the structural approach, that it is less computationally expensive, and it is easier to define and develop. (p. 185)

A detailed discussion of these approaches would take us to far afield. Suffice it to say that issues of efficiency are being addressed head-on, with what appears to be remarkable success.

Let us sum up: P&P grammars are, in many respects, improvements over previous transformational grammars. In addition to their formal virtues, they suggest a novel framework within which to approach questions about both the acquisition and the processing of language. In the processing domain, they have been used in systems that deliver fast, reliable, and detailed parses of a impressively broad range of natural-language constructions, in a variety of typologically distinct languages. These parses are, moreover, fine-grained enough to induce degrees of grammaticality, thereby replicating a key feature of human performance. Finally, the systems in question make use of the PAD framework in fleshing out the oft-repeated claim that knowledge of syntax is put to use in language processing. The most common way in which they do so is by explicitly representing grammatical rules and principles and drawing on them as data, in accordance with position **(ii)** on the psychological reality of language.

We now turn to the most recent development in the P&P framework—i.e., what Chomsky (1995) has called the Minimalist program. Like the Government and Binding theory, Minimalism has had a profound effect on the theory of sentence processing. We will see that parsing in accordance with Minimalist grammars is often a matter of directly applying the ground-level operations of *Merge* and *Move*. As with the ATN framework discussed in §4,

Minimalism provides fresh support for positions **(i)** and **(iii)** on the psychological reality issue, and also some support for position **(ii)**.

## §6. Parsing with Minimalist Grammars

### §6.1 *Minimalism: The Basics (Features, Merge, Move, and Spell-Out)*

The Minimalist Program, as the name suggests, is an attempt to achieve the explanatory power of GB, but with more modest theoretical machinery—ideally, the bare minimum. The central notions of GB, viz., D-structure and Government, are argued to be superfluous. Their explanatory functions are absorbed by components of the grammar that *all* theorists must posit—i.e., the lexicon and a recursive structure-building operation. I will not attempt to motivate the central claims of Minimalist grammars here. My aim is simply to sketch those claims, and proceed directly to a discussion of their role in parsing theory. The presentation below draws heavily on Stabler (1994), Weinberg (1999), and Harkema (2001).

As noted above, Minimalist grammars give a prominent role to the lexicon. The grammar conceives of the lexicon as the starting point of a *derivation*—a process that begins when a batch of lexical entries are delivered to the syntactic component of the grammar. The syntax is seen as nothing more than a conduit between the lexicon and both the phonological component (PF) and the “conceptual-intentional” component (CI) of the grammar. Lexical entries are treated as *bundles of features*. (Once again, Hume would have approved.) Some features can only be “read” by PF, others only by CI. PF and CI thus impose “legibility conditions” on the features of an expression; in order to be read by those components, the lexical features must be “interpretable” by them. Other features, however, are only interpretable within the *syntactic* component—hence *uninterpretable* by PF or CI. The syntactic component of the

grammar must, therefore, operate so as to output *only* the interpretable features to PF and to CI. It is the only component of the grammar that is a theory-internal posit (PF and CI being obviously necessary to characterize vocalization and thought). For this reason, parsimony dictates that its role be minimized to the extent possible. This gives rise to the so-called “Economy Conditions” discussed below.

How, then, can the syntax rid expressions of the purely syntactic features that they get from the lexicon? The answer is: *feature checking*—i.e., relating lexical items that “need” some feature, *f*, to lexical items that “need” to get rid of that very feature. (Informally, one can think of it as a kind of valence picture.) Once “checked,” a purely syntactic feature is deleted, hence no longer problematic for the PF or CI components. By contrast, purely syntactic features that are left unchecked engender problems at PF or CI, causing the derivation to “crash.”

The kinds of features that a lexical item might have include its parts of speech (complementizer, tense marker, verb, determiner, noun, adjective, etc.), its subcategorization and argument-selecting theta-features, Case-assigning or Case-requiring features, agreement features for person, number, and gender, and others. In symbolic form:

|                              |                          |
|------------------------------|--------------------------|
| Part of speech:              | {c, t, v, d, n, a, ...}  |
| Argument selecting features: | {=x, x=, =X, X=}         |
| Licensors (the providers):   | {+case, +wh, + focus...} |
| Licensees (the needy):       | {-case, -wh, -focus...}  |

(Note that the direction of the symbol ‘=’ indicates whether the item needs its features checked on its left or on its right.<sup>91</sup>) Here, then, are some concrete examples of lexical items:

---

<sup>91</sup> Note also that the features rendered above in capital letters are “Strong,” while those rendered in lower-case are “Weak”—a distinction that will play no role in subsequent discussion. To use language which we have not yet introduced, the Strong features, unlike the Weak ones, require *phonological* features to be displaced within the structure. (I discuss displacement shortly.)



| Word                | Feature Bundle           | Interpretation of the feature symbols         |
|---------------------|--------------------------|-----------------------------------------------|
| <i>The</i>          | { d , =n , - case }      | (determiner, selects a noun, needs case)      |
| <i>student</i>      | { n }                    | (noun)                                        |
| <i>wrote</i>        | { v , =d , + case , =d } | (verb, selects 2 DPs, assigns case to one DP) |
| <i>a</i>            | { d , =n , - case }      | (determiner, selects a noun, needs case)      |
| <i>dissertation</i> | { n }                    | (noun)                                        |

Now that we have a lexicon in place, we need an “abstract mechanism” to implement the feature-checking operation discussed above. This mechanism must itself be an operation that relates the features of two lexical items, and it must be recursive, so as to account for the productivity of natural language (ch. I). The simplest possible such operation is *Merge*, which takes two lexical items X and Y, and places them in a structure {X, Y}. The structure inherits the features of its members, some of which can then enter into a checking relationship. (They are said to be “in construction” with one another.) Because *Merge* is recursive, it can take as input its own outputs, combining {X, Y} with W to form the object {W, {X, Y}}, and then with Z to form {Z, {W, {X, Y}}}, and so on. In this way, hierarchical structures are generated. (We can represent these, as usual, with binary-branching trees, as in fig. 23 below.) Here is a simple example, using the lexical items displayed above.

|                                | Words        | Feature Structures            |
|--------------------------------|--------------|-------------------------------|
| <b>Input to <i>Merge</i>:</b>  | ‘wrote’, ‘a’ | { v , =d , +case , =d } , {d} |
| <b>Output of <i>Merge</i>:</b> | ‘wrote a’    | { v , =d }                    |

When the two feature bundles (i.e., words) are merged, the {d} feature of the determiner ‘a’ “checks” the {=d} feature of the verb ‘wrote’. It’s as though ‘wrote’ calls out and says “I need a determiner,” and the determiner ‘a’ comes along to supply that feature. The result is a structure that has features { v , =d }, i.e., a verb that “needs” one more determiner. The example in fig.

23 illustrates the same idea, except also introduces the pointers ‘>’ and ‘<’ to indicate the location of the head of the phrase.

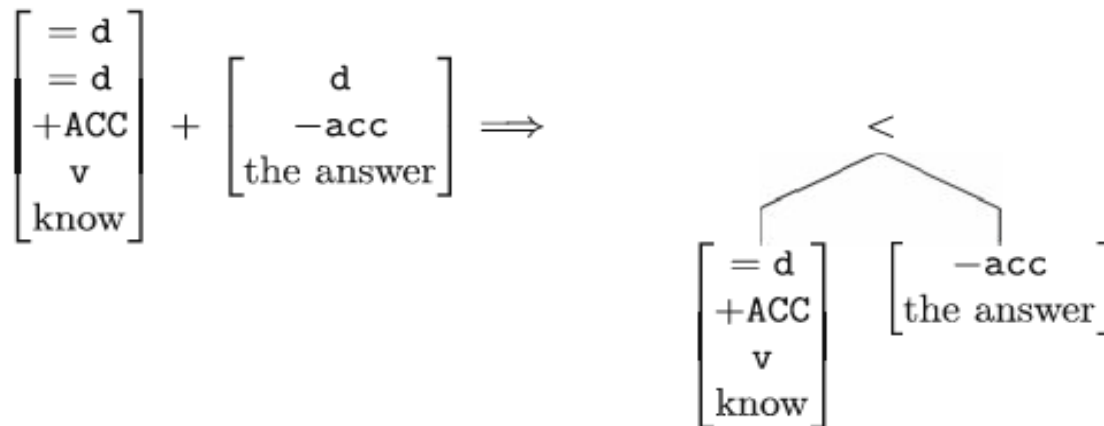


**Figure 23:** The two feature bundles associated with ‘know’ and ‘the answer’ are merged, creating a tree-structure whose head is ‘know’ (as indicated by the direction of the ‘<’ pointer). *Merge* is triggered by the need to check the {=d} feature on ‘know’ and the {d} feature on ‘the answer’. *Source:* Gerth and beim Graben (2009)

Some Minimalist grammars make a distinction between *internal Merge* and *external Merge*; others posit two separate operations, *Merge* and *Move*. I pass over the differences between these competing approaches, adopting the *Merge* and *Move* option on account of its familiarity. *Move* is an operation that displaces the most-embedded element in the structure, in order to check its syntactic features before any further *Merge* operations can apply. Fig. 24 provides an example in which the phrase ‘the answer’ moves up in the tree, in order to check its Accusative Case features {-acc} against the feature {+ACC} on ‘know’.

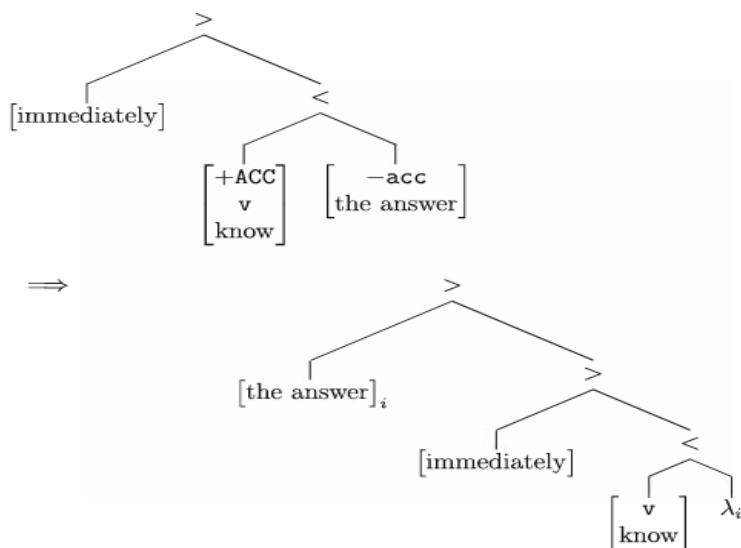


**Figure 24:** The expression ‘the answer’ is moved from its most-embedded in the tree position (top panel) to its new least-embedded position (bottom panel), leaving behind a trace ( $\lambda_i$ ) in its initial position. The movement results in a checking of the feature  $\{-acc\}$  on ‘the answer’ and the feature  $\{+ACC\}$  on ‘know’. (Gerth and beim Graben, 2009)

As noted above, a central tenet of the Minimalist Program is its insistence on what Chomsky calls “Economy Conditions.” Recall that, in GB, the rule *Move- $\alpha$*  applied to D-structure trees indiscriminately, generating a batch of structures, most of which are then filtered out by the principles of the grammar. From the Minimalist perspective, this is a profligate system. Minimalist grammars take the opposite approach: nothing is generated or operated upon unless something else makes it absolutely necessary. And, even then, the most economical operation is the only one that applies. (This is the principle of Last Resort.) Economy thus dictates that *Merge* and *Move* are always and only triggered by the need to check features.

Another example of economy constraints has to do with how the syntactic component of the grammar outputs material to PF—an operation known as *Spell-Out*. Given that we speak, sign, and write in linear structures that do not explicitly reflect hierarchical relations (ch. IV, §3.1.3), the output of the syntax to PF is a syntactically unstructured string. As Weinberg (1999) puts it, “Spell-Out turns a syntactic structure with relevant constituent relationships into a string

ready for phonological interpretation” (p. 287). The question arises: When in the derivation does Spell-Out take place? And how often? Uriagereka (1999) argues that Spell-Out applies whenever neither *Merge* nor *Move* can apply in the course of a derivation.<sup>92</sup> Weinberg (1999) capitalizes on this idea in providing an account of ambiguity resolution preferences, which we will examine in §6.2.2. Her account relies crucially on the following idea:

Within the context of the Minimalist Program, Spell-Out is a grammatical operation, on a par with movement transformations. As such, it is governed by conditions on transformations, in particular by the economy conditions discussed above. These conditions establish a preference for derivations that utilize the fewest operations possible. An operation is applied only to satisfy some independent grammatical condition. In this case, this means that the system spells out or linearizes only when it cannot otherwise establish a chain of precedence. (p. 288)

We have now introduced the basic components of Minimalist grammars: the syntax, PF, CI, legibility at the interfaces, lexical features, feature checking, *Merge*, *Move*, Economy conditions, and Spell-Out. We end this crash course by pointing out that the P&P component of the Minimalist grammars consists of various principles regarding feature checking, and language-specific parameters that specify which functional categories—Tense Phrase (TP), Agreement Phrase (AgrP), and so forth—will carry which features. Given the prominent role that features play in determining hierarchical structure, it turns out that it may well be possible to characterize the differences between all languages in just these terms—a very exciting prospect.

---

<sup>92</sup> Weinberg (1999) summarizes Uriagereka’s argument as follows: “Uriagereka uses Spell-Out as a repair mechanism to retain one-to-one correspondence between dominance and precedence. He assumes that both precedence and dominance must be established between terminal elements at all points of the derivation. Precedence implies merger, and merger is only possible when a chain of dominance can be established. When merger is not possible, the string is linearized (turned into an unstructured string where only previously established precedence relations are preserved). Since the elements that have been linearized are invisible in the syntax, precedence does not have to be established between them and other items in the structure. Thus, when two categories cannot be combined through merger or movement (the only syntactic operations) to form a dominating category, the material that has been given structure so far is “spelled out” or linearized... This idea preserves the one-to-one mapping between precedence and dominance, but only at the cost of never building single phrase markers. Instead, the system builds blocks...where all elements stand in a c-command relation to each other. When this c-command relation is interrupted, the unit is spelled out...” (pp. 287-8).

## §6.2 Parsing with Minimalist Grammars

In formal syntax, the Minimalist Program was underway in the early 1990s. As with the other formalisms that we've examined in this chapter, it only took about a decade for Minimalist grammars to make an impact on theories of sentence processing. In what follows, we will look first at a simple application, due to Berwick (1997). Berwick sees *Merge* as corresponding to the reduction step in shift-reduce parsers, of the sort that we encountered in §2.4. It seems clear that this idea will be the core of any Minimalist parser. Unfortunately, this is not very informative, as it does not amount to an explicit parsing algorithm. Nor does it tell us much about how to deal with phrasal movement and gaps. Finally, it sheds no light on our central question, regarding how (or whether) the grammar is mentally represented.

To fill in these details, we look to Harkema (2001), who provides a detailed algorithmic treatment, within the PAD framework that we discussed above (§§2.7 and 5). Harkema specifies a top-down, a bottom-up, and an Earley parser (§2) for Minimalist grammars, and proves that these can run in polynomial time. However, he does not implement it. Physical implementations in a classical framework have been carried out by Fong (2011).<sup>93</sup> Gerth and beim Graben (2009) implement a Minimalist grammar in a connectionist network and even go some way toward demonstrating the psychological plausibility of their system. Unfortunately, limitations of space preclude a full discussion of these developments. Instead, I will focus in §6.2.2 on what, to my knowledge, is the most sustained argument for the psychological plausibility of Minimalist parsers, due to Weinberg (1999). As noted in §1, Weinberg's model provides support for positions **(i)** and **(iii)** on the psychological reality issue. In addition, Weinberg advances an

---

<sup>93</sup> Fong's parser is described, with visual aids, here: <http://dingo.sbs.arizona.edu/~sandiway/mpp/mons2011.pdf>

argument against two popular frequency-based constraint-satisfaction models of ambiguity resolution, due to Macdonald, *et al.* (1994) and Filip, *et al.* (2002).

### §6.2.1 Berwick and Harkema

Berwick (1997) extols the virtues of Minimalist grammars, of which the most significant for our purposes is their applicability in the domain of sentence processing. Berwick's examples of Minimalist parsing treat *Merge* as corresponding to the reduction step in shift-reduce parsers. Recall from §2.4 that such systems consist of a working memory stack that holds a small number of input items, as well as the results of prior parsing operations—i.e., prior “reductions,” which merge two items from the input stream. Berwick illustrates this scheme in fig. 25.

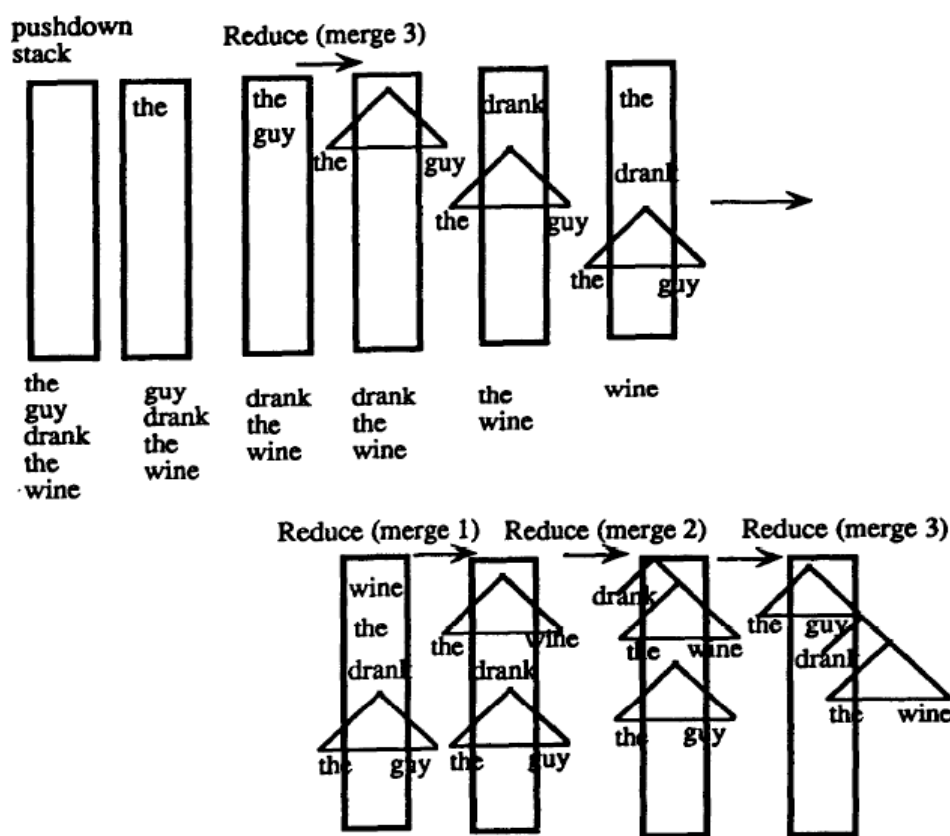


**Figure 25:** A bottom-up parse of the sentence *The guy drank the wine*. The vertical bars constitute the contents of the system's working memory buffer at each step of the parse. At first, the buffer is empty. In the first two steps, the items 'the' and 'guy' are shifted onto the stack. In

the next step (top panel), the operation *Merge* applies, creating a simple hierarchical structure. In the final step, two such structures are merged, thus completing the parse.

Although this is not represented in fig. 25, each of the lexical items is a feature bundle. The

parser reduces the items in the memory buffer in accordance with their feature structures,

checking syntactic features with every merger. Unfortunately, Berwick does not discuss the

details of the algorithm, and says nothing about how the parser would handle input that contains

gaps—e.g., wh-questions or relative-clause constructions.

Harkema (2001) supplies these details, working within the Parsing as Deduction (PAD) framework discussed above. As we've seen, the first step in applying this framework is always to formalize the rules of the grammar as inference rules for the purposes of formal deduction. In doing so, Harkema draws a distinction between different varieties of the transformational rules *Merge* and *Move*, yielding the five transformation types illustrated in fig. 26.

Merge-1:

$$\frac{[(p, q):=x\gamma]_s \quad [(q, v):x, \alpha_1, \dots, \alpha_k]_t}{[(p, v):\gamma, \alpha_1, \dots, \alpha_k]_c} \text{ Merge-1}$$

Two lexical items merge

Merge-2:

$$\frac{[(p, q):=x\gamma, \alpha_1, \dots, \alpha_k]_c \quad [(v, p):x, \iota_1, \dots, \iota_l]_t}{[(v, q):\gamma, \alpha_1, \dots, \alpha_k, \iota_1, \dots, \iota_l]_c} \text{ Merge-2}$$

A lexical item merges with a complex tree, but won't move later

Merge-3: ( $\delta \neq \emptyset$ )

$$\frac{[(p, q):=x\gamma, \alpha_1, \dots, \alpha_k]_{t_1} \quad [(v, w):x\delta, \iota_1, \dots, \iota_l]_{t_2}}{[(p, q):\gamma, \alpha_1, \dots, \alpha_k, (v, w):\delta, \iota_1, \dots, \iota_l]_c} \text{ Merge-3}$$

A lexical item merges with a complex tree, but will move later

Move-1:

$$\frac{[(p, q):+Y\gamma, \alpha_1, \dots, \alpha_{i-1}, (v, p):-Y, \alpha_{i+1}, \dots, \alpha_k]_c}{[(v, q):\gamma, \alpha_1, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_k]_c} \text{ Move-1}$$

An item moves, but will move no further

Move-2: ( $\delta \neq \emptyset$ )

$$\frac{[(p, q):+Y\gamma, \alpha_1, \dots, \alpha_{i-1}, (v, w):-Y\delta, \alpha_{i+1}, \dots, \alpha_k]_c}{[(p, q):\gamma, \alpha_1, \dots, \alpha_{i-1}, (v, w):\delta, \alpha_{i+1}, \dots, \alpha_k]_c} \text{ Move-2}$$

An item moves, and will move again later.

**Figure 26:** The five inference rules operative in Harkema's deductive parsing scheme. The cumulative effects of these rules simulate the *Move* and *Merge* operations in Minimalist syntax. Each bracketed item is a feature bundle. Some further details of the formalism are discussed in the main text. *Source:* Harkema (2001)

The inference rules in fig. 26 take lexical items as input and yield others as output. To take an example, the transitive verb *praise*, which has a *vt* feature and a *-v* agreement feature, merges with the noun *Lavinia*, to check their respective  $\{=d\}$  and  $\{d\}$  features. The result is this:

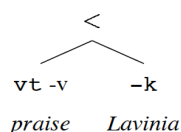


**Figure 27:** The result of a merger. The head of the phrase is *praise*, as indicated by the pointer '<', and *Lavinia* still needs Case  $\{-k\}$ . The structure can be represented as a tree (top) or in bracket notation (bottom). *Source:* Harkema (2001)

$[(1, 2):vt \ -v, (3, 4):-k]_c$

Repeated applications of Merge and Move comprise the derivation encoded in the tree diagram in fig. 28, which Harkema discusses in the following passage.

The bottom-up, top-down, and Earley-style parsers that will be presented in the remainder of this dissertation are defined on derivation trees of the sort given in [fig. 28 below]. A bottom-up parser will start from the lexical items at the leaves of the tree, which are provided by the words of the sentence to be parsed, and try to derive the expression at the root of the tree by repeatedly applying the functions Merge-1 through Move-2. A top-down parser will start from the expression at the root of the derivation tree and, by repeatedly applying the functions Merge-1 through Move-2 in reverse, will try to break it down into lexical items at the leaves matching the words of the sentence to be parsed. The Earley-style parser essentially is a top-down parser with some additional book-keeping machinery which makes it applicable to a wider range of grammars than a basic top-down algorithm.

The process that Harkema describes can be represented as a deductive proof procedure:



| Step | Lexical Item | Features            | Interpretation                                                                                                      |
|------|--------------|---------------------|---------------------------------------------------------------------------------------------------------------------|
| 1.   | Lavinia:     | {d, -k}             | <i>Lavinia</i> is a determiner (on this analysis) and needs Case                                                    |
| 2.   | Titus:       | {d, -k}             | <i>Titus</i> is a determiner (on this analysis) and needs Case                                                      |
| 3.   | praise:      | {=d, vt, -v}        | The transitive verb <i>praise</i> needs a determiner and a verbal agreement feature.                                |
| 4.   | -s:          | {=pred, +v, +k, i}  | The inflectional suffix <i>-s</i> needs a predicate to attach to, has a verbal agreement feature, and assigns Case. |
| 5.   | ε:           | {=i, c}             | An empty complementizer that needs an inflectional phrase                                                           |
| 6.   | ε:           | {=vt, +k, =d, pred} | An empty category that needs a determiner to assign Case to                                                         |

| Step | Lexical input to the operation          | Operation | Steps that serve as inputs |
|------|-----------------------------------------|-----------|----------------------------|
| 7.   | [praise: vt -v, Lavinia:-k]             | Merge-3   | (3, 1)                     |
| 8.   | [ε:+k =d pred, praise:-v, Lavinia:-k]   | Merge-3   | (6, 7)                     |
| 9.   | [Lavinia:=d pred, praise:-v]            | Move-1    | (8)                        |
| 10.  | [Lavinia:pred, Titus:-k, praise:-v]     | Merge-3   | (9, 2)                     |
| 11.  | [s Lavinia:+v +k i Titus:-k, praise:-v] | Merge-1   | (4, 10)                    |
| 12.  | [praise s Lavinia:+k i, Titus:-k]       | Move-1    | (1)                        |
| 13.  | [Titus praise s Lavinia:i]              | Move-1    | (12)                       |
| 14.  | [Titus praise s Lavinia:c]              | Merge-1   | (5, 12)                    |

This derivation amounts to a bottom-up parse of *Titus prais es Lavinia*, which can also be represented as a upward ascent along the following derivation tree.

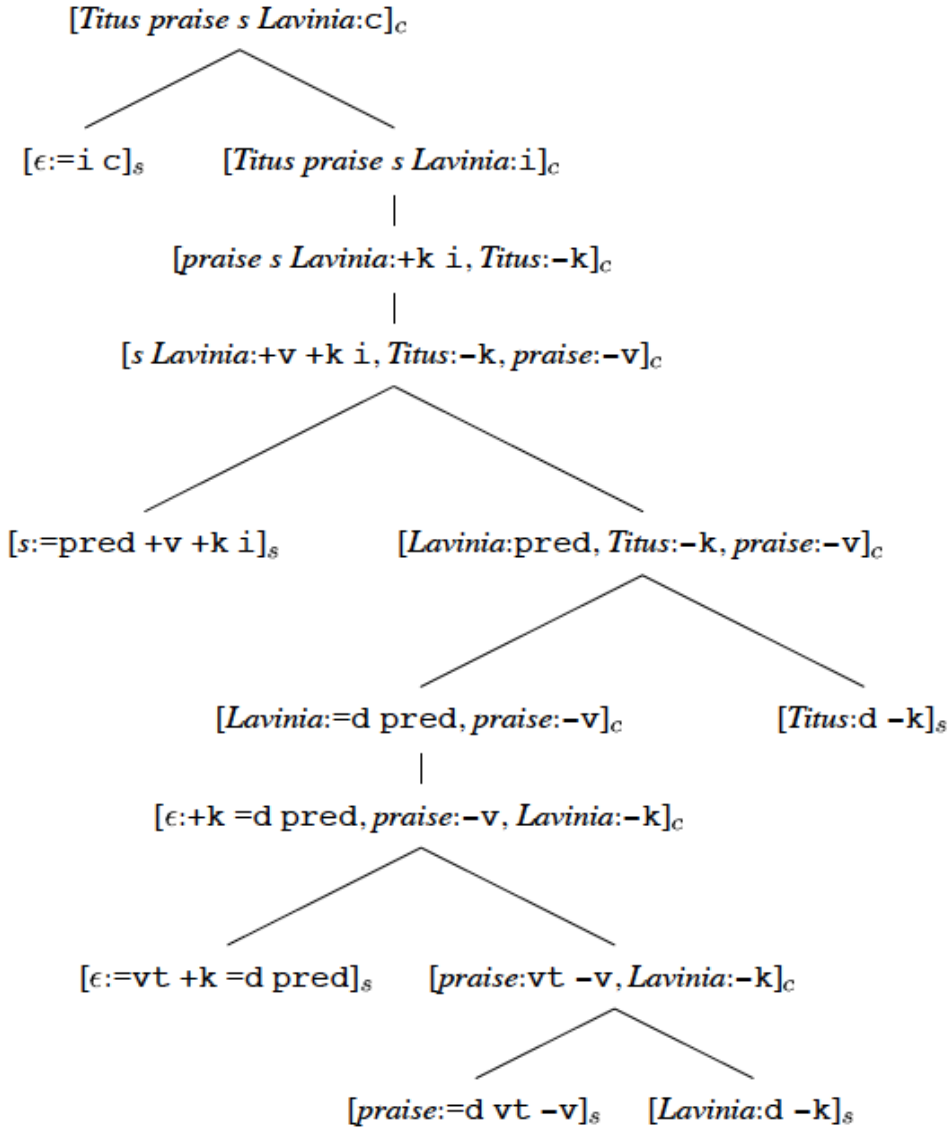


**Figure 28:** A derivation tree. Note that this is *not* a phrase marker. In this tree, every node is a structured item, which is itself representable by a tree, but is collapsed here using bracket notation (cf. fig. 27). The derivation tree represents the hierarchical order of the Merge and Move operations. As Harkema remarks in the passage quoted above, a bottom-up algorithm will traverse this tree from the leaves up, whereas a top-down algorithm will start at the top and attempt to break the input down into individual lexical items. *Source:* Harkema (2001).

Note that Harkema's proposal makes an explicit provision for various empty categories ('ε') that serve to facilitate the recognition of movement operations. For instance, the empty category that we see in step 8 of the (bottom-up) proof-theoretic derivation, where *Lavinia* moves to pick up Case, is what facilitates Case assignment. The empty category thus gives rise to the movement;

other kinds of movement are triggered by different features. Wh-movement is triggered by the need to check the  $\{-wh\}$  and  $\{+wh\}$  features that some lexical items have in their “bundles.” Notice, finally, that this gives rise to a need for the algorithm to recognize the presence of empty categories. For instance, the very first step that a top-down algorithm must make in traversing the tree in fig. 28 is to posit an empty complementizer that needs an inflectional phrase.

Harkema proves that the top-down, bottom-up, and Earley algorithms can parse Minimalist languages in polynomial time—a significant result, given the mild context-sensitivity of Minimalist grammars. Moreover, he points out (pp. 211-214) that the algorithms are compatible with a wide variety of approaches to ambiguity resolution, both in ranked parallel models (e.g., Gibson, 1991) and in serial models (e.g., Frazier, 1978). Depending on the details of the architecture, a Minimalist parser like the ones described here can be made to conform to Minimal Attachment and Late Closure (§1).

As before, the PAD framework allows the rules of the grammar to be explicitly represented and used as data. But Harkema’s presentation does not assume this. Given that the inference rules in fig. 26 can be seen as *procedural imperatives*, it seems that the algorithm can also be implemented by a hardwired circuit, whose operations carry out the inference rules directly, without needing to consult them in a separate data structure. The internal states of such a circuit would explicitly represent only its lexical input (steps 1-6 in the derivation above) and the complex tree structures in steps 7-14. Harkema’s algorithm, therefore, leaves open the question of which position psychological reality issue, (i), (ii), or (iii), we should adopt. We return to this issue in §7.

### §6.2.2 Weinberg

Steedman (1985) characterized the Competence Hypothesis as follows:

For a theory of grammar to be psychologically explanatory, it must, under the Competence Hypothesis, allow for a very direct relation of grammatical rules to the operations of a processor. Ideally, it should require nothing more to turn it into a complete theory of processing than the addition of a mechanism for local ambiguity resolution, to tell it which rule of the grammar to apply at a given point in analysis or generation. (p. 386)

In her paper, “A Minimalist Theory of Human Sentence Processing,” Amy Weinberg puts forward a proposal that seeks to implement an even stronger version of this idea. Steedman allowed for the grammar to be (psychologically) explanatory even if ambiguity resolution preferences had to be added to it in order to turn it into a processing model. Following the lead of theorists such as Pritchett (1992) and Philips (1996), Weinberg proposes to derive ambiguity resolution preferences *directly from the grammar*, thus maintaining the tightest possible connection between the grammar and the parser.<sup>94</sup> As noted in §1, if her proposal works, then it obviates the need for both resource-based parsing principles, like MA, LC, and MCP, and an excessive reliance on frequency information in a probabilistic lexicon or grammar.

Weinberg offers the following specification of a parsing algorithm:

A derivation proceeds left to right. At each point in the derivation, Merge using the smallest number of operations needed to check a feature on the category about to be attached. If Merger is not possible, try to Move within the current [unit]. If neither merger nor movement is licensed, Spell-Out the [unit]. Repeat until all terminals are incorporated into the derivation.

---

<sup>94</sup> There is a bit of irony here. Early on, Weinberg was skeptical about the Competence Hypothesis. See Berwick and Weinberg (1983, 1984). Perhaps the change of heart came with the shift from old-school transformational grammars to the new-fangled Minimalist formalism, which lends itself more easily to an interpretation as a parsing model.

This subsumes Berwick's proposal (§6.2.1), and can be implemented in the ways that Harkema suggests. The interest of Weinberg's proposal is not, however, the algorithm, but, rather, the way in which she derives well-known ambiguity resolution heuristics from the independently-motivated syntactic principles of Minimalist grammar.

Insertion or movement is governed by the Economy Conditions... The preference to attach a category using minimal structure follows immediately from this notion of Economy. At each point a category is inserted using the least number of operations necessary for feature transfer or merger. This ban on unnecessary operations subsumes Frazier and Rayner's (1982), Minimal Attachment and Gorrell's (1995) simplicity condition and has the advantage of following from independently motivated grammatical principles. (p. 289)

Moreover, the application of Economy conditions to the operation of Spell-Out (§6.1) opens up the possibility of using the grammar itself to predict well-attested garden-path phenomena, including the difficulty of reanalysis.

Following Uriagereka, we assume that Spell-Out occurs whenever a derivation [cannot proceed by the operations of *Merge* and *Move*]. The Spell-Out conditions thus also provide us with an independently motivated theory of reanalysis. If a preferred reading induces a precedence/dominance mismatch, the category that precedes but does not dominate will be spelled out. Again following Uriagereka, this means that the material inside the spelled out category is linearized and all internal syntactic structure is removed, creating a nondecomposable syntactic word. Given this, reanalysis from the preferred to dispreferred reading that requires either extraction of material from, or insertion of material into this syntactic word, will be impossible. As a lexical item, the spelled out material is an atomic unit, which can no longer be decomposed into its component pieces. If however, reanalysis occurs within a domain where Spell-Out has not applied, then material can be accessed and the preferred reading can be transformed into the dispreferred structure. Incorporating Spell-out and Economy conditions into the grammar also explains the preference for right branching derivations without the need for extra explicit principles which favor this type of derivation. ... As a grammatical operation, Spell-out is governed by Economy. Since it does not allow the checking of any features it is an operation of the last resort. As such, it will only be invoked when no other feature checking operation can apply and the minimal number of spell-outs to guarantee satisfaction of the [Linear Correspondence Axiom] will operate at each time step in the derivation. A right branching structure insures that an element that proceeds will also dominate a category and thus minimize the need for Spell-Out. Therefore, right branching structures will be preferred because they economize on the need for Spell-Out. (pp. 289-90)

The core of the idea is not new. We saw, for instance, that the Sausage Machine model of Fodor and Frazier (1978, 1980) distinguishes two processing levels, one of which (the SSS) operates on structured material that is, from its perspective, nondecomposable (§4.3). The explanation of the difficulty of reanalysis consisted in pointing out that it takes additional computational resources to “unpack” and “look inside” these structured units. Weinberg’s proposal is essentially the same, treating the output of the Spell-Out operation as precisely that kind of nondecomposable unit. What is new here is that Spell-Out is an operation of the *grammar*, not an independent aspect of the parser’s computational architecture (e.g., storage space in the memory buffers).

On the basis of the slim resources provided by the Economy conditions on Move, Merge, an Spell-Out, Weinberg’s model predicts the human attachment preferences for all the following kinds of ambiguity. (Here, I reproduce Weinberg’s categorization as well as her examples. Note that the numbering here does not reflect hers.)

**Argument/Adjunct attachment ambiguities:**

Direct object/complement subject ambiguity

(86a) The man believed his sister would win the Nobel Prize.

(86b) The man believed his sister.

Preposed object/matrix subject ambiguity

(87a) After Mary mended the socks fell off the table.

(88b) After Mary mended the socks they fell off the table.

Ditransitive/complex transitive object ambiguity

(89a) John gave the man the dog for Christmas.

(89b) John gave the man the dog bit a bandage.

### Subcategorized PP/NP modifier ambiguities

(90a) I put the book on the table

(90b) I put the book on the table into my bag.

### Main clause/relative clause ambiguity<sup>95</sup>

(91a) The horse raced past the barn fell.

(91b) The horse ridden past the barn fell.

(91c) The horse that was raced past the barn fell.

### Adjunct/Adjunct attachment ambiguities:

Adverb or particle placement:

(92a) I told Mary that I will *come* yesterday

(92b) I called to *pick* the box up

(92c) I yelled to *take* the cat out.

Normally relative clause attachments are dispreferred, but here they are the favored reading.

(93a) Although Erica hated the house she had owned it for years.

(93b) Although Erica hated the house she owned her family lived in it for years.

Suppose Weinberg is right in thinking that Minimalist Economy Conditions can absorb all of the explanatory functions of the least-effort parsing principles MA and LC. Still, there is another competing account of the HSPM's "oracle"—i.e., its ambiguity resolution principles—which appeals solely to frequency information. On this account,

---

<sup>95</sup> According to Weinberg, both Pritchett (1992) and Stevenson and Merlo (1997) have pointed out that that these types of ambiguities do not cause processing difficulties when the unergative verb 'raced' is replaced by transitive and unaccusative verbs, as in (94). Weinberg claims that her Minimalist account predicts these subtle data.

(94a) The student found in the classroom was asleep.

(94b) The butter melted in the pan was burnt.

“[p]rocessing involves factors such as the frequencies of occurrence and co-occurrence of different types of information, and the weighing of probabilistic and grammatical constraints. Our approach has suggested ... that syntactic parsing, including ambiguity resolution, can be seen as a lexical process” (MacDonald, Pearlmutter, and Seidenberg, 1994: p. 700).

To illustrate the importance of lexical frequency, consider the fact that verbs like ‘decide’ occur more frequently with sentential complements, and hence are correctly predicted to violate Minimal Attachment, as in (95).

(95) John decided the contest was fair.

Needless to say, one wants to know more about how this lexical frequency interacts with other types of frequency information. For instance, the fact that simple sentences occur more frequently in a typical corpus than sentences with embeddings conflicts with the prediction that (95) will be easy to process. Lexical and structural frequencies must be balanced in some principled way (cf. fn. 9). Still, it is plain that the HSPM makes use of frequency information. The question before us is whether that is *all* that it uses. Weinberg (1999) goes on to argue that

while speakers may very likely track frequency, *this variable works in tandem with independent grammatical constraints*. If a structure occurs very frequently in a given construction, it can influence the initial preferred analysis, but once an analysis is chosen, based on an amalgam of frequency and grammatical variables, *the grammatically driven reanalysis principles decide what will or will not be a garden path*. (p. 306)

To this end, she examines two well-known frequency-based proposals: Macdonald, Pearlmutter, and Seidenberg (1994) and Filip, Tanenhaus, and Carson (2002).

After reviewing the data in Stevenson and Merlo (1997), concerning the statistical properties of unergative, unaccusative, “ordinary,” transitive, intransitive, passive, and active verbs, Weinberg points out that

[e]ven though the class of [ordinary verbs] appears in the transitive construction more frequently than either the ergative or unaccusative class ..., it is difficult for speakers to interpret occurrences of these verbs as reduced relatives. If this is correct, it poses a problem for frequency-based approaches, which would predict that



this class should be the least difficult to interpret as it most frequently appears in the constructions that relative-clause interpretations presuppose.

In the example she provides, the frequency-based theory predicts, contrary to fact, that (96) will be easier to parse than (97) and (98). (It's actually the hardest, both experimentally and introspectively. Readers encounter trouble at the disambiguation point, which is underlined.)

- (96) The dictator *fought* in the violent coup was hated throughout the country.
- (97) The dictator *chased* in the violent coup was hated throughout the country.
- (98) The dictator *overthrown* in the violent coup was hated throughout the country.

Weinberg then turns to the frequency-based constraint-satisfaction theory in Filip, Tanenhaus, and Carson (2002), (henceforth, FTC). On their view, “the acceptability of reduced relative clauses, headed by passive participles derived from unaccusative and unergative verbs, increases when the passive participle and the main verb in a matrix clause assign their subject-NPs more proto-Patient and fewer proto-Agent properties” (p. 17). Following a proposal due to David Dowty, FTC define *proto-Patient* and *proto-Agent* as follows:

Proto-Agent: Volitional, sentient causers of the event described, who move and exist independently of the verb described (*sic*).

Proto-Patient: Undergo a change of state, serve as incremental themes, are causally affected by another event participant, are stationary relative to another event participant, and do not exist independently of the event described.

Weinberg points out that, in addition to the problem illustrated by (96)-(98), FTC's proposal is disconfirmed by the fact that (99) is a strong garden-path for most readers, despite the fact that the verbs *move* and *be afraid* assign all of the following proto-patient properties to the noun phrase *the cattle*: (i) undergo a change of state, (ii) serve as incremental themes, and (iii) are causally affected by another event participant.

- (99) The cattle moved into the crowded room were afraid of the cowboys.

FTC predict, contrary to fact, that (99) will cause no trouble for readers. Weinberg takes “[r]esults like these [to] suggest that frequency has a role to play but is filtered through grammatically justified constraints” (p. 311).

Let us sum up: Minimalist grammars are designed to be compact, making use of only the bare minimum of theoretical machinery. To this end, they posit the simple operation *Merge*, which can be made to serve as the primary operation of a parsing model. This idea has been formalized in a variety of computationally efficient algorithms due to Harkema (2001), and has been implemented in both classical computational models (Fong, 2011) and connectionist networks (Gerth and beim Graben, 2009). Most importantly, it serves as the basis of a theory of sentence processing that derives well-attested processing principles directly from the independently necessary principles of the grammar. This model thus draws on the *grammar* to explain a wide range of behavioral and neurocognitive data, including data on which frequency-based constraint-satisfaction models seem to falter.

What bearing does this have on issues to do with psychological reality and mental representation? As noted in §1, the advances in both formal syntax and parsing theory bolster the view that grammatical rules are of a suitable type to govern sentence processing. This, I think, puts positions **(i)** and **(iii)** on the psychological reality issue back into the running, at least pending further argument. The former of these is positions committed to RT, hence a strong form of psychological reality. The latter, too, puts the structure rules of language into the mind, but only by treating them as *embodied* processing rules—i.e., hardwired, procedural dispositions.

## §7. Philosophical Issues

### §7.1 *Internalism and Externalism about Representation*

Recall that Devitt (2006) argues against the representational thesis (RT), repeated here:

- (RT) A speaker of a language stands in an unconscious or tacit propositional attitude to the rules or principles of the language, which are represented in her language faculty (p. 4).

Evaluating RT requires first getting clear on what notion of representation is at issue. Devitt addresses this matter by first distinguishing the sense of ‘represent’ that he has in mind from other senses—e.g., political and theatrical representation—which are legitimate but not relevant in the present context. Having done so, he goes on to identify the relevant notion.

[T]alk of representing rules raises a question: What sense of ‘represent’ do I have in mind in RT? The sense is a very familiar one illustrated in the following claims: a portrait of Winston Churchill represents him; a sound /the President of the United States/ represents George W. Bush; an inscription, ‘rabbit’, represents rabbits; a certain road sign represents that the speed limit is 30 mph; the map on my desk represents the New York subway system; the number 11 is represented by ‘11’ in the Arabic system, by ‘1011’ in the binary system, and by ‘XI’ in the Roman system; and, most aptly, a (general-purpose) computer that has been loaded up with a program represents the rules of that program. Something that represents in this sense has a *semantic content*, a *meaning*. When all goes well, there will exist something that a representation refers to. But a representation can fail to refer; thus, nothing exists that ‘James Bond’ or ‘phlogiston’ refer to. Finally, representation in this sense is what various theories of reference—description, historical-causal, indicator, and teleological—are attempting to partly explain. (p. 5)

This sense of ‘represent’ surely has currency in cognitive science. For instance, J. A. Fodor (1983, pp. 4-8) has argued that representation, in very much this sense, is what Chomsky has in mind when he speaks of performance systems “*access[ing] the generative procedure*” and structural descriptions “*providing instructions to the performance systems*” (*The Minimalist Program*, p. 15, *emphases mine*).

To claim that Fodor and Devitt are operating with the same notion of representation—i.e., that they are both talking about the same phenomenon—is importantly different from claiming

that they agree in their *theories* of that phenomenon. Indeed, Devitt points out that, for the purposes of the debate over RT, we can stay neutral on which theory of representation is correct.

I am attempting to clarify my sense of ‘represent’ (and its cognates) by distinguishing that sense from other ordinary ones. I am not, of course, attempting a theory of representation, a theory of what ‘represent’, in that sense, refers to. The concerns of this book do not require such a theory (which is just as well given how difficult it is proving to come up with one!). (*Ignorance*, p. 6, fn. 7)

This approach seems to me appropriate; one need not defend a specific theory of every phenomenon that one mentions in affirming or denying a particular thesis. By the same token, RT should not be held hostage to one or another theory of representation.

While I am sympathetic to Devitt’s approach, I am aware that others are not. In presenting the arguments of this chapter to live audiences, I am routinely pressed to elaborate my own theory of representation, or to comment on the differences between the various theories currently on offer. The concern seems to be that my claims regarding RT either beg substantive questions or reduce to merely verbal points. To quell this sort of concern, I will offer a brief sketch of the logical space with respect to theories of representation, and make several very tentative remarks about my own theoretical commitments. My hope is that this brief discussion will be sufficient to make clear that the distinctions on which my further arguments rely—in particular, the distinction between declaratively represented rules and embodied procedures—do not beg substantive questions against other parties to the debate over RT.

Fodor’s theory of representation (Fodor, 1990; 1994) can be characterized as *fully externalist*, for it depicts representation as a particular relation between the mind/brain and various objects, properties, and events in a creature’s environment.<sup>96</sup> However, alternative theories of representation have been proposed and defended. Chomsky (2000) argues forcefully

---

<sup>96</sup> In cases of metacognition (e.g., introspection), the relation would hold between one internal state and another.

that perceptual psychologists are engaged in “internalist” or “individualist” inquiries, which deal solely with the intrinsic properties of the mind/brain, irrespective of the relations that it bears to anything in the environment. I quote Chomsky at length.

Is there a problem with internalist (or individualist) approaches to other domains of psychology? So it is widely claimed, but on dubious grounds, I think. ... [T]he study of visual perception along lines pioneered by David Marr (1982), ... has been much discussed in this connection. This work is mostly concerned with operations carried out by the retina or, loosely put, the mapping of retinal images to the visual cortex. Marr’s famous three levels of analysis—computational, algorithmic, and implementation—have to do with ways of construing such mappings. Again, the theory applies to a brain in a vat exactly as it does to a person seeing an object in motion. The latter case has indeed been studied, in work of Marr’s collaborator Shimon Ullman (1979). His studies of the determination of structure from motion used tachistoscopic presentations that caused the subject to see a rotating cube, though there was no such thing in the environment; “see,” here, is used in its normal sense, not as an achievement verb. If Ullman could have stimulated the retina directly, he would have done that; or the optic nerve. The investigation, Ullman writes, “concerns the nature of the internal representations used by the visual system and the processes by which they are derived.” The account is completely internalist. There is no meaningful question about the “content” of the internal representations of a person seeing a cube under the conditions of the experiments, or if the retina is stimulated by a rotating cube, or by a video of a rotating cube; or about the content of a frog’s “representation of” a fly or of a moving dot in the standard experimental studies of frog vision. No notion like “content,” or “representation of” figures within the theory, so there are no answers to be given as to their nature. The same is true when Marr writes that he is studying vision as “a mapping from one representation to another, and in the case of human vision, the initial representation is in no doubt—it consists of arrays of image intensity values as detected by the photoreceptors in the retina” (Marr 1982: 31)—where “representation” is not to be understood relationally, as “representation of.” ... The internalist study of language also speaks of “representations” of various kinds, including phonetic and semantic representations at the “interface” with other systems. But here too we need not ponder what is represented, seeking some objective construction from sounds or things. The representations are postulated mental entities, to be understood in the manner of a mental image of a rotating cube, whether it is the consequence of tachistoscopic presentations or a real rotating cube, or stimulation of the retina in some other way; or imagined, for that matter. Accessed by performance systems, the internal representations of language enter into interpretation, thought, and action, but there is no reason to seek any other relation to the world, as might be suggested by a well-known philosophical tradition and inappropriate analogies from informal usage. (Chomsky, 2000: pp. 158-160)

There are several ways of interpreting Chomsky’s remarks in this passage. On one

interpretation, Chomsky is claiming that there are two different notions of representation—the traditional philosophical notion (which Devitt and Fodor seem to have in mind) and a distinct notion that plays a role in cognitive science. This interpretation has its attractions. For one, it squares with the remark that “No notion like ‘content,’ or ‘representation of’ figures within the theory [of visual perception]” (*ibid*). Nevertheless, it is difficult to sustain this interpretation, in view of the fact that Fodor and Chomsky are both explicitly concerned with the internal states involved in perceptual processing—states that they agree should be characterized not only in physiological terms, but also in terms of their representational properties. The disagreement that Chomsky is registering seems, rather, to be about over what *theory* best characterizes those properties. On this interpretation, the notion of representation is shared—i.e., all parties are discussing the same phenomenon—but their theories of that phenomenon are in competition. Chomsky’s theory appears to be that the representational properties of the states involved in vision and language processing are determined *not* by the causal or nomological relations that these states bear to the environment, but *solely* by the role that such states play in internal cognitive processes. This reading squares with Chomsky’s claim that “internal representations of language enter into interpretation, thought, and action, but there is no reason to seek any other relation to the world...” (*ibid*).

Adopting the latter interpretation, we can see Fodor’s and Chomsky’s opposing views as two poles of a spectrum, the latter offering a fully internalist theory of representation and the former a fully externalist one. Between these two extremes, we can locate a number of views that attempt to accommodate the considerations that militate in favor of *both* internalism *and* externalism. One example of such “two-factor” theories can be found in Block (1986), who distinguishes between two kinds of representational property—“wide content,” which consists of

objects and properties in the environment, and “narrow content,” conceived as an inferential or computational role that determines a function from a state’s environmental context to its wide content (cf. Fodor, 1987).

Another popular “two-factor” approach takes the representational properties of only *some* states—e.g., perceptual judgments and motor commands—to be determined solely by their causal or nomological relations to the environment. Once the representational properties of these semantically basic states are fixed by (or “grounded in”) the environment, the representational properties of *non*-basic states—e.g., standing beliefs, desires, suppositions, and fears—can be accounted for by their causal or inferential relations to the basic states, which make more direct contact with the environment. One might also argue that particular *concepts*, whether used in perception, inference, or action, have their representational properties determined by causal relations to the world, while others’ are determined by inferential relations. Devitt (2006) expresses such a view in this following passage:

I think, though Fodor (1987) does not, that some [mental] words will surely not be basic but rather covered by a description theory; <bachelor> is a likely example. Meanings for these words will come from inferential associations with others; for example, <bachelor>’s association with <adult>, <unmarried>, and <male>. (p. 156)

A third type of two-factor view, which stems from the work of Wilfrid Sellars, has it that the representational properties of *every* internal state (and, by extension, every concept) are exhaustively determined by three factors: i) the environmental conditions under which it is typically elicited, ii) the causal relations it typically bears to other internal states, and iii) the behaviors that it disposes a creature to perform. Contrary to the standard versions of behaviorist, covariationist, and information-theoretic accounts of representation, Sellars believed that none of the three factors is individually sufficient; only a combination of all three can be both necessary

and sufficient for something to have the representational properties that it does.<sup>97</sup> Following familiar (though perhaps not universal) nomenclature, let us refer to this sort of view as ‘functional-role semantics’. Since this is the view that I find most attractive, I will say a bit more about it, locating it in logical space relative to the other theories mentioned above.

Unlike some of its competitors, functional-role semantics is not committed to the claim that there are two kinds of content, narrow and wide. Rather, the idea is that a state’s causal relations to the environment—both on the input and the output side—together constitute one *determinant* of its representational properties, while the internal computational or inferential role of the state constitutes the other determinant.<sup>98</sup> Thus, although both the external and the internal factors are relevant to the issue of what representational properties a given state has—like the components of a single vector—those representational properties are themselves univocal; there is no distinction between narrow and wide content.

Similarly, functional-role semantics does not draw a distinction between basic and non-basic states or concepts. The theory entails that no state or concept has its representational properties fixed *solely* by their causal or nomological relations to the environment. While it is true that perceptual judgments and motor commands enjoy more direct causal relations to the environment and to behavior than do other states, the functional-role theorist claims that such states would not have any representational properties if they entered into *only these* causal relations. On the functional-role view, in order for an internal state to have representational

---

<sup>97</sup> For a number of compelling arguments to this effect, see Brandom (1998, 2009: ch. 7).

<sup>98</sup> Neale (forthcoming) has pointed out that there is a crucial ambiguity in the word ‘determine’. Sometimes, the term has a metaphysical reading, as in “the molecular structure of this glass determines the degree of its fragility.” But, at other times, it can be used epistemologically, as in “The veterinarian determined that the goat’s ailment is curable.” The two senses of the word are close to ‘constitute’ and ‘ascertain’, respectively. In discussing what determines the representational properties of an internal state, I have in mind the metaphysical reading. However, I also believe that in ascertaining what representational properties a state has, we typically take into account external conditions, the internal role of that state vis-à-vis other states, and the behavioral consequences of being in that state.



properties at all, it must, in addition to bearing causal relations to the environment, also enter into computational or inferential relations with other internal states. Thus, even a perceptual judgment has its representational properties partly in virtue of the fact that it serves as a premise in inference and exerts a particular kind of influence on action.

An initially counterintuitive aspect of the functional-role view has been explored in detail by Sellars (1974) and, more recently, by Brandom (1998, 2009) and (in another guise) Matthews (1994, 2007). On this view, to say of some state, S, that it represents X is not merely to claim that S bears a particular relation to X. Rather, to say that S represents X is to mark that state as *belonging to a particular type*—i.e., as playing a distinctive role in the creature’s cognitive and behavioral economy.<sup>99</sup> There are, of course, many important causal and nomological relations between a creature’s psychological states and its environment. And on the functional-role view, these relations constitute two thirds of an account of those states’ representational properties. Nevertheless, no direct mind-world relation is exhaustive of a state’s having the representational properties that it does. Brandom (1998) makes this point in his discussion of the closely-related notion of reference.

No doubt, as with any other two items in the causal order, there are many relations that can correctly be said to obtain between a term tokening and what it refers to. But the present considerations show that talk about referring and referents provides no reason whatever to conclude that some one of these could be singled out as *the* reference *relation*—that unique semantically significant word-world relation in virtue of which the nonexpression is the referent of the expression. Various word-world relations play important explanatory roles in theoretical semantic projects, but to think of any one of these as what is referred to as “the reference relation” is to be bewitched by surface syntactic form.

---

<sup>99</sup> This is what Sellars (1974) means to capture with the slogan that “meaning is functional classification.” Matthews (1994, 2007) works out the idea in impressive detail, under the rubric of his “measurement theory of the attitudes.”

Brandom goes on to argue that referring is best seen as something that people (and perhaps other creatures) *do*—a kind of communicative act, not a relation between linguistic expressions and objects, properties, or events.<sup>100</sup> His target is, most likely, the popular “direct reference” theory inspired by Kripke (1980). But the functional-role view that we have been exploring extends such remarks to the notion of representation, where the contrast with Fodor’s externalist position becomes clear.<sup>101</sup> Indeed, as I have characterized it here, functional-role semantics is a middle-ground position that contrasts most starkly with the two poles of the abovementioned spectrum—Fodor’s full externalism and Chomsky’s full internalism. From the perspective of the functional-role theorist, each of these views leaves out a crucial component of what makes an internal state a representation. Let me say a word about both.

The functional-role theorist holds that full externalism does not sufficiently acknowledge the importance of the internal structures and processes that, on his view, partly determine the representational properties of, e.g., perceptual judgments. Of course, to show this conclusively requires a great deal of argument, which I cannot pursue here. But a forceful case has been made by Dennett (1969), who argues that a neural event’s being reliably elicited by, say, food is not sufficient for that neural event’s being a representation of food (pp. 76-8; cf. Dennett 1978,

---

<sup>100</sup> To be clear, no “idealism” or “anti-naturalism” follows from this position alone. Indeed, I believe that naturalism should not be held hostage to the view that any of the numerous relations that we bear to the extra-mental world in perception, thought, and behavior is usefully labeled “the reference relation.” Nor need there be such a relation for us to sustain the eminently reasonable doctrine of scientific realism. See Devitt (1997) and Chomsky (2000) for two very different defenses of these claims.

<sup>101</sup> The discussion in the main text presumes that representation as a more inclusive notion than reference. IN defense of that assumption, the following can be said: Whereas talk of reference has its home in personal-level descriptions of speech acts and perhaps propositional attitudes, extending the notion to other cases—e.g., tree rings, computer programs, maps, pictures, and subpersonal psychological states—yields awkward consequences. For instance, the following claims would not fall comfortably on untutored ears: “This tree ring refers to the age of the tree.” “This portrait refers to Plato.” “This computer program refers to the restaurants in the neighborhood.” “This (state of your) HSPM refers to a noun phrase.” By contrast, the notion of representation extends comfortably to all such cases. Ordinary-language intuitions aren’t worth much, of course, but they can, on occasion, be indicative of theoretically important distinctions.

1987). In order to be that, it must, in addition, cause the right sorts of representations downstream and, eventually, behaviors that are, in some sense, appropriate specifically to food (for that creature).<sup>102</sup> Brandom (2009: ch. 7) makes a similar point, in his discussion of the important difference between reliable *detection* and genuine *representation*.<sup>103</sup> Finally, it is worth noting that the project of naturalizing a fully externalist notion of representation has given rise to a number of theories in recent decades,<sup>104</sup> many of which are known to face potentially insuperable difficulties.<sup>105,106</sup>

---

<sup>102</sup> Dennett is, no doubt, following Wilfrid Sellars on this point, for whom so-called “language-language” and “language-exit” transitions (i.e., inference and action) were as important as “language-entries” (i.e., perception) in determining the intentional content of a mental state.

<sup>103</sup> Brandom’s position has the unfortunate feature—characteristic of his whole philosophy, incidentally—of requiring that the state in question be usable, not only as a premise, but also as a conclusion in full-blooded *inference*. Dennett, by contrast, leaves open the possibility that the downstream consequences of the state are not always personal-level processes such as inference, but are sometimes subpersonal analogs of inference. In the next section, I will argue that the states and processes involved in parsing and comprehension are subpersonal.

<sup>104</sup> The “asymmetric dependence” approach is an alternative developed by J. A. Fodor (1990). Fodor’s theory builds on the “informational semantics” introduced by Dretske (1981). Teleosemantic theories owe much of their popularity to Millikan (1984). An interestingly different version of teleosemantics is advanced by Cummins (1996).

<sup>105</sup> An interesting blog posting from David J. Chalmers (June, 2011) describes the current state of play as follows: “On the sociology, for what it is worth: my sense is that the view attributed to the three ‘Rutgers people’ in the initial post [i.e., that “it is accepted wisdom that psychosemantics was a failure; no one ever properly solved the disjunction problem let alone naturalize semantic content”] captures the received wisdom in many parts of the profession, at least concerning causal and teleological approaches to psychosemantics. Of course received wisdom is often wrong; and even the received wisdom here allows that some future approach to psychosemantics might succeed. I don’t think there’s any received wisdom as to what is missing, but certainly one strand of thinking (with which I’m sympathetic) is that any successful psychosemantics, even externalist psychosemantics, must put more weight than these programs on internal factors such as inferential role...”  
<http://philosophyofbrains.com/2011/06/20/was-psychosemantics-a-failure.aspx>

<sup>106</sup> Nearly every publication on the topic of intentional content begins with a series of objections to each of the other available theories. Cummins (1991) and Loewer (2002) provide useful summaries of several key objections to most extant views; a more up-to-date catalogue of objections can be found in Neander (2006). Godfrey-Smith (2004) captures what seems to be a widespread sentiment in the following passage: “In the 1980s the problem of giving a naturalistic theory of mental content beckoned young philosophers like myself; this looked like a philosophical problem that was both fundamental and solvable. ... Roughly twenty years on, how has the project fared? With some sadness and much caution, I suggest that things have not gone well for the Dretske-Fodor program. I doubt that we will ever see a satisfactory version of the kind of theory that Fodor’s *Psychosemantics* and Dretske’s *Explaining Behavior* tried to develop. Despite this, I do think we have learned a lot from the development of this literature. Some good partial answers may have been given to important questions—but not the exact questions that Dretske and Fodor were trying to answer. So I think it is time to start looking at different approaches to the network of

What about full internalism—the position that Chomsky articulates in passages like the one quoted above? Chomsky focuses solely on psychological processing, ignoring the role that the environment and behavior play in determining the representational properties of internal states.<sup>107</sup> His reason for doing so seems to be that the range of stimuli that gives rise to particular representation is highly variable and unsystematic, and therefore irrelevant to the scientific classification of mental phenomena. This reasoning, however, can be resisted.

First, one might argue that the range of stimuli is not as variable as Chomsky's remarks suggest. In discussing visual processing, he mentions tachistoscopic displays, direct manipulations of the retina or the optic nerve, and the endogenously-caused phenomena that we sometimes refer to as “visualizing” or “imagining.” The last of these seems importantly different from cases of perceptual judgment, hence not obviously relevant to the present discussion.<sup>108</sup> The other two cases involve *highly specific* manipulations of the mechanisms of the visual system. Not just *any* tachistoscopic presentation will elicit the same perceptual judgment, and neither will just any manipulation of the neurophysiology. The commonalities between the

---

questions surrounding belief and representation. This rethinking will involve looking again at some of the ideas of the nay-sayers of the 1980s, like Dennett and Stich, but looking further afield as well” (pp. 147-148).

<sup>107</sup> I take this to be a version of what Harman (1987) calls “solipsistic conceptual-role semantics.” Brandom (1998) calls the view “hyperinferentialism”—a position that he contrasts with his own view in the following passage: “There are three different ways in which one might take inference to be of particular significance for understanding conceptual content. The weak inferentialist thesis is that inferential articulation is *necessary* for specifically *conceptual* contentfulness. The strong inferentialist thesis is that broadly inferential articulation is *sufficient* for specifically conceptual contentfulness—that is, that there is nothing more to conceptual content than its broadly inferential articulation. ... [S]trong inferentialism as it is worked out in the rest of [*Making It Explicit*] is *not* committed to the hyperinferentialist thesis, which maintains that *narrowly* inferential articulation is sufficient for conceptual contentfulness of all sorts. [T]he broad conception includes the possibility of noninferential circumstances and consequences of application. In this way, the specifically *empirical* conceptual content that concepts exhibit in virtue of their connection to language entries in *perception* and the specifically *practical* conceptual content that concepts exhibit in virtue of their connection to language exits in *action* are incorporated into the inferentialist picture. ... The hyperinferentialist about conceptual content (adopting a position not endorsed here) would allow only *inferential* circumstances and consequences of application” (pp. 131-132).

<sup>108</sup> It is, in any event, arguable that the representational properties of endogenously-caused visual imagery can only be fully cashed out in terms of their relations to sensations, perceptual judgments, thoughts, and other phenomena, including mind-independent states of affairs and outward behavior.

stimuli and physiological interventions that elicit the same perceptual judgment are very much of interest to the vision scientist. Nothing in the passages that Chomsky quotes from Ullman or Marr seems to contravene this.

Second, suppose that we concede that the range of stimuli that elicit any perceptual judgment is simply too heterogeneous to, *by itself*, underwrite a scientifically useful account of that judgment's representational properties. Would it follow that the stimuli are strictly *irrelevant* to determining the content of the representation? I do not see that it would. The strongest conclusion that would be warranted is that our theory of perceptual representation must take into account *both* stimuli *and* factors internal to the processing mechanism (and perhaps also the downstream behavioral consequences), the latter serving to explain why such otherwise varied stimuli give rise to such uniform behavioral and psychological effects. This conclusion is, of course, precisely the claim of the functional-role semanticist.

Much more needs to be said—and, indeed, has been said—to flesh out and motivate the functional-role approach.<sup>109</sup> Again, I will not attempt to undertake that project here. I do, however, wish to forestall a common complaint against theories of representation that take inferential or computational role as a determinant of a state's representational properties. Doing so requires appealing to a distinction between personal and subpersonal psychological states, which I hope to elucidate in the following section.

---

<sup>109</sup> Different versions of functional-role semantics are discussed in Sellars (1953, 1963a, 1963b, 1974), Dennett (1969, 1978, 1987), Field (1978), Stich (1985), Jackson and Braddon-Mitchell (1996), Horwich (1999, 2006), Harman (1975, 1982, 1987, 1999), Brandom (1998, 2009), Rosenthal (2005), and Greenberg and Harman (2006).

## §7.2 *The Personal/Subpersonal Distinction*

Following Sellars (1963a), I hold that the functional-role theory of *intentional content*—one type of representational property, but certainly not the *only* type<sup>110</sup>—applies, in the first instance, to speech acts (ch. IV, §3.1.4). But the need to account for speech and rational action gives rise to the theoretical posit of internal states—the propositional attitudes. These are personal-level states, whose role in a creature’s cognitive economy is captured by a suitable formulation of folk psychology (Lewis, 1972). The folk-psychological posit of propositional attitudes suffices only for a relatively coarse-grained way of describing a creature. Digging deeper, one wants to know how a creature can so much as *have* such states. Here, the strategy of attributing *subpersonal* mechanisms becomes useful.<sup>111</sup> Focusing specifically on the case of language comprehension, we can say the following: Whereas folk psychology lets us talk about the sensation of a sound giving rise to an act of linguistic comprehension—judging that a speaker said that *p*—cognitive psychology tells us what happens *between* the sensation and the judgment.

Dennett (1987) draws a useful distinction between taking “the intentional stance”—i.e., using folk psychology to predict, explain, and describe a creature’s behavior—and adopting “the design stance,” which involves thinking of a creature as an aggregate of purposeful mechanisms, each of which has the function of performing a specialized task. It’s from the *design* stance that we isolate the HSPM and attribute subpersonal states to it. These states have *some* of the features that we take to be characteristic of personal-level propositional attitudes. In particular, they bear systematic relations to the environment, to behavior, and to one another. This is what

---

<sup>110</sup> There are other types of mental representation—e.g., sensory representation. Rosenthal (2005: ch. 7) argues, compellingly to my mind, that our account of the representational properties of purely qualitative sensations must differ from our account of the representational properties of purely intentional states.

<sup>111</sup> Theorists who maintain a firm distinction between the personal and subpersonal levels of description (though not always in those terms) include: Dennett (1969), Stich (1978), Egan (1991), and Collins (2008).

makes it both reasonable and useful to think of them as representations—i.e., to interpret them in something like the way that we interpret the diagrams in a syntax textbook. Doing so allows us to abstract away from the largely unknown neural mechanisms that underpin subpersonal states and to see the causal relations between these states as resembling the *inferential* relations that hold between propositional attitudes—as, most notably, in the PAD framework that we discussed in §2 and §§5-6. This, in turn, allows us to *rationalize* subpersonal mechanisms—i.e., to understand them as being engaged in purposeful activities and as taking reasonable cognitive steps toward accomplishing their goals. These similarities between personal and subpersonal states make it tempting to think of the latter as *propositional attitudes* with intentional content. But that would be a mistake, for there are at least eight salient differences between the two kinds of state.

First, subpersonal states are not “inferentially integrated” with personal-level states. We cannot draw personal-level inferences whose premises are the rules or principles of the grammar that we subpersonally represent (or, to use Chomsky’s neologism, “cognize”).<sup>112</sup> There may well be evidential relations between, say, the Empty Category Principle (§5.1) and the price of oil futures, but no one except a trained linguist will ever make that connection. The same goes for mental phrase markers, like [ε:+k =d pred, praise:-v, Lavinia:-k]. It’s hard to credit the idea that anyone—even exceptional linguists—reasons with these representations at the personal level.

Second, subpersonal states are *always* nonconscious, whereas personal-level states are sometimes conscious and sometimes not. For this reason, the conscious/nonconscious distinction should not be confused with the personal/subpersonal distinction. This is a pernicious error that pervades many introductory chapters in linguistics textbooks (e.g., Larson, 2010, p.

---

<sup>112</sup> See Stich (1978) for further elucidation of this point.

12). An explanation of why subpersonal states are never conscious is beyond the scope of this discussion. But, for present purposes, suffice it to say that the reasons for this are *very* different from those that are operative in cases of Freudian repression, which involves emotions and psychological defense mechanisms, and in cases of subliminal perception (e.g., in masked priming experiments), which invite still another type of explanation, having mostly to do with weakness of activation in the primary sensory cortex.

Third, subpersonal states seem to be susceptible to a computational description (§§2-6). By contrast, there are well-known and potentially insurmountable problems—loosely captured under the label “the frame problem”—for the enterprise of giving a computational description of personal-level states and processes. Recall our discussion in ch. IV (§3.1) of “Fodor’s First Law of the Nonexistence of Cognitive Science.”<sup>113</sup>

Fourth, whereas personal-level states are *paradigmatically* conceptual, there is no overwhelming reason to believe that subpersonal states are composed of *concepts*, in any sense of that vexed term. True, mental phrase markers represent linguistic stimuli *as* having one property or another.<sup>114</sup> In that sense, they *describe* rather than merely *label*, to use the language

---

<sup>113</sup> “I should like to propose a generalization; one which I fondly hope will some day come to be known as ‘Fodor’s First Law of the Nonexistence of Cognitive Science’. It goes like this: the more global (e.g. the more isotropic) a cognitive process is, the less anybody understands it.” (J. A. Fodor, 1983: p. 107). For extended arguments against the viability of a computational treatment of belief revision, see Putnam (1988) and Brandom (2008: ch. 3).

<sup>114</sup> Recall the case from J. D. Fodor (1989). In sentence (100), the noun phrase ‘the children’ is an antecedent of the NP-trace that appears after the passive verb, ‘observed’.

(100) These are the children who<sub>i</sub> *wh-trace<sub>i</sub>* were only very rarely observed *NP-trace<sub>i</sub>* in the classroom.

Fodor notes that sentences like (100) present special difficulties for the HSPM, because “the antecedents of NP-trace and PRO are typically normal argument NPs, which cannot, when they are encountered, be recognized *as* antecedents” (p. 197, emphasis on ‘as’ in the original). According to Fodor, when the passive verb in (100) is integrated into the mental phrase marker, the HSPM posits an NP-trace and proceeds to search for a suitable antecedent. The search takes time and computational resources, as reflected in a distinctive ERP signature (Featherston, 2001), as well as slower response times on cross-modal priming tasks (Nicol and Swinney, 1989) and visual probe recognition tasks (MacDonald, 1989; McElree and Bever, 1989). In order to explain the processing lags mentioned above, Fodor takes the HSPM to initially represent ‘the children’ *as* a noun phrase, but not *as* an



of Sellars (1957) and Brandom (2009).<sup>115</sup> But Sellars's and Brandom's conditions on a state's being conceptual would certainly rule out subpersonal states. (See the brief discussion of normativity below.) This is not to say that subpersonal states have "nonconceptual content," in the sense that is operative in the recent philosophical literature on sensation and perception (e.g. Kelly, 2001). For, as many have noted, discussions of *that* topic leave it strikingly unclear what the notion of nonconceptual content amounts to (Bermúdez and Cahen, 2010).

Fifth, subpersonal states have no mental attitude. Personal level states have assertoric and desiderative attitudes, which divide in subtle ways: suspicion, prediction, expectation, anticipation, and discovery are all assertoric, whereas wishing, longing, desire, need, and preference are all desiderative. (Even finer-grained distinctions can be drawn; see F. Dostoyevsky.) Subpersonal states have nothing like this richness. Psycholinguists' occasional talk of the HSPM having, e.g., frustrations, is quite amusing, but obviously fanciful. Of course, one can dig in one's heels and insist that the mental attitude of subpersonal states is always assertoric. But it's not clear what the point of this maneuver might be. Folk psychology derives

---

antecedent of the NP-trace. Moments later, the HSPM cottons on to the fact that 'the children' belongs to latter category as well. Since the stimulus expression is the same, the only way to describe this transition is as a change in the *representational content* of the HSPM's classification of that expression.

<sup>115</sup> "Classification as the exercise of reliable differential responsive dispositions (however acquired) is not by itself yet a good candidate for *conceptual* classification, in the basic sense in which applying a concept to something is *describing* it. Why not? Suppose one were given a wand, and told that the light on the handle would go on if and only if what the wand was pointed at had the property of being *grivey*. One might then determine empirically that speakers are grivey but microphones are not, doorknobs are but windowshades are not, cats are and dogs are not, and so on. One is then in a position reliably, perhaps even infallibly, to apply the *label* 'grivey'. Is one also in a position to *describe* things *as* grivey? Ought what one is doing to qualify as applying the *concept* grivey to things? Intuitively, the trouble is that one does not know what one has found out when one has found out that something is grivey, does not know what one is taking it to be when one takes it to be grivey, does not know what one is describing it *as*. The label is, we want to say, uninformative. ... What more is required? Wilfrid Sellars gives this succinct, and I believe correct, answer: 'It is only because the expressions in terms of which we describe objects, even such basic expressions as words for the perceptible characteristics of molar objects, locate these objects in a space of implications, that they describe at all, rather than merely label' ["Counterfactuals, Dispositions, and the Causal Modalities," in *Minnesota Studies in the Philosophy of Science*, vol. 2, ed. H. Feigl, M. Scriven, and G. Maxwell (Minneapolis: University of Minnesota Press, 1957), sec. 107.]" (Brandom, 2009: pp. 202-3)

a massive payoff from drawing the assertoric/desiderative distinction. What comparable payoff is there in the subpersonal case?

Sixth, subpersonal states like those involved in the early stages of language comprehension are not expressible in speech. They are, of course, *causally involved* in speech production, but that is not sufficient for being *expressed*, let alone conveyed or communicated. For a speech act to *express* an internal state, the two must have similar intentional contents, and the mental attitude of the internal state must correspond to the illocutionary force of the speech act (Rosenthal, 2005). But subpersonal states do not seem to have mental attitudes. And their representational properties are arguably not usefully assimilated to intentional content. Still, suppose we agree to slum it and endow those states with both assertoric mental attitude and some intentional content, perhaps as a kind of courtesy. It *nevertheless* remains plain that these properties of the subpersonal state will come apart, often dramatically, from the contents and illocutionary forces of the speech acts that they facilitate.

Seventh, subpersonal states are best characterized by their *natural functions*—i.e., the biological purposes for which they were selected, either genetically or during the learning period. By contrast, the possession of a great many personal-level states (e.g., thoughts about quarks) cannot be explained by a straightforward appeal to natural selection.<sup>116</sup> As Dennett makes clear, adopting the design stance involves assuming only that a system has a purpose, and that it is not currently malfunctioning. By contrast, taking the intentional stance involves making weightier

---

<sup>116</sup> Perhaps an elaboration of the teleosemantic framework (Millikan, 1984) can handle these cases. But this is far from obvious.

assumptions—e.g., that the system is a rational agent whose terms mostly refer, whose judgments are mostly true, and whose inferences are mostly good.<sup>117</sup>

Finally, there are theorists (e.g., Dennett, 1987; Brandom, 1998) who insist that representational properties are, in some sense, *inherently normative*. I am not one such, for I am not confident that we can draw a fact/norm distinction in a way that animates such claims. But suppose that one is fully committed to that sort of view. One must then say something about the *kind* of normativity that's in play, or perhaps about the naturalistic *ground* of that normativity. In the case of speech acts and personal-level, conceptual states—the propositional attitudes—the strategy of choice seems to be an appeal to community sanctions (very crudely speaking).<sup>118</sup> But should we credit the idea that community sanctions *also* ground the representational properties of subpersonal states? I think we should not.

It is implausible that human communities reward or punish their members for implementing one parsing algorithm rather than another—especially as long as both algorithms preserve, to a similar degree, the socially relevant behavioral effects. If there is normativity to the representational properties of subpersonal states, then it must, I think, be a biological kind of normativity, not a social one. The ground of this kind of normativity would have to be something about our evolutionary history (Millikan, 1984), or perhaps some notion of fitness. It is much less plausible that it will have anything to do with our “reciprocal relations of authority” to other members of the community (Brandom, 2009).

---

<sup>117</sup> I hasten to point out that honoring the distinction between the design-stance and the intentional stance—hence a distinction between personal and subpersonal levels of description—does *not* require taking on board Dennett's quasi-instrumentalism, nor his views on the indeterminacy of meaning or his appeals to principles of interpretive charity.

<sup>118</sup> For a detailed treatment, see Brandom (1998).

We have now reviewed eight reasons for distinguishing personal-level states—i.e., propositional attitudes that bear intentional contents—from the subpersonal states that include mental phrase markers and, if RT is true, mental representations of syntactic principles. I think we have strong grounds for rejecting any view that fails to honor this distinction. One very influential view that arguably falls into this category is due to J. A. Fodor. Consider, for instance, the following passage from his classic monograph, *Modularity of Mind*.

It is, perhaps, not very important to this Neocartesian story that what is innate should be, strictly speaking, *knowledge*. After all, knowledge is—or so many philosophers tell us—*inter alia* a normative notion, having much to do with the satisfaction of standards of justification. Chomsky is himself quite prepared to give up the claim that the universal linguistic principles are innately *known* in favor of the explicitly neologistic (hence sanitized) claim that they are innately “cognized.” ... It is, however, important to the Neocartesian story that what is innately represented should constitute a bona fide object of propositional attitudes; what’s innate must be the sort of thing that can be the value of a propositional variable in such schemas as ‘x knows (/believes, /cognizes) that P’. ... Here is why this is important. As previously remarked, it is the fate of the (presumed) innate information to interact with the child’s primary linguistic data, and this interaction is assumed to be *computational*. Now, the notion of computation is intrinsically connected to such semantical concepts as implication, confirmation, and logical consequence. Specifically, a computation is a transformation of representations which respects these sorts of semantic relations. (See Fodor, 1975; Haugeland, 1981.) It is, however, a point of definition that such semantic relations hold only among the sorts of things to which propositional content can be ascribed; the sorts of things which can be said to *mean that P*. The idea that what is innate has propositional content is thus part and parcel of a certain view of the ontogeny of mental capacities—viz., that in cognitive development, what is endogenously given is computationally deployed” (J. A. Fodor, 1983: pp. 5-6).

In this passage, Fodor moves directly from the claim that the states of the language faculty are *computational* to the very different claim that they are on a par with belief and knowledge, solely on the grounds that both kinds of state have representational properties. Chomsky (2000), Matthews (2007), and Collins (2008: ch. 5) expose the unpalatable consequences of this view. I will not pause to review their conclusions here.

Functionalism in the philosophy mind is arguably the reigning orthodoxy. But there are, broadly speaking, two versions of that view, and it matters for present purposes which of the two we adopt.<sup>119</sup> One version stems from the work of Sellars and Lewis, who emphasize personal-level states and folk-psychological descriptions, pitched from the intentional stance. The other, which stems from the early work of Putnam and J. A. Fodor, deals with computational states, of the sort that cognitive psychologists posit. The latter brand of functionalism has, I think, led philosophers astray, by blurring the distinction between personal and subpersonal levels of description and making it appear as though the states involved in language processing are propositional attitudes.<sup>120</sup>

One upshot of getting clear on the personal/subpersonal distinction, and the corresponding distinction between the two types of functionalism, is that doing so paves the way for a reconsideration of a common complaint against functional-role theories of representation. Functional-role semantics is sometimes rejected on the grounds that it underwrites an argument for semantic holism, which many see as having implausible consequences with regard to communication, theory change, and other phenomena (Fodor and Lepore, 1992). There is, however, a question about whether these arguments against holism about personal-level states carry over to the *subpersonal* case. In some cases, it seems that they do not. For instance, issues pertaining to the success of communication among ordinary speakers cannot arise in the subpersonal case, for the simple reason that subpersonal states are never directly expressed, communicated, or conveyed in speech. (If they were, psycholinguistics would be much easier.)

---

<sup>119</sup> The distinction I have in mind is a descendent of the one that Block (1978) drew, between “psychofunctionalism” and “analytic functionalism.” I would not, however, accept several of the claims that Block makes in the course of drawing that distinction.

<sup>120</sup> A paradigmatic case of this phenomenon is the introductory chapter in Fodor, Bever, and Garrett (1974).

Nor do any issues arise about the possibility of interpretation, agreement, or disagreement across divergent or allegedly incommensurable theories, paradigms, or conceptual schemes.

One problematic issue does, however, arise in connection with the holistic characterization of the representational properties of subpersonal states. One might reasonably wonder whether such a characterization would make it impossible for the psychologist to formulate lawlike generalizations about these properties. I do not think that the functional-role view has this consequence. First, subpersonal systems are, in all likelihood, *designed* to be as they are—i.e., they are adaptations, in the sense of Pinker (1994). It follows that they are species-wide properties, largely similar across speakers, both neurophysiologically and behaviorally. As with any biological system, there will almost certainly be exceptional cases. But the psycholinguist can freely idealize away from these, treating them as cases where *ceteris* is not *paribus*. Given that the laws of psychology are not *strict* laws, this type of idealization is perfectly appropriate. Alongside the point about adaptation, an appeal to idealization is, I believe, sufficient to quell concerns about the consequences of adopting holism in the subpersonal case. The predicates of subpersonal psychology would remain projectable and psychological laws would remain true (*ceteris paribus*).

### §7.3 Computational Psychology, Hardwired Rules, and Procedural Knowledge

Taking on board the main claims of the foregoing discussion, let us restrict our focus solely to the subpersonal level of description, and employ a view of representation as “functional classification.” We are now in a position to be more precise about the kinds of commitments that a psychologically plausible computational model might make about the representations involved in language processing. In particular, we can draw what I think is a crucial distinction between two species of psychologically reality—declarative representation and procedural embodiment.

Having done so, I will argue that we have no grounds for thinking that syntactic rules or principles are represented in the HSPM, as against being embodied.

Stabler (1983) defines a variety of claims that can be made within the framework of computational psychology. On his view, a computational system is one that goes into a physical state that represents the output of some function whenever the system is in a state that represents the corresponding input to that function.<sup>121</sup> We say that a system computes a function when this causal pattern is “regular and predictable” enough for it to be “convenient” for us to so describe it; the description is used because it is “clear and useful.” (Stabler, 1983: fn. 1.) Departing slightly from Stabler’s terminology, let us call such a function the system’s “target function.”

In Stabler’s terms, giving a **first-level theory** of a computational system involves saying what target function it computes. Such a theory does not, however, tell us *how* the system performs its computations. Seeking a more detailed description, we may, in some cases, formulate a **second-level theory**—a specification of the *program* that the system uses in computing that function. A program consists of more basic functions, whose sequential computation produces the output of the target function.

Some systems—e.g., the modern-day personal computer—compute their target functions by explicitly representing, as a separate data structure in their memory banks, the very programs that they are running. Such systems have “control states” that explicitly represent the *instructions* of a program—its processing rules, so to speak—and are causally involved in the inner workings of the machine. Each such instruction encodes one of the basic functions that, together with other basic functions, conspire to produce the output of the target function. Saying

---

<sup>121</sup> Cf. Chalmers, 1995: “A physical system implements a given computation when there exists a grouping of physical states of the system into state-types and a one-to-one mapping from formal states of the computation to physical state-types, such that formal states related by an abstract state-transition relation are mapped onto physical state-types related by a corresponding causal state-transition function” (p. 392).

which programs are explicitly *represented* in the system (rather than merely computed by it) is giving a **third-level theory**.

Importantly, not all computers are “stored-program” systems of this kind. Some are hardwired circuits, for which we can give only a second-level, not a third-level theory. Such circuits, Stabler says, “embody” a program, but do not explicitly represent it. The notion of a hardwired or embodied rule is a difficult one, so it will be helpful to consider a number of examples of systems that embody rules without representing them. We can then get a better sense of how to apply Stabler’s theoretical framework to the case of language processing.

Horgan and Tienson (1996) recommend that we think of some types of connectionist networks as hardwired circuits, susceptible only to a second-level theory. They write:

It is sometimes assumed that connectionist representations cannot exhibit syntactic structure, and that lack of syntax therefore constitutes an essential difference between connectionism and classicism (e.g., Churchland, 1995). On this view, connectionist models depart from the classical “rules and representations” conception of cognition because they eschew traditional symbolic, syntactically structured representations. But in fact some connectionist models do employ syntactically structured representations and exhibit structure-sensitive processing—although syntactic constituency in these models is not a simple part-whole relation. Examples of such models include Pollack (1990), Smolensky (1990), and Berg (1992). So connectionism need not eschew syntax—and arguably should not (Horgan and Tienson 1996). In connectionist models, rules for processing representations are not explicitly represented. It is sometimes assumed that classicism is committed to explicitly represented rules, and that lack of such rules therefore constitutes an essential difference between classicism and connectionism (e.g., Hatfield 1991). But although programs are explicitly represented as stored “data structures” in the ubiquitous general-purpose computer, stored programs are not an essential feature of the classical point of view. In some computational devices—including, for example, many hand-held calculators—the rules are all hardwired into the system and are not explicitly represented” (p. 725)

In §3.1 of *Ignorance of Language*, Devitt picks up on this discussion and claims that the rule for, e.g., addition is not explicitly represented in a simple mechanical calculator.

Think of a really simple calculator, a mechanical one. When the calculator adds it goes through a mechanical process that is governed by the rules of an algorithm for



addition. We have already noted that we cannot assume that the calculator embodies a particular rule simply because the calculator behaves as if it does. The present point concerns the rules that do, as a matter of fact, govern the operations of the calculator. Perhaps these rules operate on representations of numbers like 28 and functions like addition, but the rules themselves are not represented in the calculator. The rules are hardwired but not encoded in the calculator. And, of course, the calculator does not know that the governing rules are rules for addition, subtraction, etc. We can be quite confident about this because the calculator is not the sort of thing that can know about anything. Finally, in virtue of the calculator being governed by those rules we can say, if we like, that the arithmetical “information” that those rules reflect is embodied in the calculator even though the calculator does not represent the information. [fn. 8: Dwyer and Pietroski wonder, surprisingly, how having a mechanism in a speaker that is “correctly described by a linguistic theory is not tantamount to [the speaker’s] having the relevant beliefs” (1996: 341). It is not tantamount because the theory could correctly describe the algorithm that is embodied in the mechanism in the way that an arithmetical algorithm is embodied in the calculator.]

Finally, consider the account of “implicit knowledge” that one finds in Davies (1987, 1989, 1995). Davies elaborates this key notion as follows: A device that effects transitions between a set of inputs and their respective outputs can be said to have implicit knowledge of a rule just in case the device contains a state or a mechanism that serves as a *common causal factor* in all of those transitions. In a series of clever examples, Davies demonstrates that there are ways of satisfying this definition which are logically weaker than explicit representation (Stabler’s third-level theory) but, at the same time, *stronger* than mere “conformity” with a rule.

The internal wiring of a simple pocket calculator can provide an intuitive example of such common causal factors. The transitions between ‘7 + 1’ and ‘8’, ‘8 + 1’ and ‘9’, and so on, might all have a common causal explanation—i.e., a single feature of the calculator’s internal wiring that mediates each one of these transitions. There may be a mechanism in the circuit devoted solely to computing the function of addition,  $f(x, y) = (x + y)$ , which is activated when the ‘+’ button is pressed. A component of that mechanism has the task of incrementing by one whatever number serves as the argument to the function,  $f(x) = (x + 1)$ . Another component increments by 2 whatever number serves as the argument to function,  $f(x) = (x + 2)$ . And so on.

Applying Davies' idea that such common causal factors are cases of implicit knowledge, the whole addition mechanism is revealed as having implicit knowledge of a rule for addition. It seems that what Davies calls "implicit knowledge" is a case of what I, following Devitt (2006), have been calling "embodiment."

Turn now to the psychologist's notion of procedural knowledge. In §§3.1 and 11.5 of *Ignorance of Language*, Devitt asks us to consider a language processing systems like ACT-R, which is described here by Vasishth and Lewis (2006).

All procedural knowledge is represented as production rules—asymmetric associations specifying conditions and actions. Conditions are patterns to match against buffer contents, and actions are taken on buffer contents. All behavior arises from production rule firing; the order of behavior is not fixed in advance but emerges in response to the dynamically changing contents of the buffers. ... In summary, this sentence processing model (a) provides an integrated, quantitative account of both length and structural complexity effects in both ambiguous and unambiguous constructions; (b) predicts reading times in both ambiguous and disambiguating regions; (c) probabilistically predicts both parsing failures and reading times; and (d) provides single-parameter quantitative predictions across multiple experiments and paradigms. (pp. 414-415)

In considering this sort of model, Devitt argues that procedural knowledge does not consist in representational states.

ACT-R theory, the most recent ACT theory, "accounts for a wide range of cognitive phenomena, including perceptual-motor tasks, memory, and problem-solving" (Johnson et al. 2003: 32). The key question for us is whether this achievement requires that the production rules that are learnt be represented. Descriptions of ACT are often vague on this matter but the received view is that the rules are represented: "in ACT both declarative and procedural knowledge are represented in an explicit, symbolic form (i.e. semantic networks plus productions)" (Sun et al. 2001: 235; see also Masson 1990: 223). Yet ... since the production rules constitute procedural not declarative knowledge there seems to be no immediate and pressing need to take them as represented. Because ACT theories are based on general-purpose computer models (Anderson 1983: 2) it is perhaps not surprising that the cognitive architecture they propose involves the representation of production rules. Still, we wonder whether we should take this aspect of the model seriously if we are looking for a simulation of skills that exist in real organisms. Is there any reason to think that the IF-THEN rules that become embodied in an organism as a result of practice are represented rather than merely embodied? Perhaps we can suppose that the organism has simply learnt to respond to the working memory representation of a certain

condition with the appropriate action. Is there any explanatory gain in supposing further that it does this by representing the rule governing this response and applying it? Anderson himself remarks that “the production is very much like the stimulus-response bond” (Anderson 1983: 6). And Pinker calls it “a knee-jerk reflex” which is “triggered” (1997: 69). (*Ignorance*, p. 215)

It is important to contrast Devitt’s notion of procedural embodiment with a related but crucially different notion of procedural representation. Consider, for instance, the following passage from J. A. Fodor.

Though it gets thrown around a lot in cognitive science, the notion of a procedural representation isn’t itself transparent. At its least tendentious, however, a procedural representation is just a representation of a procedure. This is the construal that’s suggested by examples like sentence parsing. It may be that what underlies the child’s ability to assign syntactic forms to utterances is something like an algorithm for mapping sentence tokens under acoustic representation onto their structural descriptions. A grammar of the child’s language may be “explicit” in the parser, or it may be merely “implicit” in the algorithm in the sense that, whereas the latter contains “declarative” representations (like “‘the’ is a word”), the former contains “imperative” representations (like “if you find a phonological sequence #/t/ /h/ /e/#, label it a token of the word-type ‘the’.”). Notoriously, parsers and grammars needn’t be trivially interconvertible. Going in one direction, there are grammars for which parsers aren’t constructible; and, going in the other direction, there needn’t be any fact of the matter about which of an infinity of extensionally equivalent grammars a given parser realizes. So, on the construal of “procedural representation” as “representation of a procedure,” there really can be something of substance to the distinction between procedural representations and others. (Fodor, 1998: p. 136)

Fodor’s distinction is between representations that have, as it were, different mental attitudes.<sup>122</sup>

The declarative representation has an assertoric attitude—it purports to give a description of how things are; the procedural representation has an imperative attitude—it tells the system what to do in a particular circumstance. But regardless of these differences, Fodor seems to take both kinds of representation to be stored in a data structure and accessed (i.e., read) by the system—a separate processing step. Devitt’s contrast between represented and embodied rules is different.

---

<sup>122</sup> I argued above that subpersonal states do not have mental attitudes. Thus, respecting the personal/subpersonal distinction requires viewing talk of, e.g., imperative subpersonal representations as mere metaphor. As I noted in §7.2, Fodor tends to elide this distinction.

He would consider states with both assertoric and imperative mental attitudes to be representations, on account of their being stored and accessed. What is distinctive about embodied rules, on his view, is precisely that they are *not* stored in a separate data structure and *not* read by the system. Rather, they are *directly implemented*—i.e., hardwired into the architecture of the system, and carried out without further ado.

To clarify this contrast further, let us consider Devitt's example of the kingfisher—a bird that feeds on fish by diving (at an angle) and reliably catching its prey without pursuing it underwater. In order to do this by sight alone, the kingfisher has to compensate for the way that light refracts as it exits the water and enters air. Otherwise, the image of the fish would be systematically misleading as to its actual location (fig. 29).

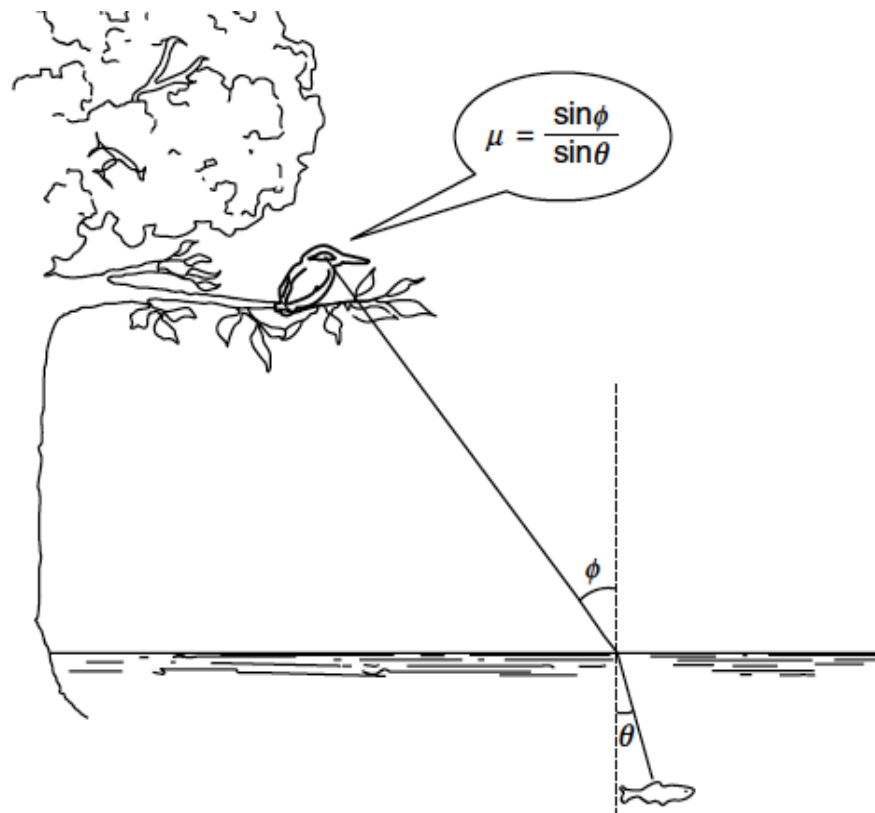


**Figure 29:** The kingfisher. *Source:* Devitt (2006), p. 48. (Original in Boden, 1984: p. 153.)

The question arises: How does the kingfisher manage this feat? Devitt writes:

[I]t would be absurd to suggest that the kingfisher goes through the following process: it represents (tacitly knows) that  $\mu$  for water-to-air refraction is such and such; it represents that the angle  $\phi$  is so and so; it represents that  $\mu = \sin \phi / \sin \theta$ ; it uses this information to calculate the angle  $\theta$ . ... Those mechanisms, reflecting information about water-to-air refraction, are simply built into the kingfisher “by nature” just as the rules for the calculator are simply built into it by an engineer” (p. 49)

Now, if we hold fast to the distinction between personal and subpersonal levels of description, or (what comes to the same thing) if we make it clear that we are describing the kingfisher from the design stance, not from the intentional stance, then it is perhaps not absurd to suggest that some subsystem within the kingfisher constructs perceptual representations of angles. But there is no obvious reason for holding that this subsystem also represents, in a declarative fashion, the value of  $\mu$  or the equation describing water-to-air refraction ( $\mu = \sin \phi / \sin \theta$ ). That is, there are no grounds for claiming that it draws on such a representation as a premise in its computations.

A more parsimonious way to characterize the kingfisher’s internal processes, while still abstracting away from neurophysiological details, is to say that a subsystem within the kingfisher embodies a certain procedure. The procedure is to token a representation of the appropriate value for the unperceived angle  $\theta$ , as a direct consequence of tokening a perceptual representation of the value of  $\phi$ . Such a procedure would be, as Devitt says, built into the very structure of kingfisher’s visuomotor system, rather than being represented in a declarative fashion in a separate data structure.

Of course, we can be more parsimonious still. Instead of positing representations of perceived and unperceived angles within the kingfisher, we could posit a direct reflex-like sensory-motor link. We would thus claim the system transforms activations in the visual cortex, occasioned by a refracted angle  $\phi$  (within some range of magnitudes), into motor neurons that directly control diving behavior that is appropriate to angle  $\theta$ . If the only behavior that we were

concerned to explain were the kingfisher's diving for fish, then this sort of account—which posits no representations whatsoever—would seem to be suitable. We would thus say that the kingfisher behaves *as if* it stores representations, though in actuality it does not. (One is reminded here of the hydraulically-driven mechanical statues that once stood in the royal French gardens—the early automata that so fascinated Descartes.)

Note that this would constitute a major disanalogy between the kingfisher's diving capacity and the case of human sentence processing. In ch. IV, I argued that the HSPM constructs mental phrase markers in order to keep track of the hierarchical syntactic relations that hold between stimuli dispersed over discontinuous regions of time. If it were not for this aspect of human sentence processing, we would be able to view the states of the HSPM as mere responses to linguistic stimuli, not as genuine information-bearing representations. In the case of the kingfisher, however, it is possible that this crucial element of the story is missing. It may be that the kingfisher's motor outputs are directly responsive, in a reflex-like manner, to its immediate sensory inputs, rather than being sensitive to a rich set of contextual factors that extend arbitrarily far back in time. In this cases, we can view its internal states as mere responses, rather than genuine representations. Whether this is so is, of course, an empirical issue. Though my own guess is that the kingfisher's behavior is not rigid enough to be susceptible to this sort of treatment, I am not in a position to pronounce on the issue.

Devitt discusses a slightly different case, involving the visuomotor skills of humans. Experiment show that, when running to catch a fly ball,

skilled fielders “ran at a speed that kept the acceleration of the tangent of the angle of elevation of gaze to the ball at 0” (McLeod and Dienes 1996: 531). It is not known how they manage this (p. 542) but they surely don't manage it by representing the algorithm for being in the right place and applying it to a series of representations of the acceleration of the tangent of the angle. (*Ignorance*, p. 50)

Surely, it is correct to say that the *fielders* don't represent the algorithm. Should we say that a

subpersonal system inside of them does? Devitt suggests that the algorithm for computing the target position is embodied in them, not represented. This seems to be the most parsimonious explanation of the data. Still, a question remains, parallel to the two raised above regarding mental phrase markers in parsing and representations of angles in the kingfisher: Does the fielder—or anything inside her—represent “the acceleration of the tangent of the angle of elevation of gaze to the ball,” or any of the mathematical determinants of that complex value? Again, our verdict will depend partly on the anatomical facts regarding whether there is a direct neural connection between the dorsal horn of visual stream and the motor areas that are responsible for coordinating the fielders’ direction of motion and gaze. It also partly depends on whether the mechanisms involved in this task are routinely recruited by a variety of other seemingly disparate tasks, and whether these mechanisms are relatively flexible in their operation. These are empirical issues whose resolution awaits further experimental research.

To round off this discussion, let us consider the phenomena that psychologists call implicit learning and implicit memory. The issues surrounding these notions are, unfortunately, somewhat murky, in part because psychologists are not in agreement about how to draw the relevant distinctions. As Devitt points out:

our view of the nature of skills depends on our view of the nature of procedural knowledge and implicit memory. Here there are lots of interesting ideas but no consensus. Indeed, our knowledge of this matter is at an early stage: “a great deal remains to be learned about the cognitive and neural mechanisms of implicit memory” (Schacter 1999: 395); ... “there is no consensus regarding the details of the dichotomies” (Sun 2003: 698). ... Axel Cleeremans (2003) refers to “as many as eleven different definitions... [of] implicit learning” (p. 491).

Davies (1995) examines the same issue and argues that the psychologist’s implicit/explicit distinction may not be helpful in addressing the general topic of what he calls implicit knowledge (which, again, seems to map neatly onto what we have called embodied rules).

[T]he distinction in experimental psychology between explicit and implicit memory

tests does not help us. In an implicit memory test “memory for a recent experience is inferred from facilitations of performance, generally known as repetition or direct priming effects, that need not and frequently do not involve any conscious recollection of the prior experience” (Schacter, 1989, p. 695). In contrast, explicit memory tests “make explicit reference to and demand conscious recollection of a specific previous experience.” (ibid.) Given that we are considering human information processing much of which is unconscious, and that we also want our notions of knowledge of rules to be applicable to systems—such as small connectionist networks—for which the question of conscious recollection does not even arise, this ... usage of “explicit” is of no help to us... (pp. 158-9)

Davies’ remarks in this passage point to a problem in the literature on implicit and explicit memory and, incidentally, also in discussions of the declarative/procedural distinction. Theorists sometimes conflate both of those distinctions with the conscious/nonconscious distinction and fail to take note of the personal/subpersonal distinction.<sup>123</sup> The moral that I draw from such considerations is that the relationships between these four distinctions (implicit/explicit, procedural/declarative, conscious/nonconscious, and personal/subpersonal) must be clarified before we can discuss their import for the issues regarding psychological reality of syntax.

Putting aside the difficult issues pertaining to implicit and explicit memory, my hope is that the examples above give us a tolerably firm grip on the notion of an embodied rule. With this notion in mind, we can now consider the case of language comprehension.

In the discussion above, we saw an example of embodiment when we considered the hardwired ATN circuit (§4). And I have hinted at how this idea can be applied in the context of the PAD framework—i.e., by seeing the grammar not as a set of (axiomatic) *premises* in the parser’s formal deductions, but as a set of *inference rules*, which the parser *directly implements*,

---

<sup>123</sup> Focusing for the moment on the declarative/procedural distinction, rather than the (possibly orthogonal) implicit/explicit distinction, we can compare two examples of declarative memory representations: (1) the factual knowledge that dogs are mammals—clearly a personal-level state, which is sometimes conscious (especially when expressed in speech), and (2) lexical memory—e.g., information about the subcategorization frames of verbs—a *subpersonal* state and hence *never* conscious. It appears that some declarative knowledge is accessible to consciousness, but some isn’t.



by moving from one mental phrase marker to another in precisely the ways that the inference rules in fact license. Devitt puts the issue as follows:

This is a distinction between two ways in which certain processing rules might be psychologically real, two ways in which the rules might be embodied in the speaker. ... [I]t is natural, although perhaps not right, to interpret Chomsky as holding RT. According to RT linguistic rules are psychologically real in that they are represented (in the sense illustrated) in the speaker. If these rules were processing rules then the representations of them would be applied in processing; they would be “read”. [The distinction we are now contemplating] draws attention to another possibility: the rules might govern processing without being represented. (pp. 45-6)<sup>124</sup>

It is on this basis that Devitt distinguishes between positions **(i)** and **(iii)** on the psychological reality of language, repeated below.

- (i)** The structure rules of a language—i.e., the syntactic rules or principles that comprise a grammar—are also processing rules that are *represented* in the speaker and *applied* in language use.
- (iii)** The structure rules of a language are also processing rules for language use but the processing rules are merely embodied in the HSPM.

Now, as Stabler (1983) points out, it is provable that any computation performed by a stored-program computer can also be performed by an assembly of hardwired circuits that don’t explicitly *represent* any instructions—a system for which can give only a second-level theory. Hence, it may turn out that the HSPM is such a device. In that case, only position **(iii)** would be warranted. Moreover, Stabler goes on to argue that we do not, at present, have any behavioral or neurophysiological data to support position **(i)**. To my knowledge, his argument is still compelling today. The past three decades of psycholinguistic research have shed a great deal of light on the nature of the human sentence processing routines, but I am aware of no experimental results or theoretical considerations that would militate decisively in favor of the idea that the rules or principles of grammar are represented in the mind, rather than being embodied in the

---

<sup>124</sup> Devitt also usefully distinguishes both of these from a case where the system merely behaves *as if* it is governed by a rule (p. 46).

form of procedural dispositions. Devitt (2006) makes use of this point in setting out his argument against the representational thesis. He writes:

According to Stabler, “no [neurophysiological] evidence has been found . . . either for linguistic processing or any other cognitive processing” governed by representations of the rules (Stabler 1983: 399). We have noted before that should a rule govern a cognitive process, it could be the case that it governs by being embodied without being represented (3.1). So, where we have evidence that a certain rule does govern, Pylyshyn’s Razor demands further evidence before we conclude that it does so by being represented; we need further evidence that the rule plays its role like a soft-wired rule in a general-purpose computer rather than like a hardwired rule in a special-purpose computer. I suggest that there is a striking lack of this further evidence with human cognitive processing. Indeed, as I shall try to demonstrate, the evidence is much more against represented rules than for them. (p. 204).

Recall that in §4 we did consider one possible source of evidence that could settle the issue of whether grammars are mentally represented or merely embodied. Suppose that we take the least-effort parsing principles, MA, LC, and MCP, as being descriptively true of the HSPM’s internal operations. Suppose also that it could be shown that no model that implements a grammar as a set of embodied procedures can account for the predictive and explanatory power of those parsing principles. In that case, we could reject all such models in one fell swoop. However, such an argument faces a number of obstacles.

While there are cogent arguments against the psychological plausibility of some models that treat grammars as procedural knowledge—specifically, Augmented Transition Networks (§4) and the Marcus Parser—such arguments may not generalize to future implementations. For instance, the Parsing as Deduction (PAD) framework allows us to view the rules of the grammar not only as axioms—i.e., declaratively represented claims that are used as explicit premises in a deduction—but also as *inference rules* that are embodied in a parsing system without being represented. In the latter sort of system, the only representations that are used as premises in a deduction are the mental phrase markers that describe past input and the lexical categorizations

of incoming material. Together, these representations give rise *directly* to a new mental phrase marker, in virtue of the system's having been hardwired to perform such transitions. Now, since the PAD framework has been shown to subsume most, if not all, parsing algorithms—the top-down Earley algorithm, the bottom-up CYK algorithm, the left-corner parsing algorithm, and the various principles-based algorithms—we must conclude that all of these are susceptible to a procedural treatment. The procedural approach to parsing has not, therefore, been ruled out. This being the case, we have, at present, no reason to believe that the HSPM houses a declarative representation of a grammar, rather than a procedural embodiment of it.

Having reached this conclusion, we must ask to what extent the psychological reality of syntactic rules or principles is threatened by the possibility that grammars serve merely as embodied procedures. Certainly, the strongest positions on the psychological reality issue—positions **(i)** and **(ii)** do not fare well. Still, position **(iii)** is left standing. So we must ask whether this position is robust enough to sustain all of the sober claims that one finds in generative linguistics. I believe that it can. For, position **(iii)** licenses the claim that formal theories of syntax can play a significant role in illuminating both the structure of a cognitive process—viz., language comprehension—and the mental representations involved in that process. Even if they are embodied in a procedural fashion, rather than represented in a declarative data structure in the human brain, syntactic rules and principles figure in predictions and explanations of behavioral phenomena that abstract away from neurophysiological detail. This, it seems to me, is all that could reasonably have been expected in advance of sustained empirical efforts to discover the inner workings of the HSPM.

We may conclude, then, on the following ecumenical note: While a wholesale skepticism concerning the psychological reality of syntactic principles is probably unwarranted,

a common claim in generative linguistics, i.e., that grammars are *represented* in the minds of competent language users, must be seen either as loose talk—a conflation of the notions of embodiment and representation—or as simply ungrounded pending further research.

#### §7.4 *A Useful Analogy*

I want to end this discussion by focusing in on the distinction between mental phrase markers (MPMs) and mental syntactic principles (MSPs). In chs. III and IV, I argued for the psychological reality of MPMs. These are states of the human parsing mechanism that are analogous to perceptual judgments, in that their function is to represent the syntactic properties of incoming linguistic stimuli. MPMs are relatively transient. Though priming data suggest that they remain active for a short time after coming into existence, their “lifespan” is typically measured in milliseconds.<sup>125</sup> In this chapter, we explored the grounds for attributing psychological reality to a *different* kind of representational state—what we might call *mental syntactic principles* (MSPs). These are standing structures—architectural features of the human parsing mechanism—that embody the rules or principles of a person’s language. When the language faculty is inactive—i.e., when no comprehension or production processes are taking place—these structures are, so to speak, *dormant*. In the jargon of contemporary metaphysics, we can say that they are *dispositional*, rather than occurrent. One way of getting a grip on this distinction between MPMs and MSPs is by focusing on an analogous distinction between perceptual judgments and beliefs. I think this analogy can be helpful in clarifying the issues at an intuitive level, so it will be useful to say a bit more about it here.

---

<sup>125</sup> I discussed structural priming in ch. III, §3. The classic experiments in this paradigm are reported in Bock (1989, 1991, 1992). For a recent review of the literature on structural priming, see Pickering and Ferreira (2008).

Philosophers sometimes apply the term ‘belief’ to *any* assertoric propositional attitude, regardless of what other properties it has. This usage elides an important distinction between *occurrent* assertoric attitudes, on the one hand, and *dispositions to have* occurrent assertoric attitudes, on the other. The dispositional/occurrent distinction is not only significant for the purposes of theory construction, but is also clearly marked in folk psychology, and frequently makes a difference to everyday practices of prediction and explanation.<sup>126</sup> Thus, on the more careful usage that I recommend here, beliefs are dispositions to have occurrent thoughts or judgments when the occasion arises, or to process various occurrent thoughts and judgments in specific ways.<sup>127</sup> By contrast, thoughts and judgments are states that we come to be in *occurrently*, either by way of perceptual processing or in the manifestation of dispositional beliefs in inference. Similar remarks go for non-assertoric states. There is distinction between standing preferences, on the one hand, and occurrent *wants* or *urges*, on the other.<sup>128</sup>

Some philosophers assume that the occurrent/dispositional distinction lines up with either the conscious/nonconscious distinction or with the personal/subpersonal distinction. Neither of these assumptions has any merit. The folk-psychological practices of predicting and explaining behavior rest on the ascription of both thoughts and beliefs to *whole persons*, not to subpersonal

---

<sup>126</sup> Q: “Why did the student get this question wrong on the math exam?” A: “He must have thought that the square root of 4 can only be a positive integer.” Q: “But if you had asked him straight out, he would have told you that it can be negative too.” A: “Yes, I’m sure he *believes* that it can, but he must have not *thought* of that at the time.”

<sup>127</sup> Notice that I am *not* proposing that belief is a disposition to express oneself in speech or in nonverbal behavior. There is a difference between being disposed to think something and being disposed to say it or to act on it. Beliefs are dispositions to have thoughts—i.e., occurrent psychological states. I do, however, hold the view that when one has an occurrent thought, one is thereby disposed—to whatever extent—to express that thought in speech and nonverbal behavior. This disposition is typically overridden by a number of factors, including a rudimentary sense of tact, etiquette, or moral decency.

<sup>128</sup> Q: “Do you prefer chocolate ice-cream or vanilla?” A: “I prefer chocolate, but I don’t want any right now because I just ate two portions of cake and I have no urge to eat any more dessert.”

mechanisms.<sup>129</sup> Moreover, intentional states can occur both consciously and nonconsciously. Intentions, thoughts, expectations, hopes, worries, fears, desires, volitions, and decisions, as well as many other more “exotic” propositional attitudes have all been shown in experimental settings to occur without a person’s metacognitive awareness *of* their occurrence (Rosenthal, 2005). Some of these are occurrent states and others are dispositional. We are not aware, for instance, of all of our beliefs at any given time, though we are able to answer questions like “Have you ever ridden an antelope?” without skipping a beat (Dennett, 1978). When the question is presented, a standing belief gives rise to an occurrent thought, which is then expressed in speech. If the question is presented in a clever masked-priming study or in a variety of other (ordinary and not-so-ordinary) conditions, the occurrent thought may still be formed, though nonconsciously. Being an occurrent state, then, is neither a necessary nor a sufficient condition for being a conscious state. And being a dispositional state is neither necessary nor sufficient for being a nonconscious state.<sup>130</sup>

Returning to the case of language processing, we can now put our analogy to use. Mental syntactic principles (MSPs) can be thought of as a subpersonal analogue of standing beliefs; mental phrase markers (MPMs) can then be seen as a subpersonal analogue of perceptual thoughts or judgments. At the personal level, we tell stories such as the following: Ever since he was a kid, Jeff has believed that spiders are dangerous, so when he just now saw several spiders

---

<sup>129</sup> And even if choose to extend our use of the term ‘belief’ to cases of subpersonal information processing, we can just as easily extend our use of ‘thought’ and ‘judgment’ in that way.

<sup>130</sup> Searle (1990) argues that only occurrent conscious states can have what he calls “aspectual shape”—a property whose ascription gives rise to an intensional-with-an-‘s’ context. Searle thinks of aspectual shape as that property of mental states in virtue of which intensional differences make a difference *to the person who is in the state*. But, as Rosenthal (1990) points out, there is no reason whatsoever to believe that only conscious states can make a difference to a person. A belief to the effect that water is water is a disposition to have a rather boring thought; a belief to the effect that water is H<sub>2</sub>O is a disposition to have a marginally interesting one. The latter belief can give rise to a nonconscious thought whose function in a person’s inferential economy is different from that of thought that water is water. This is a difference that makes a difference *to the person*.

crawl out of a box in his attic, he reasoned that there are probably more dangerous bugs in the box. In this story, the perceptual judgment regarding the presence of spiders interacts with a standing belief to yield an occurrent thought as a conclusion.<sup>131</sup> Language processing, though it takes place subpersonally, may have roughly the same structure. To illustrate, consider what happens when the HSPM encounters (101).

(101) Have the sick soldiers...

The HSPM has two standing belief-like representations to the effect that a sentence-initial auxiliary verb serves to introduce a question, and that a subsequent noun phrase will be the subject of the clause. As a result of encountering the first word of (101), it constructs a mental phrase marker that represents ‘Have’ as an inverted auxiliary verb. From this MPM and the two standing MSPs, the HSPM “concludes” that that the subsequent noun phrase will be the subject of the clause.

Given this analogy between the belief/judgment contrast at the personal level and the MSP/MPM contrast at the subpersonal level, we can address our concerns about the psychological reality of MSPs by asking the following question: Are beliefs, construed in the manner above, psychologically real? That is, can we attribute a psychological reality to a “mere procedural disposition”? Can we do so simply on the strength of the fact that it’s a disposition to give rise to an occurrent representation?

Let us imagine a scenario where we learn that there is nothing more to having the belief that red strawberries are ripe than having the disposition to infer ‘This is ripe’ *directly* from ‘This is a red strawberry’, and ‘It’s red’ *directly* from ‘It’s a ripe strawberry’, and so forth. I think we

---

<sup>131</sup> We can imagine two accounts of how a perceptual judgment interacts with a belief. The first account has it that the interaction is direct and unmediated. The second has it that the belief gives rise to an occurrent thought with the same intentional content, which, in turn, interacts directly with the perceptual judgment. Which of these accounts one prefers will depend on the details of one’s theory of cognitive architecture.

would still grant that the standing state underlying this cluster of dispositions is psychologically real. The “procedural” nature of this state does not hamper our ability to explain and predict a person’s behavior by positing it. And given that we have, as yet, no way of identifying the state in neurophysiological terms, it seems that the predictive and explanatory leverage that we derive from classifying it in intentional or semantic terms is enormous. Moreover, when we do eventually look at the neurophysiological details, we expect—in accordance with the view of embodiment that I endorsed above—to find that there is a common causal factor in the brain that accounts for the two aforementioned inferences, and others. For these reasons, I see no compelling grounds for resisting the ascription of psychological reality to the dispositional states of belief. Similar considerations can be adduced for the case of standing preferences.

These remarks carry over, I suggest, to the subpersonal case. Mental syntactic principles may well turn out to be nothing more than embodied procedural knowledge—a systematic disposition to move directly from one type of mental phrase markers to another. This would not, however, impugn their explanatory value as theoretical posits.

### *§7.5 Summary and Conclusion*

In §7.1, I discussed several competing views of representation. The major lines of division, I suggested, pertain to the role of internal processing factors (computational and inferential role) and external environmental factors (causal and nomological mind-world relations). On a fully externalist view, like the one championed by Fodor (1990, 1994), internal factors do not determine the representational properties of psychological states. At the opposite end of the spectrum, we find fully internalist views like those of Chomsky (2000), which deny that causal and nomological relations to environment have any bearing on the representational properties of internal states. Between these two poles, a variety of views have been developed



and defended. Of these mixed views, the one for which I have the greatest sympathy is functional-role semantics, as developed by Sellars, Harman, Brandom, and others. I went on to explore the core tenets of this view and to contrast it with other positions in logical space. Finally, I suggested that there is no overwhelming reason to adopt either a full internalism or a full externalism.

I began §7.2 by listing a number of important differences between personal and subpersonal levels of description. Unlike personal-level states, subpersonal states are not inferentially integrated, never conscious, never expressed in speech, lack mental attitude, are not composed of concepts, and do not resist a teleological and computational treatment. I went on to relate the personal/subpersonal distinction to two others—viz., the distinction between Dennett’s intentional and design stances, and the distinction between the sort of functionalism propounded by David Lewis and the sort that has been popularized by Jerry Fodor. I suggested that the latter sort of functionalism has given rise to a conflation between subpersonal and personal-level states. One upshot of this conflation is that a set of arguments commonly invoked against functional-role semantics is often not evaluated in light of the personal/subpersonal distinction. At the close of §7.2, I provided reasons for thinking that these arguments, whatever their force in the personal-level case, do not militate against a functional-role treatment of the representational properties of subpersonal states.

For the purposes of the present project, a crucial distinction must be drawn between two ways in which a set of rules or principles might be psychologically real. In §7.3, I sketched the theoretical framework of Stabler (1983), who distinguishes between computational systems in which rules are declaratively represented and ones in which those very rules are “hardwired” or embodied. Drawing extensively on Devitt (2006), I surveyed several examples of the latter type

of system, including simple mechanical calculators, some neural networks, and the visuomotor systems of the kingfisher and the skilled outfielder. These examples served to clarify the represented/embodied distinction, and to differentiate it from the distinction that Fodor (1998) draws between declarative and procedural representations. In the course of that discussion, I also appealed to a view developed by Davies (1987, 1989, 1995), according to which a rule is embodied in a system just in case there is a common causal factor within the system that mediates between the rule's domain and its range.

With this distinction between embodied and represented rules in place, I went on to examine how best to characterize the role of syntactic rules or principles in human sentence processing. Following Devitt (2006), I claimed that there is no argument against the view that grammars are embodied in the HSPM, rather than explicitly represented. While some procedural models of sentence processing have been ruled out, many more remain. It is, at present, an open empirical issue whether such models can be made to satisfy the well-confirmed least-effort parsing principles, Minimal Attachment, Late Closure, and the Minimal Chain Principle. Pending an argument to the effect that such models cannot, even in principle, account for the behavioral data garnered from psycholinguistic experiments, we must tentatively conclude that the view that grammars are mentally represented is without adequate support.

To assess the consequences of this conclusion for the issue of the psychological reality of language, I went on in §7.4 to discuss the distinction between occurrent and dispositional propositional attitudes—e.g., datable judgments and urges vs. standing beliefs and preferences. I suggested that, at the subpersonal level, we find an analogous distinction between mental phrase markers and mental syntactic principles. With this analogy in mind, I went on to ask what would follow from a discovery that beliefs are nothing more than the reliable inferential

dispositions to move procedurally from one occurrent thought or judgment to another (e.g., from “Lightning is seen now” to, e.g., “Thunder will be heard soon,”). I suggested that it would not follow from this that beliefs lack psychological reality. We would continue to think of beliefs as useful, explanatory posits that illuminate the contents of a creature’s mind and, consequently, figure in our accounts of its outward behavior. By analogy, I think that we should be willing to ascribe psychological reality to the states of the HSPM that embody syntactic rules or principles. Doing so underwrites the legitimacy of psycholinguistic explanations, of the kind that we surveyed in chs. III-V. In this sense, then, we can conclude that embodied grammars are indeed psychologically real.

## REFERENCES

- Abney, S. P. (1991). "Parsing by Chunks," in *Principle-Based Parsing*, Berwick, Abney, and Tenny (eds.), pp. 257-278.
- Abney, S. P. (1996). "Statistical Methods in Linguistics," in *The Balancing Act*, Judith Klavans and Phillip Resnik (eds.), MIT Press: Cambridge
- Abney, S. P. and Johnson, M. (1991). "Memory Requirements and Local Ambiguities of Parsing Strategies," *Journal of Psycholinguistic Research*, Vol. 20, 3, pp. 233-250.
- Aho, A. V. and Ullman, J. D. (1972). *The Theory of Parsing, Translation, and Compiling*, Vol. 1. Prentice-Hall.
- Antony, L. and Hornstein, N., eds. (2003). *Chomsky and His Critics*, Blackwell: Malden, MA.
- Armstrong, D. M. (1968). *A Materialist Theory of the Mind*, London: Routledge.
- Auer, P. and Wei, L., eds. (2007). *Handbook of Multilingualism and Multilingual Communication*, Berlin: Mouton de Gruyter.
- Aydede, M. (2010). "The Language of Thought Hypothesis," in *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/entries/language-thought/>
- Bader, M, Meng, M., and Bayer, J. (2000). "Case and Reanalysis," *Journal of Psycholinguistic Research*, Vol. 29, No. 1. pp. 37-52.
- Balog, K. (1999). "Conceivability, possibility, and the mind-body problem," *Philosophical Review* 108, pp. 497-528.
- Barber, A., ed. (2003). *Epistemology of Language*, Oxford University Press: New York
- Barber, A. (2004). "Idiolects," in *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/entries/idiolects/>.
- Bates, M. (1978). "The theory and practice of augmented transition network grammars," in *Natural Language Communication with Computers*, Leonard Bolc (ed.), Springer-Verlag: Berlin.
- Bechtel, W. (1987). "Psycholinguistics as a Case of Cross-Disciplinary Research: A Symposium Introduction," in *Synthese*, 72 (3), pp. 293-311.
- Bechtel, W. and Abrahamsen, A. (2002). *Connectionism and the Mind: Parallel Processing*,

*Dynamics, and Evolution in Networks* (2nd edn.). Oxford: Blackwell.

Bechtel, W. and Abrahamsen A. (2006). "Phenomena and Mechanisms: Putting the Symbolic, Connectionist, and Dynamical Systems Debate in Broader Perspective," in *Contemporary Debates in Cognitive Science*, Stainton (ed.), pp. 159-185

Berthouzoz, C. and Merlo, P. (1997). "Statistical Ambiguity Resolution for Principle-Based Parsing," in *Proceedings of the Recent Advances in Natural Language Processing*, pp. 179-186. Available at the following URL: <http://www.latl.unige.ch/doc/ranlp97.ps>

Bermúdez, J., & Cahen, A. (2010). "Nonconceptual mental content," In E. N., Zalta, (Ed.), *The Stanford Encyclopedia of Philosophy*.

Berwick, R.C. and Weinberg, A.S. (1982). "Parsing efficiency, computational complexity, and the evaluation of grammatical theories," *Linguistic Inquiry* 13, pp. 165-192.

Berwick, R.C. and Weinberg, A.S. (1983). "The role of grammars in models of language use," *Cognition*, 13, pp. 1-62.

Berwick, R.C. and Weinberg, A.S. (1984). *The Grammatical Basis of Linguistic Performance*, MIT Press, Cambridge, MA.

Berwick, R.C. (1991a). "Principles of Principle-based Parsing," in *Principle-Based Parsing: Computation and Psycholinguistics*, Berwick, R. C., Abney, S. P., and Tenny, C. eds., Dordrecht: Kluwer Academic Publishers, pp. 1-37.

Berwick, R.C. (1991b). "Principle-Based Parsing," in P. Sells, S.M. Shieber, and T. Wasow, editors, *Foundational Issues in Natural Language Processing*. Kluwer Academic Publishers, pp. 115-226.

Berwick, R.C. (1997). "Syntax Facit Saltum: Computation and the Genotype and Phenotype of Language," *Journal of Neurolinguistics*, Vol. 10, No. 213, pp. 231-249.

Berwick, R.C., Abney, S.P., and Tenny, C., eds. (1991). *Principle-Based Parsing: Computation and Psycholinguistics*. Dordrecht: Kluwer Academic Publishers.

Berwick, R.C. and Fong, S. (1995). "A Quarter Century of Computation with Transformational Grammar." In J. Cole, G.M. Green, and J.L. Morgan, editors, *Linguistics and Computation*. Kluwer Academic Publishers.

Berwick, R.C. and Niyogi, S. (2005). "A Minimalist implementation of Hale-Keyser incorporation theory," in DiSciullo, Anna Maria (ed.) *UG and External Systems, Language, Brain, and Computation*, Amsterdam: John Benjamins Publishing Co., pp. 269

- Bever, T.G., Carroll, J.M., and Miller, L.A., eds. (1984). *Talking Minds: The Study of Language in Cognitive Science*. MIT Press, Cambridge, MA.
- Black, M. & Chiat, S. (1981). "Psycholinguistics without 'psychological reality'," in *Linguistics*, 19, pp. 37-61.
- Block, N. (1978). "Troubles with Functionalism," *Minnesota Studies in the Philosophy of Science*, 9, pp. 261-325.
- Block, N. (1986). "Advertisement for a Semantics for Psychology," in French, P. A., *et. al.*, eds., *Midwest Studies in Philosophy, Vol. X*. University of Minnesota Press: Minneapolis, pp. 615-678.
- Block, N. (1996). "On A Confusion About a Function of Consciousness," in *Behavioral and Brain Sciences*, 18, pp. 227-247, Cambridge University Press.
- Bock, J. K., & Kroch, A.S. (1989). "The Isolability of Syntactic Processing," in G. N. Carlson & M. K. Tanenhaus (eds.), *Linguistic structure in language processing*, Dordrecht, The Netherlands: Kluwer, pp. 157–196.
- Bock, K., & Loebell, H. (1990). "Framing sentences." in *Cognition*, 35, pp. 1–39.
- Bock, K., Loebell, H., & Morey, R. (1992). "From conceptual roles to structural relations: Bridging the syntactic cleft," in *Psychological Review*, 99, pp. 150–171.
- Bod, Hay, & Jannedy, eds. (2003). *Probabilistic linguistics*, MIT Press: Cambridge, MA.
- Bornkessel-Schlesewsky, I and Schlewsky, M. (2009). *Processing Syntax and Morphology, A Neurocognitive Perspective*, Oxford University Press.
- Braddon-Mitchell, D. and Jackson, F. (1996). *Philosophy of Mind and Cognition*, Blackwell Publishers.
- Brandom, R. (1998). *Making It Explicit*. Harvard University Press: Cambridge, MA.
- Brandom, R. (2001). *Articulating Reasons*. Harvard University Press: Cambridge, MA.
- Brandom, R. (2008). *Between Saying and Doing*. Oxford University Press: New York
- Brandom, R. (2009). *Reason in Philosophy*. Harvard University Press: Cambridge, MA.
- Bresnan, J. (1978). "A Realistic Transformational Grammar," in *Linguistic Theory and Psychological Reality*, M. Halle, J. Bresnan, G. A. Miller (eds.), Cambridge, MA: MIT Press.
- Bresnan, J. (2001). *Lexical-functional Syntax*, Malden, MA: Blackwell Publishers.

Burge, T. (1979). "Individualism and the Mental," *Midwest Studies in Philosophy*, vol. IV, P. A. French, T. E. Uehling Jr., H. K. Wettstein, eds., Minneapolis: University of Minnesota Press. Reprinted in *The Nature of Mind*, ed. David Rosenthal. New York: Oxford University Press, 1991, pp. 536-567.

Byrne, A. and Logue, H., eds. (2009). *Disjunctivism: Contemporary Readings*. Cambridge, MA: MIT Press.

Carruthers, P. (2000). *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: Cambridge University Press.

Carruthers, P. (2006). "The Case for Massively Modular Models of the Mind," in *Contemporary Debates in Cognitive Science*, R. Stainton (ed.), Blackwell: Malden, MA, pp. 3-21.

Cain, M.J. (2010). "Linguistics, Psychology and the Scientific Study of Language," *dialectica* Vol. 64, N° 3, pp. 385-404.

Carnie, A. (2010). *Constituent Structure* (2<sup>nd</sup> edition). Oxford University Press.

Carter, M. (2007). *Minds and Computers: An Introduction to the Philosophy of Artificial Intelligence*, Edinburgh University Press, chs. 14, 16.

Chalmers, D. (1990). "Syntactic transformations on distributed representations," *Connection Science*, 2, pp. 53-62

Chalmers, D. (1996). *The Conscious Mind*. New York: Oxford University Press.

Chalmers, D. (1993) "Connectionism and compositionality: Why Fodor and Pylyshyn were wrong," *Philosophical Psychology* 6, pp. 305-319

Chalmers, D. (1993/forthcoming). "A Computational Foundation for the Study of Cognition," forthcoming in a special issue of the *Journal of Cognitive Science*, G. Piccinini (ed.), available at the following URL: <http://consc.net/papers/computation.html>

Chalmers, D. 1995. "On Implementing a Computation." *Minds and Machines* 4: pp. 391- 402.

Charniak, E., Goldwater, S., Johnson, M. (1998). "Edge-Based Best-First Chart Parsing," available at the following URL: <http://www.aclweb.org/anthology-new/W/W98/W98-1115.pdf>

Chesi, C. (2004). *Phases and cartography in linguistic computation: Toward a cognitively motivated computational model of linguistic competence*. PhD thesis, Università di Siena.

Chesi, C. (2007a). "An introduction to phase-based minimalist grammars: Why *move* is top-down from left-to-right," *Studies in Linguistics (CISCL WP)*, 1, 49-90.

Chesi, C. (2007b). "Five reasons for building phrase structures top-down from left to right," in *Nanzan Linguistics*, Special Issue 3.

- Chomsky, N. (1957/2002). *Syntactic Structures*, 2<sup>nd</sup> ed. De Gruyter Mouton.
- Chomsky, N. (1959). "Review of *Verbal Behavior*. By B. F. Skinner" *Language*, 35, pp. 26-58.
- Chomsky, N. (1961/1964). "Degrees of Grammaticalness," in *The Structure of Language: Readings in the Philosophy of Language*, Fodor and Katz (eds.). Excerpted from "Some Methodological Remarks on Generative Grammar," in *Word*, 17, 1961, pp. 219-39.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*, Cambridge, Mass.: MIT Press.
- Chomsky, N. (1967). "Reply to Harman," in *Language and Philosophy: A Symposium*, Hook, S., (ed.), New York: New York University Press.
- Chomsky, N. (1969a). "Linguistics and Philosophy". In Hook 1969, pp. 51-94.
- Chomsky, N. (1969b). "Comments on Harman's Reply". In Hook 1969, pp. 152-9.
- Chomsky, N. (1972). *Language and Mind*. Enlarged edition. New York: Harcourt Brace Jovanovich.
- Chomsky, N. (1975a). "Knowledge of Language," in *Minnesota Studies in the Philosophy of Science, Volume VII: Language, Mind, and Knowledge*, Gunderson (ed.), pp. 299-320.
- Chomsky, N. (1975b). *Reflections on Language*. New York: Pantheon Books.
- Chomsky, N. (1975c). "Quine's Empirical Assumptions," in *Words and Objections: Essays on the Work of W. V. Quine*, Davidson, D. and Hintikka, J., (eds.), Reidel: Dordrecht, pp. 53-68.
- Chomsky, N. (1977). *Essays on Form and Interpretation*, Amsterdam: Elsevier North Holland
- Chomsky, N. (1978) "On the Biological Basis of Language Capacities," in G. A. Miller and E. Lenneberg (eds.) *Psychology and Biology of Language and Thought*. New York: Academic Press.
- Chomsky, N. (1980a). *Rules and Representations*. New York: Columbia University Press.
- Chomsky, N. (1980b). "Rules and Representations". *Behavioral and Brain Sciences* 3: 1-14.
- Chomsky, N. (1980c). "Author's Response" to peer commentary on 1980b. *Behavioral and Brain Sciences* 3. pp. 42-58.
- Chomsky, N. (1981). *Lectures on Government and Binding*, Dordrecht: Foris Publications.
- Chomsky, N. (1986). *Knowledge of Language*. New York: Praeger.



Chomsky, N. (1988) "Language and Problems of Knowledge," in *The Philosophy of Language*, 4<sup>th</sup> ed., A. P. Martinich (ed.), ch. 39.

Chomsky, N. (1995). *The Minimalist Program*, Cambridge, Mass.: MIT Press.

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*, Cambridge, Mass.: Cambridge University Press.

Chomsky, N. (2001). "Derivation by Phase," in *Ken Hale: A Life in Language* (Current Studies in Linguistics 36), Michael Kenstowicz (ed.), pp. 1–52. Cambridge, MA: MIT Press.

Chomsky, N. (2003). "Reply to Rey," in Antony and Hornstein (eds.), pp. 274-87.

Churchland, P. (1988). *Matter and Consciousness*, Cambridge, Mass.: MIT Press.

Churchland, P. M. (1985/1989) "The Ontological Status of Observables: In Praise of Super-Empirical Virtues," in *Images of Science*, Churchland, P. M. and Hooker, C. A. (eds.). Chicago: University of Chicago Press; reprinted as ch. 8 in *A Neurocomputational Perspective: The Nature of the Mind and the Structure of Science*, Cambridge, Mass.: MIT Press.

Clapin, H. (2002). *The Philosophy of Mental Representation*. Oxford: Clarendon Press.

Clapin, H. Staines, P. and Slezak, P. (eds.) (2004). *Representations in Mind: New Approaches to Mental Representation*, Elsevier Publishing.

Clark, H. H. and Clark, E. V. (1977). *Psychology and Language*. Harcourt Brace Jovanovich.

Clifton, C., Traxler, M.J., Mohamed, M.T., Williams, R.S., Morris, R.K. and Rayner, K. (2003). "The use of thematic role information in parsing: Syntactic processing autonomy revisited," *Journal of Memory and Language*, 49, pp. 317–334

Coltheart, M. (1999). "Modularity and cognition." *Trends in Cognitive Sciences*, 3, pp. 115-120.

Collins, J. (2007). "Review of *Ignorance of Language*," *Mind*, Vol. 116, 462, pp. 416-423.

Collins, J. (2008a). *Chomsky: A Guide for the Perplexed*, Continuum Press.

Collins, J. (2008b). "Knowledge of Language Redux," in *Croatian Journal of Philosophy* 7, pp. 3-42.

Collins, J. (2009). "A Question of Irresponsibility: Postal, Chomsky, and Gödel," *Biolinguistics*, Vol. 3, No 1, pp. 99-103.

Crocker, M.W. (1996). *Computational Psycholinguistics*. Dordrecht: Kluwer.

Crocker, M.W. (1999). "Mechanisms for Sentence Processing," in S. Garrod and M. Pickering, eds., *Language Processing*, Taylor and Francis.

- Crocker, M.W. and Brants, T. (2000a). "Probabilistic Parsing and Psychological Plausibility," in *Proceedings of the 18th conference on Computational linguistics - Volume 1*, available at:
- Crocker, M.W. and Brants, T. (2000b). "Wide-Coverage Probabilistic Sentence Processing," *Journal of Psycholinguistic Research*, Vol. 29, 6, pp. 647-669.
- Crocker, M.W., Pickering, M., and Clifton C., eds. (2000). *Architectures and Mechanisms for Language Processing*, Cambridge University Press.
- Cuetos, F. and Mitchell, D. (1988). "Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish," *Cognition*, 30, pp. 73-105.
- Culbertson, J., and Gross, S. (2009). "Are Linguists Better Subjects?," *British Journal for the Philosophy of Science*, 60, pp. 721-36.
- Culbertson, J., and Gross, S. (forthcoming). "Revisited Linguistic Intuitions," *British Journal for the Philosophy of Science*.
- Cummins, R. (1985). *The Nature of Psychological Explanation*, Cambridge, Mass.: MIT Press.
- Cummins, R. (1991). *Meaning and Mental Representation*, Cambridge, Mass.: MIT Press.
- Cummins, R. (1996a). *Representations, Targets, and Attitudes*, Cambridge, Mass.: MIT Press.
- Cummins, R. (1996b). "Systematicity," *Journal of Philosophy*, Vol. 93, No. 12., pp. 591-614.
- Cummins, R. (2010). *The World in the Head*, Oxford University Press
- Davidson, D. and Hintikka, J., eds. (1975). *Words and Objections: Essays on the Work of W. V. Quine*, Springer.
- Davies, M. (1987). "Tacit Knowledge and Semantic Theory: Can a Five per cent Difference Matter?" *Mind*, 96, pp. 441-62.
- Davies, M. (1989). "Tacit Knowledge and Subdoxastic States," in *Epistemology of Language*, A. George (ed.), pp. 131-52.
- Davies, M. (1995). "Two Notions of Implicit Rules." In *Philosophical Perspectives*, 9, *AI, Connectionism, and Philosophical Psychology*, James E. Tomberlin (ed.). Cambridge MA: Blackwell
- Davis, S. and Gillon, B., eds. (2004). *Semantics: A Reader*, New York: Oxford University Press.
- Delmonte, R. (2005). "Deep & Shallow Linguistically Based Parsing: Parameterizing Ambiguity in a Hybrid Parser," in DiSciullo, Anna Maria (ed.), *UG and External Systems, Language, Brain, and Computation*, Amsterdam: John Benjamins Publishing Co..

De Mulder, H., Everaert, M., Nilsen, O, Lentz, T., and Zondervan, A., eds. (2010). *The Linguistic Enterprise: From Knowledge of Language to Knowledge in Linguistics*, Amsterdam: John Benjamins.

Dennett, D. C. (1969). *Content and Consciousness*. London: Routledge & K. Paul.

Dennett, D. C. (1978). *Brainstorms: Philosophical Essays on Mind and Psychology*, Cambridge: MIT Press

Dennett, D. C. (1987). *The Intentional Stance*, Cambridge, Mass.: MIT Press.

DeVincenzi, M. (1991). "Filler-gap dependencies in a Null Subject Language: Referential and Non-Referential WHs," *Journal of Psycholinguistic Research*, vol. 20, no. 3, pp. 197-213.

Devitt, M. and Sterelny, K. (1989). "What's Wrong with 'the Right View'," in *Philosophical Perspectives, 3: Philosophy of Mind and Action Theory*, James E. Tomberlin (ed.). Atascadero: Ridgeview Publishing Company, pp. 497-531.

Devitt, M. (1996). *Realism and Truth*. 2<sup>nd</sup> ed., Princeton University Press.

Devitt, M. (1998). "Naturalism and the A Priori," *Philosophical Studies*, 92, pp. 45-65.  
Reprinted in Devitt, 2010b.

Devitt, M. (2002). "Moral Realism: A Naturalistic Perspective," *Croatian Journal of Philosophy*, 2, pp. 1-15.

Devitt, M. (2003). "Linguistics is Not Psychology," in *The Epistemology of Language*, Barber (ed.), OUP, pp. 107-139.

Devitt, M. (2005). "There is No A Priori," in *Contemporary Debates in Epistemology* Sosa and Steup, eds., Cambridge, MA: Blackwell Publishers, pp. 105-115.

Devitt, M. (2006). *Ignorance of Language*. Oxford: Oxford University Press.

Devitt, M. (2006b). "Defending *Ignorance of Language*: Responses to the Dubrovnik Papers". *Croatian Journal of Philosophy*, 6, pp. 571-606.

Devitt, M. (2008). "Explanation and Reality in Linguistics," *Croatian Journal of Philosophy*, Vol. VIII, No. 23, pp. 203-231.

Devitt, M. (2008b). "A Response to Collins' Note on Conventions and Unvoiced Syntax," *Croatian Journal of Philosophy*, 8, pp. 249-55.

Devitt, M. (2008c). "Methodology in the Philosophy of Linguistics," in *Australasian Journal of Philosophy*, 86, pp. 671-84.

Devitt, M. (2009). "Psychological Conception, Psychological Reality: A Response to Longworth and Slezak." *Croatian Journal of Philosophy*, 9, pp. 37-46

Devitt, M. (2010a). "Linguistic Intuitions Revisited," *British Journal for the Philosophy of Science*, 61, pp. 833-65.

Devitt, M. (2010b). *Putting Metaphysics First: Essays on Metaphysics and Epistemology*. Oxford: Oxford University Press.

Devitt, M. (2010c). "What 'Intuitions' are Linguistic Evidence?" *Erkenntnis*, 73, pp. 251-64.

Devitt, M. (2011a). "Linguistic Knowledge," in *Knowing How: Essays on Knowledge, Mind and Action*, J. Bengson and M. Moffett (eds.), Oxford: Oxford University Press, pp. 314-33.

Devitt, M. (2011b). "Methodology and the Nature of Knowing How," *Journal of Philosophy*, CVIII, pp. 205-18.

DiSciullo, Anna Maria (ed.) (2005). *UG and External Systems, Language, Brain, and Computation*, Amsterdam: John Benjamins Publishing Co.

Dowty, D. R., Karttunen, L. and Zwicky, A.M., eds. (1985). *Natural Language Processing: Psychological, Computational, and Theoretical Perspectives*, Cambridge University Press: Cambridge, England.

Dretske, F. (1981). *Knowledge and the Flow of Information*, MIT Press: Cambridge, MA

Dretske, F. (1986). *Explaining Behavior*. MIT Press: Cambridge, MA

Dretske, F. (2006). "Perception without Awareness," in Tamar S. Gendler & John Hawthorne (eds.), *Perceptual Experience*. Oxford University Press.

Earley, J. (1970). "An efficient context-free parsing algorithm," *Communications of the Association for Computing Machinery*, 13(2).

Eddington, D. (2008). "2008 Presidential Address: Linguistics and the Scientific Method," *Southwest Journal of Linguistics*, 27, pp. 1-16.

Edelman, S. and Christiansen, M. H. (2003) "How seriously should we take Minimalist syntax?" in *TRENDS in Cognitive Sciences*, Vol.7, No. 2, pp. 60-61

Egan, F. (1991). "Propositional Attitudes and the Language of Thought," *Canadian Journal of Philosophy*, Vol. 21, No. 3, pp. 379-388.

Elman, J. L. (1992). "Grammatical Structures and Distributed Representations," in *Connectionism: Theory and Practice*, Vol. 3, Stephen Davis (ed), Oxford University Press.

- Elman, J. L. (1993). "Learning and development in neural networks: The importance of starting small," *Cognition*, 48, pp. 71-99.
- Epstein, S.D. and Hornstein, N. (eds.) (1999), *Working Minimalism*, Cambridge: MIT Press.
- Evans, N. and Levinson, S. C. (2009). "The myth of language universals: Language diversity and its importance for cognitive science," in *Behavioral and Brain Sciences*, 32, pp. 429-92.
- Featherston, S. (2001). *Empty Categories in Sentence Processing*, John Benjamins Publishing Company: Amsterdam.
- Ferreira, F., and Clifton, C., Jr. (1986). "The independence of syntactic processing," *Journal of Memory and Language*, 25, pp. 348-68.
- Ferreira, F., Christianson, K., Hollingworth, A. (2001) "Misinterpretations of Garden-Path Sentences: Implications for Models of Sentence Processing and Reanalysis," *Journal of Psycholinguistic Research*, Vol. 30, No. 1, 20.
- Ferreira, F. and Henderson, J. M. (1991). "Recovery of Misanalyses of Garden-Path Sentences," in *Journal of Memory and Language*, 30, pp. 725-745
- Ferreira, F. and Henderson, J. M. (1998). "Syntactic Reanalysis, Thematic Processing, and Sentence Comprehension," in *Reanalysis and Sentence Processing*, Fodor and Ferreira (eds.), Kluwer Academic Publishers, New York, , pp. xx.
- Field, H., (1978). "Mental Representation," *Erkenntnis*, 13, pp. 9-61.
- Filip, H., Tanenhaus, M. K., Carlson, G. N., Allopenna, P. D. and Blatt. J. (2002). "Reduced Relatives Judged Hard Require Constraint-Based Analyses," in Paola Merlo and Suzanne Stevenson (eds.) *Sentence Processing and the lexicon: formal, computational, and experimental perspectives*, Amsterdam: Benjamins, pp. 255-280.
- Fodor, J. A. (1968a). *Psychological Explanation*, Random House: New York
- Fodor, J. A. (1968b). "The Appeal to Tacit Knowledge in Psychological Explanation," *The Journal of Philosophy*, 65(20).
- Fodor, J. A. (1970). "Three reasons for not deriving 'kill' from 'cause to die'," *Linguistic Inquiry*, 1, pp. 429-438.
- Fodor, J. A. (1983). *The Modularity of Mind*. MIT Press: Cambridge
- Fodor, J. A. (1987). *Psychosemantics*. MIT Press: Cambridge
- Fodor, J. A. (1990). *A Theory of Content, and Other Essays*, MIT Press: Cambridge.

- Fodor, J. A. (1997). "Connectionism and the Problem of Systematicity (continued): Why Smolensky's solution still does not work," *Cognition*, 63, (1), pp. 109–119.
- Fodor, J. A. (1998). *Concepts: Where Cognitive Science Went Wrong*. OUP.
- Fodor, J. A. (2002). *The Mind Doesn't Work That Way*, MIT Press: Cambridge.
- Fodor, J. A. (2003). *Hume Variations*. New York: Oxford University Press.
- Fodor, J. A. (2008). *LOT 2: The Language of Thought Revisited*. New York: Oxford University Press.
- Fodor, J. A., Bever, T., and Garrett, M. (1974). *The Psychology of Language*, McGraw Hill: New York.
- Fodor, J., & McLaughlin, B. P. (1990). "Connectionism and the problem of systematicity: Why Smolensky's solution does not work," *Cognition*, 35, pp. 183–204.
- Fodor, J., & Pylyshyn, P. (1988). "Connectionism and cognitive architecture: A critical analysis," *Cognition*, 28, pp. 3–71.
- Fodor, J. A. and Katz, J. J. (1964). *The Structure of Language: Readings in the Philosophy of Language*. Englewood Cliffs, N.J.: Prentice-Hall, Inc.
- Fodor, J. D. (1989). "Empty Categories in Sentence Processing," in *Language and Cognitive Processes*, 4, pp. 155-209.
- Fodor, J. D. (1995). "Comprehending Sentence Structure," in *An Invitation to Cognitive Science: Volume 1, Language*, Gleitman and Liberman (eds.), Cambridge, MA: MIT Press, pp. 209-246.
- Fodor, J. D. (1998a). "Learning to Parse?" *Journal of Psycholinguistic Research*, Vol. 27, No. 2, pp. 285-319.
- Fodor, J. D. (1998b). "Parsing to Learn," *Journal of Psycholinguistic Research*, Vol. 27, No. 3, pp. 339-374.
- Fodor, J. D. and Ferreira, F. (eds.) (1998). *Reanalysis in Sentence Processing*, Kluwer Academic Publishers, New York.
- Fodor, J. D. and Frazier, L. (1980). "Is the Human Sentence Parsing Mechanism an ATN?," *Cognition*, 8, pp. 417-459.
- Fodor, J. D. and Inoue, A. (1994). "The Diagnosis and Cure of Garden Paths," *Journal of Psycholinguistic Research*, 23, pp. 407-434.

- Fodor, J. D. and Inoue, A. (1998) "Attach Anyway," in *Reanalysis and Sentence Processing*, Fodor and Ferreira (eds.), Kluwer Academic Publishers, New York, pp. xx.
- Fodor, J. D. and Inoue, A. (2000a). "Garden Path Repair: Diagnosis and Triage," *Language and Speech*, 43 (3), pp. 261-271.
- Fodor, J. D. and Inoue, A. (2000b). "Syntactic Features in Reanalysis: Positive and Negative Features," *Journal of Psycholinguistic Research*, Vol. 29, No. 1.
- Fong, S. (2005). "Computation with Probes and Goals: A Parsing Perspective," in DiSciullo, Anna Maria (ed.) *UG and External Systems, Language, Brain, and Computation*, Amsterdam: John Benjamins Publishing Co.
- Fong, S. (2011). *Minimalist parsing: Simplicity and feature unification*. Workshop on Language and Recursion. Mons, Belgium: University of Mons. March.
- Ford, M., Bresnan, J. and Kaplan, R. (1982). "A Competence-Based Theory of Syntactic Closure," in *The Mental Representation of Grammatical Relations*, J. Bresnan (ed.), MIT Press: Cambridge, Massachusetts.
- Forster, K. I. and Olbrei, I. (1973). "Semantic Heuristics and Syntactic Analysis," *Cognition*, 2, pp. 319-347.
- Forster, K. I. (1979). "Levels of processing and the structure of the language processor," in W.E. Cooper, and E.C.T. Walker (eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Erlbaum, pp. 27-85.
- Frank, R., Mathis, D. and Badecker, W. (2008). "The Acquisition of Anaphora by Simple Recurrent Networks," unpublished MS, available at the following URL:  
[http://whitney.ling.yale.edu/~rfrank/Bob\\_Frank/Publications\\_files/SRN-note-revised.pdf](http://whitney.ling.yale.edu/~rfrank/Bob_Frank/Publications_files/SRN-note-revised.pdf)
- Frank, S.L (2008). "Self-Organizing Word Representations for Fast Sentence Processing," in French and Thomas (eds.), *From Associations to Rules, Connectionist Models of Behavior and Cognition: Proceedings of the Tenth Neural Computation and Psychology Workshop*, World Scientific Publishing Co.: New Jersey, pp. 78-88.
- Frazier, L. (1979). *On Comprehending Sentences: Syntactic Parsing Strategies*, PhD Dissertation, available at: <http://digitalcommons.uconn.edu/dissertations/AAI7914150/>
- Frazier, L. (1988). "Grammar and Language Processing," in F. Newmeyer (ed.), *Linguistics: the Cambridge survey*. Cambridge. Vol. 1, pp. 15-34.
- Frazier, L. and Clifton C. (1989). "Successive Cyclicity in the Grammar and the Parser," *Language and Cognitive Processes*, 4, 2, pp. 93-126.
- Frazier, L. and Clifton C. (1996). *Construal*. Cambridge, MA: MIT Press.

- Frazier, L. and Clifton C. (1998). "Sentence Reanalysis, and Visibility," in *Reanalysis in Sentence Processing*, Ferreira and Fodor (eds.), Kluwer Academic Publishers, New York, pp. 143-176.
- Frazier, L. and Fodor, J. D. (1978). "The Sausage Machine: A New Two-Stage Parsing Model," *Cognition*, Vol. 6, pp. 291-325.
- Frazier, L., & Rayner, K. (1982). "Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences," *Cognitive Psychology*, 14, pp. 178-210.
- Frenck-Mestre, C. and Pynte, J. (1995). "Lexical Influences on Parsing Strategies: Evidence from Eye Movements," in *Studies in Visual Information Processing, Vol. 6 Eye Movement Research: Mechanisms, Processes, and Applications*, Findlay, J.M., Walker, R., and Kentridge, R.W. (eds.), pp. 433-444.
- Friederici, A.D. (1995). "The time course of syntactic activation during language processing: A Model Based on Neuropsychological and Neurophysiological Data," in *Brain and Language* 50, pp. 259-281.
- Friederici, A.D., Pfeifer, E., and Hahne, A. (1993). "Event-related brain potentials during natural speech processing: Effects of semantic, morphological, and syntactic violations," *Cognitive Brain Research*, 1, pp. 183-92.
- Friederici, A.D., Hahne, A., and Saddy, D. (2002). "Distinct Neurophysiological Patterns Reflecting Aspects of Syntactic Complexity and Syntactic Repair," *Journal of Psycholinguistic Research*, Vol. 31, No. 1.
- Friederici, A.D., Brauer, J. (2009). "Syntactic complexity in the brain," in *Syntactic Complexity: Diachrony, Acquisition, Neuro-cognition, Evolution*, T. Givón and M. Shibatani (eds.), John Benjamins Publishing Company: Amsterdam.
- Gazdar, G., Klein, E., Pullum, G., and Sag, I. (1985). *Generalized Phrase Structure Grammar*. Oxford: Basil Blackwell.
- Gardner-Chloros, P. (2009). *Code-switching*. Cambridge University Press.
- George, A. (ed.) (1989), *Reflections on Chomsky*, Oxford: Basil Blackwell.
- Gerth, S. and beim Graben, P. (2009). "Unifying syntactic theory and sentence processing difficulty through a connectionist minimalist parser," *Cognitive Neurodynamics*, 3, pp. 297-316.
- Gibson, E.A.F. (1991). *A Computational Theory of Human Linguistic Processing: Memory Limitations and Processing Breakdown*, unpublished Ph.D. dissertation, Carnegie Mellon University, available at: [tedlab.mit.edu/tedlab\\_website/researchpapers/Gibson%201991.pdf](http://tedlab.mit.edu/tedlab_website/researchpapers/Gibson%201991.pdf)



- Gibson, E. A. F. (1998). "Linguistic Complexity: Locality of Syntactic Dependencies," *Cognition*, 68, pp. 1-76.
- Gleitman, L. and Liberman, M., eds. (1995). *An Invitation to Cognitive Science: Volume I, Language*, Cambridge, MA: MIT Press.
- Godfrey-Smith, P. (2004). "On Folk Psychology and Mental Representation," in Clapin, Staines, and Slezak (eds.) *Representations in Mind: New Approaches to Mental Representation*, Elsevier Publishing, ch. 8.
- Goodman, N. (1954/1983). *Fact, Fiction, and Forecast*, 4<sup>th</sup> ed., Harvard University Press.
- Goodman, N. and Quine, W. V. (1947) "Steps Toward a Constructive Nominalism," *Journal of Symbolic Logic*, 12, pp. 105-122. Reprinted in Nelson Goodman, *Problems and Projects*, Bobbs-Merrill, 1972, pp. 173-198.
- Gorrell, P. (1991). "Subcategorization and Sentence Processing," in *Principle-Based Parsing: Computation and Psycholinguistics*, Berwick, R. C., Abney, S. P., and Tenny, C. eds., Dordrecht: Kluwer Academic Publishers, pp. 279-300.
- Greenberg, M. and Harman, G. (2006). "Conceptual-Role Semantics," in *The Oxford Handbook of Philosophy of Language*, Oxford University Press.
- Grodzinsky, Y. (2000). "The neurology of syntax: Language use without Broca's area," *Behavioral and Brain Sciences*, 23, pp. 1-71.
- Guasti, M. T. (2002). *Language Acquisition: The Growth of Grammar*. Cambridge: MIT Press.
- Gunderson, K., ed. (1975). *Minnesota Studies in the Philosophy of Science, Volume VII: Language, Mind, and Knowledge*, Minneapolis: University of Minnesota Press.
- Hadley, R. (1994) "Systematicity in Connectionist Language Learning," *Mind and Language*, ix, pp. 247-72.
- Haegeman, L. (1994). *Introduction to Government and Binding Theory*, 2nd ed., Oxford: Blackwell Publishers, 1<sup>st</sup> ed. 1991.
- Hahne, A. and Friederici, A. D. (2002). "Differential task effects on semantic and syntactic processes as revealed by ERPs," *Cognitive Brain Research*, 13, pp. 339-56.
- Hale, J. (1999). "Dynamical Parsing and Harmonic Grammar," unpublished report, available at the following URL: [courses.cit.cornell.edu/jth99/prism.ps](http://courses.cit.cornell.edu/jth99/prism.ps)

- Hale, J. (2001). "A Probabilistic Parser as a Psycholinguistic Model," in Proceedings of the NAACL, available at: <http://www.aclweb.org/anthology/N/N01/N01-1021.pdf>
- Hale, J. (2003). *Grammar, Uncertainty and Sentence Processing*, Unpublished Ph.D. dissertation, Johns Hopkins University.
- Hale, J. & Smolensky, Paul. (2006). "Harmonic Grammars and Harmonic Parsers for Formal Languages," in Smolensky, P. and Legendre, G., eds., *The Harmonic Mind*. Cambridge: MIT Press, Vol. 1, ch. 10, pp. 393–415.
- Halle, M., Bresnan, J. and Miller, G. A. (eds.) (1978). *Linguistic Theory and Psychological Reality*, Cambridge: MIT Press.
- Harkema, H. (2001). *Parsing Minimalist Languages*, PhD Dissertation, UCLA. Available at: <http://www.linguistics.ucla.edu/people/stabler/paris08/Harkema01.pdf>
- Harman, G. (1967). "Review: Psychological Aspects of the Theory of Syntax," *The Journal of Philosophy*, Vol. 64, No. 2, Some Recent Issues in Linguistics, pp. 75-87.
- Harman, G. (1969) Reply to Chomsky's Reply to the Review in *Language and Philosophy: A Symposium*, Hook, S., (ed.), New York: New York University Press.
- Harman, G. (1973). *Thought*. Princeton University Press.
- Harman, G. (1974a). "Review of Language and Mind," in *On Noam Chomsky: Critical Essays*, Gilbert Harman (ed.), Anchor Press, Doubleday: Garden City, NY, pp. 201-218.
- Harman, G., ed. (1974b). *On Noam Chomsky: Critical Essays*, Anchor Press, Doubleday: Garden City, NJ.
- Harman, G. (1986). *Change in View: Principles of Reasoning*. Cambridge, MA.: MIT Press. A pdf version is available at the following URL: [http://www.princeton.edu/\\_harman](http://www.princeton.edu/_harman)
- Harman, G. (1972/1999). "Meaning and Semantics," in *Semantics and Philosophy*, eds., M. Munitz and P. Unger, New York University Press, New York. Reprinted in *Reasoning, Meaning, and Mind*, OUP. (esp. p. 202)
- Harman, G. (1975) "Language, thought, and communication," in *Language, Mind, and Knowledge*, ed., K. Gunderson, University of Minnesota Press, Minneapolis, Minnesota,
- Harman, G. (1982). "Conceptual role semantics," *Notre Dame Journal of Formal Logic*, vol. 23, pp. 242-257.
- Harman, G. (1987). "(Nonsolipsistic) Conceptual Role Semantics," in *New Directions in Semantics*, Ernest LePore (ed.), Academic Press.

Harman, G. (1999). *Reasoning, Meaning, and Mind*, OUP

Harman, G. (2002). "Internal Critique: A Logic is not a Theory of Reasoning and a Theory of Reasoning is not a Logic," in D. M. Gabbay, R.H. Johnson, H. J. Ohlbach, and J. Woods, eds., *Handbook of the Logic of Argument and Inference: The Turn Towards the Practical*, Volume 1 in *Studies in Logic and Practical Reasoning*. Amsterdam: Elsevier, pp. 171-86.

Hodes, H. T. (1984). "Logicism and the Ontological Commitments of Arithmetic," *The Journal of Philosophy*, Vol. 81, No. 3, pp. 123-149.

Hodes, H. T. (1990). "Where Do the Natural Numbers Come from? In Memory of Geoffrey Joseph," *Synthese*, Vol. 84, No. 3, The Philosophy of Mathematics, Part II, pp. 347-407.

Hofweber, T. (2005). "Number Determiners, Numbers, and Arithmetic," *The Philosophical Review*, Vol. 114, No. 2, pp. 179-225.

Hook, S., ed. (1969). *Language and Philosophy: A Symposium*, New York: New York University Press.

Houghton, G., ed. (2005) *Connectionist Models in Cognitive Psychology*, New York: Psychology Press.

Horwich, P. (1998). *Meaning*. Oxford: Clarendon Press.

Horwich (2005). *Reflections on Meaning*. Oxford: Clarendon Press.

Hume, D. (1740/2000). *A Treatise of Human Nature*, edited by David Fate Norton and Mary J. Norton, Oxford/New York: Oxford University Press.

Hume, D. (1748/1999). *An Enquiry concerning Human Understanding*, edited by Tom L. Beauchamp, Oxford/New York: Oxford University Press.

Jackendoff, R. (2000). "Fodorian Modularity and Representational Modularity," in *Language and the Brain: Representation and Processing*, Yosef Grodzinsky, Lew Shapiro, David Swinney (eds.), San Diego: Academic Press, pp. 4-30.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press.

Johnson, M.. (1989). "Parsing as Deduction: The Use of Knowledge of Language," *Journal of Psycholinguistic Research*, Vol. 18, 1, pp. 105-128.

Johnson, M.. (1991). "Deductive Parsing: The Use of Knowledge of Language," in R. Berwick, S. Abney, and C. Tenny (eds.), *Principle-Based Parsing*, Kluwer Academic Publishers, Dordrecht.

Jurafsky, D. (1993). "A Cognitive Model of Sentence Interpretation: A Construction Grammar Approach," Technical Report TR-93-077, International Computer Science Institute, Berkeley. Available at the following URL: <ftp://ftp.icsi.berkeley.edu/pub/techreports/1993/tr-93-077.ps.gz>

Jurafsky, D. (1996). "A Probabilistic Model of Lexical and Syntactic Disambiguation," *Cognitive Science*, 20, pp. 137-194.

Jurafsky, D. (2003). "Probabilistic Modeling in Psycholinguistics: Linguistic Comprehension and Production," in *Probabilistic linguistics* Bod, Hay, & Jannedy, eds., MIT Press, Cambridge, MA.

Jurafsky, D, Martin, J. H. (2008). *Speech and Language Processing: An Introduction to Speech Recognition, Computational Linguistics, and Natural Language Processing (2nd edition)*, Prentice Hall.

Kaplan, R.M. (1973). A general syntactic processor. In Rustin, R. (Ed.), *Natural Language Processing*, pp. 193–241. Algorithmics Press, New York.

Kaplan, R. M. (1987). "Three seductions of computational psycholinguistics," in Whitelock, P., Wood, M.M., Somers, H. L., Johnson, R., and Bennett, P. (Eds.), *Linguistic Theory and Computer Applications*, pp. 149–188. Academic Press, London.

Karttunen, D. R. and Zwicky, A.M. (1985). "Introduction," in *Natural Language Processing: Psychological, Computational, and Theoretical Perspectives*, Dowty, D. R., Karttunen, L. and A. M. Zwicky (eds.), Cambridge University Press.

Katz, J. J. (1981). *Language and Other Abstract Objects*. Totowa, N. J.: Rowman & Littlefield.

Katz, J. J. (1984). "An Outline of a Platonist Grammar," in J. J. Katz (ed.) *The Philosophy of Linguistics* Oxford: Oxford University Press.

Katz, J. J. (ed.) (1985). *The Philosophy of Linguistics*. Oxford: Oxford University Press.

Katz, J. J. (1996). "The Unfinished Chomskyan Revolution," *Mind and Language*, 11, pp. 270–94.

Katz, J. (1990). *The Metaphysics of Meaning*. Cambridge: MIT Press.

Katz, J. (2000). *Realistic Rationalism*. Cambridge: MIT Press.

Katz, J., and Postal, P. M. (1991). "Realism vs. Conceptualism in Linguistics." *Linguistics and Philosophy*, vol. 14, pp. 515–54.

Kelly, S.D. (2001). "The non-conceptual content of perceptual experience: Situation dependence and fineness of grain," *Philosophy and Phenomenological Research*, 62(3), pp. 601–608.

- Kilbury, J. (1985). "Chart parsing and the Earley algorithm," in U. Klenk (ed.), *Kontextfreie Syntaxen und verwandte Systeme*, 76-89. Tübingen: Niemeyer. Available at the following URL: [user.phil-fak.uni-duesseldorf.de/~kilbury/Publ/Kilbury1985-CPEA.pdf](http://user.phil-fak.uni-duesseldorf.de/~kilbury/Publ/Kilbury1985-CPEA.pdf)
- Kimball, J. (1973) "Seven Principles of Surface Structure Parsing in Natural Language," *Cognition* 2, pp. 15-47.
- King, M. (1983). "Transformational Parsing," in *Parsing Natural Language*, M. King, ed., Academic Press.
- Kirsh, D. (1990). "When Is Information Explicitly Represented?" in *Information, Language and Cognition*, P. Hanson (ed.), University of British Columbia Press, re-issued by Oxford University Press (1992), pp. 340-365.
- Knowles, J. (2000). "Knowledge of Grammar As a Propositional Attitude." *Philosophical Psychology* 13(3), pp. 325-53.
- Koster, van Riemsdijk, Vernaad (1977). "GLOW manifesto," GLOW newsletter, available at <http://glow.uit.no/index.php?page=manifesto#glow>
- Kripke, S. (1972/1980). *Naming and Necessity*, Cambridge, MA: Harvard University Press
- Kukla, A. and Walmsley, J. (2006). *Mind: A Historical and Philosophical Introduction to the Major Theories*, Hackett Publishing.
- Kuhn, T. S. (1962/1996). *The Structure of Scientific Revolutions*, 3rd ed. Chicago, IL: University of Chicago Press.
- Lawrence, S., Giles, L., Fong, S. (2000). "Natural Language Grammatical Inference with Recurrent Neural Networks," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 12, No. 1, pp. 126-140.
- Larson, R. K. (2010). *Grammar as Science*. Cambridge, MA: MIT Press.
- Larson, R. K. and Ludlow, P. (1997). "Interpreted Logical Forms," in *Readings in the Philosophy of Language*, Peter Ludlow (ed.), Cambridge, MA: MIT Press, pp. 993-1039.
- Levin, M. (1979). "Explanation and Prediction in Grammar (and Semantics)," *Midwest Studies in Philosophy*, II, pp. 128-137.
- Levine, R. D. and Postal, P. (2004). "A Corrupted Linguistics," Collier and Horowitz (eds.) *The Anti Chomsky Reader*, San Francisco, CA: Encounter Books, pp. 203-233.
- Lewis, D. K. (1969/2002). *Convention*. Malden, MA: Blackwell Publishing.

Lewis, D. K. (1972). "Psychophysical and Theoretical Identifications," *Australasian Journal of Philosophy*, 50, pp. 249-258.

Lin, D. (1993). "Principle-Based Parsing without Overgeneration," available at the following URL: [www ldc.upenn.edu/acl/P/P93/P93-1016.pdf](http://www ldc.upenn.edu/acl/P/P93/P93-1016.pdf)

Lin, D. (1994). "PRINCIPAR: An Efficient, Broad-coverage, Principle-based Parser," available at the following URL: [www ldc.upenn.edu/acl/C/C94/C94-1079.pdf](http://www ldc.upenn.edu/acl/C/C94/C94-1079.pdf)

Loar, B. (1990). "Phenomenal states," *Philosophical Perspectives*, 4, pp. 81-108.

Loewer, B. (2002). "A Guide to Naturalizing Semantics," available at <http://www.nyu.edu/gsas/dept/philo/courses/rules/papers/Loewer.pdf>

Lyons, W. (1986). *The disappearance of introspection*, Cambridge, MA: MIT Press

MacDonald, M.C. (1989). "Priming Effects from Gaps to Antecedents," *Language and Cognitive Processes*, 4, pp. 1-72.

MacDonald, M. E., Pearlmutter, D., and Seidenberg, M. (1994). "The Lexical Nature of Syntactic Ambiguity resolution," *Psychological Review*, 101, pp. 678-703.

Marcus, Mitchell P. (1980) *A Theory of Syntactic Recognition for Natural Language*. MIT Press: Cambridge, MA.

Marcus, M. P. (1984). "Some Inadequate Theories of Human Language Processing," in *Talking Minds: The Study of Language in Cognitive Science* Bever, T.G., Carroll, J.M., and Miller, L.A., eds., MIT Press: Cambridge, MA, pp. 253-278.

Marslen-Wilson, W. D. and Tyler, L. K. (1998) Rules, representations, and the English past tense. *Trends in Cognitive Sciences* 2, pp. 428-435.

Marslen-Wilson, W. D. and Tyler, L. K. (2003) Capturing underlying differentiation in the human language system. *Trends in Cognitive Sciences*, 7, pp. 62-63.

Matthews, R. J. (1994). "The Measure of Mind," *Mind*, New Series, Vol. 103, No. 410, pp. 131-146.

Matthews, R. J. (2007). *The Measure of Mind*, Oxford University Press.

May, R. (1985). *Logical Form: Its Structure and Derivation*, Cambridge: MIT Press.

May R. (1997). "Logical form as a Level of Linguistic Representation," in *Readings in the Philosophy of Language*, P. Ludlow (ed.), Cambridge: MIT Press.

Mayberry, M. R. and Miikkulainen, R. (2000). "Combining Maps and Distributed

Representations for Shift-Reduce Parsing,” in *Hybrid Neural Systems*, Stefan Wermter and Ron Sun (eds.), Springer: Berlin, pp. 144-157.

McElree, B. and Bever, T.G. (1989). “The Psychological Reality of Linguistically Defined Gaps,” *Journal of Psycholinguistic Research*, 18, pp. 21-35.

McRoy, S. and Hirst, G. (1990). “Race-based parsing and syntactic disambiguation,” *Cognitive Science*, 14, pp. 313-353.

Meng, M., and Bader, M. (2000a). “*The role of case and number features in syntactic ambiguity resolution*,” in M. de Vincenzi & V. Lombardo (eds.), *Cross-linguistic perspectives on language processing*, Dordrecht: Kluwer Academic Publishers, pp. 175-204.

Meng, M., and Bader, M. (2000b). “Mode of Disambiguation and Garden-path Strength: An Investigation of Subject-Object Ambiguities in German,” *Language and Speech*, 43(1), pp. 43-74.

Merlo, P. (1995) “Modularity and Information Content Classes in Principle-Based Parsing.” *Computational Linguistics*, 21, (4), pp. 515-541

Merlo, P. and Stevenson, S., eds. (2002). *The Lexical Basis of Sentence Processing: Formal, Computational, and Experimental Issues*, Amsterdam: John Benjamins Publishing Company.

Millikan, R. G. (1984). *Language, Thought and Other Biological Categories: New Foundations for Realism*. Cambridge, Mass: MIT Press

Millikan, R. G. (2004). “In Defense of Public Language,” in Antony and Hornstein, (eds.), pp. 215-37.

Milner, Jean-Claude (1979). “Reply to GLOW Manifesto,” GLOW Newsletter, available at <http://glow.uit.no/index.php?page=milner#glow>

Moore, R.C. (2000). “Improved left-corner chart parsing for large context-free grammars,” in *Proceedings of the Sixth International Workshop on Parsing Technologies*, available at the following URL: <http://research.microsoft.com/apps/pubs/default.aspx?id=68866>

Morawietz, F. (2000). “Chart Parsing and Constraint Programming,” in *Proceedings of the 18th conference on Computational linguistics - Volume 1*, available at: <http://portal.acm.org/citation.cfm?id=990900>

Neander, K. (2006). “Naturalistic Theories of Reference,” in *The Blackwell Guide to the Philosophy of Language*, Devitt, M. and Hanley, R. (eds.), Blackwell Publishing, pp. 374-391

Neale, S. (1997). “Meaning, Grammar, and Indeterminacy,” *Dialectica* Vol. 41, N°4.

Neeleman, A., & van de Koot, H. (2010). “Theoretical validity and psychological reality of the

grammatical code,” in De Mulder, H., Everaert, M., Nilsen, O, Lentz, T., and Zondervan, A. (eds.), *The Linguistic Enterprise: From Knowledge of Language to Knowledge in Linguistics*, Amsterdam: John Benjamins.

Neville, H.J. Nicol, J., Barss, A., Forster, K., and Garrett, M.F. (1991). “Syntactically based sentence processing classes: Evidence from event-related potentials,” *Journal of Cognitive Neuroscience*, 6, pp. 233–55.

Newell, A & Simon, H. A. (1963). “GPS: A Program that Simulates Human Thought,” in Feigenbaum, E.A. and Feldman, J., eds., *Computers and Thought*, New York: McGraw-Hill.

Ni, W., Fodor, J. D., Crain, S., and Shankweiler, D. (1998). “Anomaly Detection: Eye Movement Patterns,” *Journal of Psycholinguistic Research*, Vol. 27, 5, pp. 515-539

Nicol, J. & Swinney, D. (1989). “The role of structure in coreference assignment during sentence comprehension,” *Journal of Psycholinguistic Research*, 18, pp. 5-19.

Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.

O'Donnell, M. (1995). *Sentence Analysis and Generation: A Systemic Perspective*, Unpublished Ph.D. Dissertation, University of Sydney, available at: <http://www.wagsoft.com/Papers/Thesis/>

Ogihara, T. (2003). "Adverbs of Quantification and Sequence of Tense Phenomena," in Javier Gutierrez-Rexach (ed.), *Semantics: Critical Concepts*, Volume 4, pp. 319-337, Routledge: London.

Orenstein, A. (2002). *W. V. Quine*, Princeton University Press.

Osterhaut, L. and Swinney, D.A. (1992). “On the Temporal Course of Gap Filling During Comprehension of Verbal Passives,” *Journal of Psycholinguistic Research*, Vol. 2, No. 2.

Papineau, D. (2002). *Thinking about Consciousness*. New York: Oxford University Press.

Peacocke, C. (1989) “When is Grammar Psychologically Real?” in *Reflections on Chomsky*, A. George (ed.), Oxford: Basil Blackwell, pp. 111- 130.

Pereira, F.C.N. (1985). “A New Characterization of Attachment Preferences,” in *Natural Language Processing*, Dowty, D. R., Karttunen, L. and Zwicky, A.M., (eds.), Cambridge University Press: Cambridge, England.

Pereira, F.C.N. and Warren, D.H.D. (1983). “Parsing as Deduction.” In *Proceedings of the 22nd Annual Meeting of the Association for Computational Linguistics*, pp. 137-144.

Pereplyotchik, D. (in progress), “An Argument for Semantic Holism,” MS

Pereplyotchik, D. (2008), “Language Before Thought?” MS



Pereplyotchik, D. (2011), "Psychological and Computational Models of Language Comprehension: In Defense of the Psychological Reality of Syntax," *Croatian Journal of Philosophy*, Vol. XI, No. 31, pp. 31-72.

Peters, P.S. and Ritchie, R.W. (1973). "On the Generative Power of Transformational Grammars." *Information Sciences*, 10.

Phillips, C. (1994). *Order and Structure* MIT Dissertation

Phillips, C. (1996). "Summary of *Order and Structure*, MIT Dissertation" in GLOT International. Available at <http://www.ling.umd.edu/colin/research> .

Phillips, C. and Lewis, S. (forthcoming). "Derivational Order in Syntax: Evidence and Architectural Consequences," in *Directions in Derivations*, C. Chesì (ed.), Elsevier.

Pickering M. & Barry G. (1991). "Sentence processing without empty categories," *Language and Cognitive Processes*, 6, pp. 229–259.

Pickering, M. J. and van Gompel, R. P. G. (2006). "Syntactic Parsing," in M. J. Traxler & M. A. Gernsbacher (eds.), *The Handbook of Psycholinguistics*, San Diego CA: Elsevier, pp. 455-503.

Pickering, M. J. and Ferreira, V. S. (2008). "Structural Priming: A Critical Review," *Psychological Bulletin*, Vol. 134, No. 3, pp. 427-459.

Pickering, M. J., Traxler, M. J. and Crocker, M. W. (2000). "Ambiguity Resolution in Sentence Processing: Evidence against Frequency-Based Accounts," *Journal of Memory and Language* 43, 3, pp. 447-475.

Pietroski, P. (2008). "Think of the Children: Comments on Ignorance of Language, by Michael Devitt," *Australasian Journal of Philosophy*, 86, pp. 657-69.

Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*. New York: William Morrow and Co.

Pinker, S. (1997). *How the Mind Works*. New York: W. W. Norton and Company.

Pollard, C., and Sag, I. A. (1994). *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.

Postal, P. (2009). "The Incoherence of Chomsky's 'Biolinguistic' Ontology," *Biolinguistics*, Vol. 3, No 1.

Prinz, J. J.(2000). "A Neurofunctional Theory of Visual Consciousness," *Consciousness and Cognition*, 9, pp. 243-259.

Prinz, J. (2006). "Is the Mind Really Modular?" in *Contemporary Debates in Cognitive Science*, R. Stainton (ed.), Malden, MA: Blackwell, pp. 22-36.

Pritchett, B. (1992). *Grammatical Competence and Parsing Performance*, University Of Chicago Press.

Pullum, G. K. (1991). "Review of Principle-Based Parsing: Computation and Psycholinguistics," in *Studies in Linguistics and Philosophy*, 44, Dordrecht: Kluwer Academic Publishers, pp. 393-396. Available at: [www.aclweb.org/anthology/I93-2010](http://www.aclweb.org/anthology/I93-2010)

Pullum, G. K. and Scholz, B. C. (2007). "Systematicity and natural language syntax," *Croatian Journal of Philosophy*, 7, 375-402.

Pullum, G. K. and Scholz, B. C. (forthcoming). "Recursion and the Infinitude Claim," to appear in Harry van der Hulst (ed.), *Recursion in Human Language*, Berlin: Mouton de Gruyter.

Putnam, H. (1960). "Minds and Machines," in *Dimensions of Mind: A Symposium*, S. Hook, ed., New York University Press: New York.

Putnam, H. (1967). "Psychological Predicates," in *Art, Mind, and Religion*, W. H. Capitan and D. Merrill, eds., University of Pittsburg Press pp. 37-48. Reprinted as "The Nature of Mental States," in *The Nature of Mind*, D. Rosenthal, ed. Oxford University Press: New York, 1991, pp. 197-203.

Putnam, H. (1973). "Meaning and Reference." *The Journal of Philosophy*, Vol. 70, No. 19, Seventieth Annual Meeting of the American Philosophical Association Eastern Division. (November, 1973), pp. 699-711.

Putnam, H. (2004). *The Collapse of the Fact/Value Dichotomy, and Other Essays*. Harvard University Press.

Pylyshyn, Z. (1986). *Computation and Cognition: Toward a Foundation for Cognitive Science*. MIT Press.

Quine, W. V. (1953). "The Problem of Meaning in Linguistics," in *From a Logical Point of View*, Cambridge, MA: Harvard University Press, pp. 47-64. Reprinted in Katz, ed. (1985).

Quine, W. V. (1960). *Word and Object*. Cambridge, MA: MIT Press.

Quine, W. V. (1969). *Ontological Relativity and Other Essays*. Columbia University Press.

Quine, W. V. (1970). "Methodological Reflections on Current Linguistic Theory" *Synthese* 21, pp. 386-398. Reprinted in *Semantics of Natural Language* D. Davidson and G. Harman (eds.).

Quine, W. V. (1974). *The Roots of Reference*. Open Court Press.

- Quine, W. V. (1975a). "Response to Chomsky," in *Words and Objections*, Davidson and Hintikka (eds.), Dordrecht: Kluwer Academic Publishers.
- Quine, W. V. (1975b). "Mind and Verbal Disposition," in *Mind and Language*, S. Guttenplan (ed.), Oxford: Oxford University Press, pp. 83-95. Reprinted in W. V. Quine, *Quintessence*, R. Gobson (ed.), Cambridge: Harvard University Press, 2004, pp. 313-325; and again in W. V. Quine, *Confessions of a Confirmed Extensionalist and Other Essays*, D. Follesdal and D. B. Quine (eds.), Cambridge: Harvard University Press, 2004, pp. 244-257.
- Quine, W. V. (1976). *The Ways of Paradox, and Other Essays* (Revised and Enlarged Edition), Cambridge, MA: Harvard University Press.
- Quine, W. V. (1981). "The Pragmatists' Place in Empiricism," in Mulvancy and Zeltner (ed.), *Pragmatism its Sources and Prospects*, University of South Carolina Press, pp. 23-39.
- Quine, W. V. (1986). *Philosophy of Logic* (2<sup>nd</sup> ed.). Harvard University Press.
- Quine, W. V. (1992). *Pursuit of Truth*. Harvard University Press.
- Rattan, G. (2002). "Tacit knowledge of grammar: a reply to Knowles." *Philosophical Psychology* 15(2): 135-54.
- Rayner, K., (1998). "Eye Movements in Reading and Information Processing: 20 Years of Research," *Psychological Bulletin*, Vol. 124, No. 3, pp. 372-422.
- Rayner, K., Carslton, M., and Frazier, L. (1983). "The Interaction of Syntax and Semantics During Sentence Processing: Eye Movements in the Analysis of Semantically Biased Sentences," *Journal of Verbal Learning and Verbal Behavior*, 22, pp. 358-374.
- Rayner, K., Garrod, S., & Perfetti, C. A. (1992). "Discourse influences during parsing are delayed," in *Cognition*, 45, pp. 109-139.
- Rey, G. (2003a), in *Chomsky and His Critics*, Antony and Hornstein (eds.)
- Rey, G. (2003b), in *Epistemology of Language*, A. Barber (ed.)
- Rey, G. (2006a). "The Intentional Inexistence of Language—But Not Cars," in R. Stainton (ed.), *Contemporary Debates in Cognitive Science* (Oxford: Blackwell Publishers), pp. 237-55.
- Rey, G. (2006b). "Conventions, Intuitions and Linguistic Inexistents: A Reply to Devitt." *Croatian Journal of Philosophy*, 6, 549-69.
- Rey, G. (2008). "In Defense of Folieism," in *Croatian Journal of Philosophy*, pp. 177-202.

- Roark, B. and Johnson, M. (1999). "Efficient probabilistic top-down and left-corner parsing," in *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pp. 421-428. Available at the following URL: [www ldc.upenn.edu/acl/P/P99/P99-1054.pdf](http://www ldc.upenn.edu/acl/P/P99/P99-1054.pdf)
- Rogers, Hartley (1967). *Theory of Recursive Functions and Effective Computability*, MIT Press.
- Rohde, D.L.T. (1999). A Connectionist Model of Sentence Comprehension and Production. Unpublished PhD thesis proposal, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.
- Rohde, D.L.T. (2002). *A Connectionist Model of Sentence Comprehension and Production* Ph.D. Dissertation, Department of Computer Science, Carnegie Mellon University
- Rosenthal, D. (1968). *Intentionality: A Study of the Views of Chisholm and Sellars*, PhD. Dissertation, Princeton University.
- Rosenthal, D. (1986). "Intentionality," *Midwest Studies in Philosophy*, x, pp. 151–184; postscript in *Rerepresentation: Readings in the Philosophy of Mental Representation*, ed. Stuart Silvers, Dordrecht: Kluwer Academic Publishers, 1989, pp. 341–344. Reprinted as ch. 3 in Rosenthal (2005).
- Rosenthal, D. (2005). *Consciousness and Mind*. Oxford University Press.
- Rosenthal, D. (2008). "Consciousness and Its Function," *Neuropsychologia*, 46, 3, pp. 829-840
- Sadeh-Leicht, O. (2007). *The Psychological Reality of Grammar, The Theta Principle in Parsing Performance*, Doctoral Dissertation, Universitat Utrecht
- Sag, I. A. and Fodor, J. D. (1994). "Extraction without traces," *Proceedings of the 13th West Coast Conference on Formal Linguistics*, pp. 365–384.
- Sahoo, K. (2011). "Review of Constituent Structure, by Andrew Carnie," in Linguist List, available at the following URL: <http://linguistlist.org/issues/22/22-2415.html>
- Schabes, Y., Abeille, A., Joshi, A.K. (1988). "Parsing Strategies with 'Lexicalized' Grammars: Application to Tree Adjoining Grammars," in *Proceedings of the 12<sup>th</sup> International Conference on Computational Linguistics*, available at: [acl ldc.upenn.edu/C/C88/C88-2121.pdf](http://acl ldc.upenn.edu/C/C88/C88-2121.pdf)
- Schank, R., Tesler, L, and Weber, S, (1970) "Spinoza II: Conceptual Case-based Natural Language Analysis," Research Report AIM-109, Stanford, CA: Dept. of Computer Science.
- Schank, R. (1972). "Conceptual Dependency: A Theory of Natural Language Understanding," *Cognitive Psychology*, 3, pp. 552-631.
- Schank, R. (1975a). *Conceptual Information Processing*. Amsterdam: North-Holland Publishing.

Schank, R. and the Yale AI Project (1975a). "SAM: A Story Understander," Research Report 43, New Haven, CT: Yale University Dept. of Computer Science.

Schank, R. and Abelson, R. (1977). *Scripts, Plans, Goals, and Understanding*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schank, R. and Riesbeck, C. (1978). "Comprehension by Computer," in *Studies in the Perception of Language*, W. Levelt and G. Flores d'Arcais, (eds.), Chichester, UK: John Wiley and Sons.

Schank, R. (1979a), "Language and Memory," *Cognitive Science*, 4, pp. 243-284.

Schank, R. (1979b), "Reminding and Memory Organization: An Introduction to MOPs," Research Report 170, New Haven, CT: Yale University Dept. of Computer Science.

Schank, R., Lebowitz, M., and Birnbaum, L. (1980). "An Integrated Understander," *American Journal of Computational Linguistics*, 6, pp. 13-30.

Schank, R. and Riesbeck, C. (1981). *Inside Computer Understanding*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schank, R. and Birnbaum, L. (1984). "Memory, Meaning, and Syntax," in *Talking Minds: The Study of Language in Cognitive Science* Bever, T.G., Carroll, J.M., and Miller, L.A., eds., MIT Press: Cambridge, MA, pp. 209-251.

Schwartz, S. (1980). "The Search for Pronominal Referents," Research Report 10, New Haven, CT: Yale University, Cognitive Science Program.

Searle, J. (1990). "Consciousness, Explanatory Inversion, and Cognitive Science," *The Behavioral and Brain Sciences*, 13, 4, pp. 585-696, and Author's Response, "Who is Computing with the Brain?," pp. 632-642.

Sellars, W. (1953). "Inference and Meaning," *Mind*, 62, No. 247, pp. 313-338.

Sellars, W. (1963a). "Empiricism and the Philosophy of Mind," in *Science Perception and Reality*. Atascadero, CA: Ridgeview Publishing, pp.

Sellars, W. (1963b). "Some Reflections on Language Games," in *Science, Perception, and Reality*. Atascadero, CA: Ridgeview Publishing, pp. 321-358.

Sellars, W. (1963c). "Empiricism and Abstract Entities," in *The Philosophy of Rudolf Carnap*, Paul Schilpp (ed.), La Salle, Illinois: Open Court.

Sellars, W. (1974). "Meaning as Functional Classification," *Synthese*, 27, pp. 417-37.

- Shieber, S.M. (1985). "Evidence against the Context-Freeness of Natural Language," in *Linguistics and Philosophy*, 8.
- Shieber, S.M. (1988). "A uniform architecture for parsing and generation," in *Proceedings of the 12th International Conference on Computational Linguistics*, Budapest, Hungary, pp. 614-619, available at the following URL: <http://dash.harvard.edu/handle/1/2265286>
- Shieber, S.M. and Johnson, M. (1993). "Variations on Incremental Interpretation," *Journal of Psycholinguistic Research*, Vol. 2, No. 2.
- Shieber, S.M., Schabes, Y., and Pereira, F.C.N. (1993). "Principles and implementation of deductive parsing," Technical Report CRCT TR-11-94, Computer Science Department, Harvard University, Cambridge, Massachusetts. Available at: <http://arXiv.org/>.
- Slobin, D. (1966). "Grammatical Transformations and Sentence Comprehension in Childhood and Adulthood," *Journal of Verbal Learning and Verbal Behavior*, 5, pp. 219-227.
- Smolensky, P. (1988/1995) "Connectionism, Constituency, and the Language of Thought," in *Connectionism: Debates on Psychological Explanation*, ed. C. Macdonald & G. Macdonald. Oxford, UK: Blackwell Publishers;
- Smolensky, P. (1995) "Reply: constituent structure and explanation in an integrated connectionist/symbolic cognitive architecture," in *Connectionism: Debates on Psychological Explanation*, ed. C. Macdonald & G. Macdonald. Oxford, UK: Blackwell Publishers, pp. 222-290.
- Smolensky, P. (2006). "Harmony Optimization and the Computational Architecture of the Mind/Brain," ch. 1 in *The Harmonic Mind*. Smolensky and Legendre, eds., Cambridge: MIT Press, pp. 1-61.
- Smolensky, P. and Legendre, G., eds. (2006). *The Harmonic Mind*. Cambridge: MIT Press.
- Soames, S. (1984). "Linguistics and Psychology". *Linguistics and Philosophy* 7, pp. 155-79.
- Soames, S. (1985). "Semantics and Psychology," in Katz (ed.), *Philosophy of Linguistics*.
- Sprouse, J. and Almeida, D. (forthcoming). "The role of experimental syntax in an integrated cognitive science of language," in *The Cambridge Handbook of Biolinguistics*., Kleanthes Grohmann and Cedric Boeckx (eds.) Available at <http://www.socsci.uci.edu/~jsprouse/#papers>
- Stainton, R. (ed.) (2006). *Contemporary Debates in Cognitive Science*. Malden, MA: Blackwell.
- Stabler, E. P. (1983). "How are Grammars Represented?" in *Behavioral and Brain Sciences* 6: 391-402.
- Stabler, E. P. (1984). "Review of Berwick and Weinberg" in *Cognition*, 17, pp. 155-179.

- Stabler, E. P. (1992). *The Logical Approach to Syntax: Foundations, Specifications, and Implementations of Theories of Government and Binding*. Cambridge: MA, MIT Press
- Stabler, E.P. (1994). "The Finite Connectivity of Linguistic Structure," in *Perspectives on Sentence Processing*, C. Clifton, Jr., L. Frazier, and K. Rayner, (eds.), Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 303-336.
- Stabler, E.P. (1997). "Derivational Minimalism," in C. Retoré, editor, *Logical Aspects of Computational Linguistics*, vol. 1328 of *Lecture Notes in Artificial Intelligence*. Springer Verlag, Berlin, Heidelberg, Germany.
- Stabler, E.P. (2001). "Minimalist Grammars and Recognition," in Christian Rohrer, Antje Rossdeutscher, and Hans Kamp, eds., *Linguistic Form and Its Computation*. CSLI Publications, Stanford, California, 2001.
- Steedman, M. (1985). "LFG and Psychological Explanation," in *Linguistics and Philosophy*, Vol. 8, No. 3 (Aug., 1985), pp. 359-385.
- Steedman, M. (1992). "Grammars and Processors," University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-92-53. Available at: <http://repository.upenn.edu/cis-reports/475>
- Steedman, M. (1999). "Connectionist Sentence Processing in Perspective," *Cognitive Science*, Vol. 23, (4), pp. 615-634.
- Steedman, M. (2000). *The Syntactic Process*. Cambridge: MIT Press
- Steedman, M. (2002). "Connectionist and Symbolist Sentence Processing," in Merlo, P. and Stevenson, S., eds., *The Lexical Basis of Sentence Processing: Formal, Computational, and Experimental Issues*, John Benjamins Publishing Company, pp. 95-108.
- Sterelny, K. (1991). *The Representational Theory of Mind: An Introduction*. Wiley-Blackwell.
- Stevenson, S. (1993) "A Competition-based explanation of syntactic attachment preferences and garden path phenomena." *Proceedings of the 31st Annual Association for Computational Linguistics*, pp. 266-273.
- Stevenson, S. and P. Merlo (1997) "Lexical Structure and Processing Complexity," in *Language and Cognitive Processes* vol. 12.2/3, pp. 349-399.
- Stevenson, S. (1999). "Bridging the Symbolic-Connectionist Gap in Language Comprehension." in E. Lepore, Ed., *What is Cognitive Science?*, pp. 336-355, Oxford, UK: Blackwell Publishers.
- Stevenson, S. and Smolensky, P. (2006). "Optimality in Sentence Processing," ch. 19 in *The Harmonic Mind*. Smolensky and Legendre, eds., Cambridge: MIT Press, pp. 307-337.

- Stich, S. (1971). "What Every Speaker Knows," *The Philosophical Review*, 80 (4), pp. 476-496.
- Stich, S. (1972). "Grammar, Psychology and Indeterminacy," *Journal of Philosophy*, LXIX, 22, pp. 799-818.
- Stich, S. (1973). "What Every Grammar Does: A Reply to Prof. Arbini," *Philosophia*, Vol. 3 No. 1, pp. 85-96.
- Stich, S. (1974). "Review of *The Underlying Reality of Language and its Philosophical Import* by Jerrold J. Katz," *The Philosophical Review*, Vol. 83, No. 2, pp. 259-263.
- Stich, S. (1975). "Competence and Indeterminacy," in *Testing Linguistic Hypotheses*, Cohen, D. and Wirth, J. R., (eds.) Washington: John Wiley, 1975.
- Stich, S. (1978a). "Empiricism, Innateness, and Linguistic Universals," *Philosophical Studies*, Vol. 33, No. 3, pp. 273-286.
- Stich, S. (1978b). "Beliefs and Subdoxastic States," *Philosophy of Science*, Vol. 45, No. 4., pp. 499-518.
- Stich, S. (1979). "Between Chomskian Rationalism and Popperian Empiricism," *British Journal for the Philosophy of Science*, 30, pp. 329-47.
- Stich, S. (1980). "What Every Speaker Cognizes," *Behavioral and Brain Sciences*, 3, pp. 39-40.
- Stich, S. (1981). "Can Popperians Learn to Talk?" in *The British Journal for the Philosophy of Science*, Vol. 32, No. 2, pp. 157-164.
- Stich, S. (1985). *From Folk Psychology to Cognitive Science: The Case against Belief*, Cambridge: MIT Press.
- Stowe, L.A. (1986). "Parsing WH-constructions: Evidence for on-line gap location," *Language and Cognitive Processes*, 3, pp. 227-245.
- Strawson, G. (1994). *Mental Reality*. Cambridge: MIT Press.
- Stroud, C. and Phillips, C. (2010). "Examining the evidence for an independent semantic analyzer: an ERP study in Spanish"
- Sturt, P., Pickering, M.J., Scheepers, C. and Crocker, M.W. (2001). "The Preservation of Structure in Language Comprehension: Is Reanalysis the Last Resort?" in *Journal of Memory and Language* 45, pp. 283-307.
- Tanenhaus, M.K., Carlson, G.N. Seidenberg, M.S. (1985). "Do Listeners Compute Linguistic Representations?" in *Natural Language Processing: Psychological, Computational, and Theoretical Perspectives*, Dowty, Karttunen, and Zwicky (eds.), Cambridge University Press.



- Thomas, J. D. (1995). "Center-embedding and Self-embedding in Human Language Processing," MS Thesis, MIT.
- Tomasello, M. (2005). "Beyond Formalities: The Case of Language Acquisition," *The Linguistic Review*, 22, pp. 183–197.
- Traxler, M.J. and Pickering, M.J. (1996). "Case Marking in the Parsing of Garden-Path Sentences: Evidence from Eye Movements," *The Quarterly Journal of Experimental Psychology*, 49A, 4, pp. 991-1004.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). "Semantic influences on parsing: Use of thematic role information in syntactic disambiguation," *Journal of Memory and Language*, 33, pp. 285–318.
- Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge MA: MIT Press.
- Tyler, L. (1980). "Serial and Interactive Parallel Theories of Syntactic Processes," MIT Center for Cognitive Studies Occasional Paper No. 8
- Tyler, L. and Marslen-Wilson, W. (1977). "The On-line Effects of Semantic Context on Syntactic Processing," *Journal of Verbal Learning and Verbal Behavior*, 16, pp. 683-692.
- Ullman, M.T. et al. (1997). "A neural dissociation within language: evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system," *Journal of Cognitive Neuroscience*, 9, 266–276
- Van Gelder, T. (1990). "Compositionality: A Connectionist Variation on a Classical Theme," *Cognitive Science*, 14, pp. 355-384.
- Van Gompel, R. P. G., Pickering, M. J., Pearson, J., and Liversedge, S. P. (2005). "Evidence against competition during syntactic ambiguity resolution." *Journal of Memory and Language*, 52, pp. 284–307.
- Vasishth, S. and Lewis, R. (2006). "Symbolic Models of Human Sentence Processing," in Brown, K., ed., *Encyclopedia of Language and Linguistics*, 2nd ed., vol. 5, Elsevier, pp. 410-419. available at: <http://www-personal.umich.edu/~rickl/pubs/vasishth-lewis-2006-ell-article.pdf>
- Walenski, M. (2002). *Relating Parsers and Grammars: On the structure and real-time comprehension of English infinitival complements*, PhD Dissertation, UC San Diego
- Wanner, E. (1980). "The ATN and the Sausage Machine: Which One is Baloney?" *Cognition*, 8, pp. 209-225.
- Wanner, E. and Maratsos, M. (1978). "An ATN Approach to Comprehension," in *Linguistic Theory and Psychological Reality*, Halle, Bresnan, and Miller (eds.), Cambridge: MIT Press.

Wasow, T. (1978). "On constraining the class of transformational languages," in *Synthese* 39, pp. 81-104.

Weinberg (1999). "A Minimalist Theory of Human Sentence Processing," in Epstein and Hornstein (eds.), *Working Minimalism*, MIT Press, pp. 283-315.

Woods, W. (1973). "An Experimental parsing System for Transition Network Grammars," in R. Rustin (ed.), *Natural Language Processing*, New York: Algorithmics Press, 111-154.

Yang, C. and Berwick, R.C. (1996). "Principle-Based Parsing for Chinese," in *Language, Information and Computation* (PACLIC 11), pp. 363-371. Available at:  
[www.aclweb.org/anthology-new/Y/Y96/Y96-1038.pdf](http://www.aclweb.org/anthology-new/Y/Y96/Y96-1038.pdf)

Zorzi, Marco, (2005) "Computational Models of Reading," pp. 403-444, in Houghton, George, ed., *Connectionist Models in Cognitive Psychology*, New York: Psychology Press