From Molecule to Metaphor

A Neural Theory of Language

Jerome A. Feldman

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 55 Hayward Street, Cambridge, MA 02142.

This book was set in Stone Sans and Stone Serif by SNP Best-set Typesetter Ltd., Hong Kong and was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Feldman, Jerome A.
From molecule to metaphor : a neural theory of language / by Jerome A. Feldman.
p. cm.
Includes bibliographical references and index.
ISBN 0-262-06253-4 (hard: alk. paper)
1. Language and languages—Philosophy. 2. Psycholinguistics. 3. Neurobiology.
4. Cognitive science. I. Title.

P107.F445 2006
401—dc22

2005058045

10 9 8 7 6 5 4 3 2 1

Preface

I hear and I forget
I see and I remember
I do and I understand
—Attributed to Confucius, 500 BCE

Many years ago, I was browsing through books on learning how to draw. One of them, said, after a brief introduction, put down this book and start drawing. This book is like that—it will frequently suggest a simple mental exercise to help you personally experience a phenomenon. If this appeals to you, you might like the book.

By now, virtually everyone agrees that the scientific explanation for human language and cognition will be based on our bodies, brains, and experiences. The major exception is Noam Chomsky whose dominance of twentieth-century linguistics is unparalleled in any other academic field. I will later quote from Chomsky's 1993 book, *Language and Thought*, and he repeatedly stated the same idea in his 2003 Berkeley lectures: "We don't know nearly enough about the brain for cognitive science to take it seriously." Chomsky has focused on linguistic *form*; since this book deals first with *meaning*, we won't encounter him again until chapter 22.

As a first mental exercise, try expressing to yourself what you know about how your own thoughts work. *How do our brains compute our minds?* When I ask Berkeley students, on the first day of class, to write a page on this question, most of the students express mystification. Even people who know a great deal about neuroscience, psychology, linguistics, philosophy, and artificial intelligence often have no clear idea of how the findings of these fields could combine to yield even a preliminary understanding of how language is embodied in us.

This book proposes to begin integrating current insights from many disciplines into a coherent *neural theory of language*. It might seem that no such effort is needed. Isn't language obviously a function of our brains—what else could it be? Certainly other human abilities such as motor control, hearing, and especially vision have been studied as neural systems for many decades. But language is still often treated as an abstract symbol system not particularly tied to human brains or experience.

A great deal of permanent value has been learned from formal studies of language, but it is surprising that the notion of disembodied language persists. This is partly an historical artifact, but it also arises from the fact that other animals share our visual and motor abilities but not our language skills. Much of the progress in neural theories of vision and motor control have come from invasive animal experiments that are thankfully prohibited on people. Until recently, very little has been known about how our brains process language.

Currently no one knows the details of how words or sentences are processed in the brain, and there is no known methodology for finding out. Many scientists believe it is premature (perhaps by centuries) to formulate explicit theories linking language to neural computation. Even theoreticians are usually content with suggestive models, which can't actually be right, but do suggest interesting experiments. However, the cognitive sciences reveal a great deal about how our brains produce language and thought. And we have a long and productive tradition, going back at least to the Greek atomic theories of matter, of postulating "bridging theories" in advance of the detailed evidence. Brian Greene's *The Elegant Universe* offers a wonderful description of the fundamental nature of matter, though science might never deliver experimental verification.

In contemporary science, it is not unusual to have quite extensive knowledge at both ends of a causal chain and to build and test theories to explicate the bridging links. For example, astrophysics is concerned with linking fundamental particle physics with astronomy. In economics and other social sciences, a principle concern is how individual preferences give rise to group behavior. Similarly, much of molecular biology is concerned with how genetic material yields the various proteins and resultant organisms. Higher levels of biology also try to develop bridging theories. We can see the search for a neural theory of language as one such attempt, albeit an unusually ambitious one. These bridging theories are often developed as computer simulations, and this book follows this tradition.

I treat the mind as a biological question—language and thought are adaptations that extend abilities we share with other animals. For well over a century, this has been the standard scientific approach to other mental capacities such as vision and motor control. But language and thought, even now, are usually studied as abstract formal systems that just happen to be implemented in our brains. Instead, we pursue the great ethologist Nico Tinbergen's (Tinbergen 1963) four questions that must be asked of any biological ability:

1. How does it work?
2. How does it improve fitness?
3. How does it develop and adapt?
4. How did it evolve?

The first three of these questions are covered in considerable detail. The origins of language are still largely unknown and are discussed briefly in chapter 26.

There is a sufficiently large gap between brain and language to contain ecological niches for many theories, especially if their proponents are satisfied to ignore inconvenient findings. Understanding language and thought requires combining findings from biology, computer science, linguistics, and psychology. A theory that seems perfectly adequate from one perspective may contradict what is known in another field. Problems that seem intractable in one discipline might be quite approachable from a different direction. Taking all the constraints seriously is the only way to get it right.

But this requires us to understand the essential ideas from several quite different scientific domains. In any of these fields, keeping up with technical advances and doing original work are extremely demanding pursuits and require focused effort. There are some endeavors at the boundaries between subfields, but very little scientific work that attempts to encompass the full range needed for our task. I will need to synthesize a bridging theory from separate fields, all of which have their focus elsewhere. My approach is to pick out key findings and theories from various disciplines and show how, in combination, they constrain the possible bridging theories of language to a narrow family of possibilities.

Each discussion is an oversimplification of some research field, often involving thousands of active investigators, and thus is inherently incomplete. The usual references suggest more detailed discussions of various points, but these are most useful as key words for search engines. By the time you read this book, important new developments will have occurred in each of these areas. Books for further reading are included for people who would like additional background in one or another direction.

While we are far from having a complete neural theory of language, enormous scientific advances have occurred in all the relevant fields. Taken together, these developments provide a framework in which everything we know fits together nicely. The goal of this book is simple: I would like you, at the end, to say, *This all makes sense. It could explain how people understand language*. I will make no attempt to convince you other theories are wrong—in fact, I assume that most of them are partially right. The book can be seen as part of a general effort to construct a Unified Cognitive Science that can guide the effort to understand our brains and minds. I try to present a story here that is consistent with all the existing scientific data and that also seems plausible to you as a description of your own mind.

Except for one thing. One part of our mental life is still scientifically inexplicable—subjective experience. Why do we experience everything in the way we do? The pleasure of beauty, the pain of disappointment, and even the awareness of being alive . . . these do not feel like they are reducible to neural firings and chemical reactions. Almost everyone believes that his or her own personal experience has a quality that goes beyond what this book, and science in general, can describe. If I had anything technical to say about subjective experience, it would be the highlight of the book, to say the least.

People use terms like *personal experience*, *subjective experience*, and *phenomenology* to label this idea. Philosophers have coined a technical term, *qualia*, to refer to these phenomena that are currently beyond scientific explanation. Antonio Damasio (Damasio 2003), who in my opinion is doing the best scientific work on subjective experience, distinguishes measurable *emotions* from subjective *feelings*. Aside from a brief discussion in chapter 26, this book focuses on what can be learned from studying the physiological and behavioral correlates of experience—that is, what can be measured and modeled objectively.

My undertaking of this quixotic enterprise came as the result of a year of explicit soul-searching around the time of my fortieth birthday. I had the good luck of entering the field of computer science in its infancy, and I believed this gave me the opportunity to move in almost any direction, exploiting insights into information processing not available to previous generations. My long-term interests in language and the brain and work on various computer systems including some of the earliest robots, led me to focus on the question that I just asked you—How does the brain compute the mind? Twenty-five years later, due to advances in all fields that were inconceivable to me at the time, the outlines of an answer seem to be emerging.

A Brief Guide to the Book

This book is designed to be read in order; each chapter provides some of the underpinnings for later ideas. But it should also be possible to look first at the parts that interest you most and then decide how much effort you wish to exert. Many forward and backward pointers are included to help integrate the material.

Information processing is my organizing theme. Language and thought are inherently about how information is acquired, used, and transmitted. Chapter 1 lays out some of the richness of language and its relation to experience. The central mechanism in my approach to the neural language problem is neural computation. Chapters 2 and 3 provide a general introduction to neural computation. Chapters 4 through 6 provide the minimal biological background on neurons, neural circuits, and how they develop. We focus on those properties of molecules, cells, and brain circuits that determine the character of our thinking and language.

Chapters 7 and 8 consider thought from the external perspective and look at the brain/mind as a behaving system. With all of this background, chapter 9 introduces the technical tools that are used to model how various components of language and thought are realized in the brain. A fair amount of mechanism is required for my approach, which involves building computational models that actually exhibit the required behavior while remaining consistent with the findings from all disciplines. I refer to such systems as *adequate* computational models, which I believe are the only hope for scientifically linking brain and behavior. There is no guarantee that an adequate model is correct, but any correct model must be adequate in the sense defined here.

The specific demonstrations begin with a study of how children learn their first words. This involves some general review (chapter 10) and a more thorough study of conceptual structure (chapter 11) needed for word learning. The first detailed model is presented in chapter 12, which describes Terry Regier's program that learns words for spatial relation concepts across languages. This theme of concrete word learning is then extended to cover words for simple actions in chapters 13 and 14, which describes David Bailey's demonstration system.

The next section extends the discussion to words for abstract and metaphorical concepts. In chapters 15 and 16, we look further at the structure of conceptual systems and how they arise through metaphorical mappings from direct experience. Chapter 17 takes the informal idea of understanding as imaginative *simulation* and shows how it can be made the basis for a concrete theory. This theory is shown in chapter 18 to be sufficiently rich to describe linguistic aspect—the shape of events. This is enough to capture the direct effects of hearing a sentence, but for the indirect consequences, we need one more computational abstraction of neural activity—belief networks, described in chapter 19. All of these ideas are brought together in Srinivas Narayanan's program for understanding news stories, discussed in chapter 20.

Chapters 21 through 25 are about language *form*, that is, grammar—how grammar is learned and how grammatical processing works. Chapter 21 lays out the basic facts about the form of language that any theory must explain. Chapter 22 is partly a digression; it discusses the hotbutton issues surrounding how much of human grammar is innate. We see that classical questions become much different in an explicitly embodied neural theory of language and that such theories can be expressed in standard formalisms (chapter 23).

Chapter 24 shows how the formalized version of neural grammar can be used scientifically and to build software systems for understanding natural language. The poster child for the entire theory is Nancy Chang's program (chapter 25) that models how children learn their early grammar—as explicit mappings (constructions) relating linguistic form to meaning. Chapter 26 discusses two questions that are not currently answerable: the evolution of language and the nature of subjective experience. Finally, chapter 27 summarizes the book and suggests that further progress will require a broadly based unified cognitive science. But the scientific progress to date does support a range of practical and intellectual applications and should allow us to understand ourselves a bit better.

A version of the material in this book has been taught to hundreds of undergraduate students at the University of California, Berkeley over the years. There were weekly assignments, and most of the students actually did them. The course did not work for all the students, but a significant number of them came out of the class with the basic insights of a neural theory of language. If you want to understand how our brains create thought and language, there is a fair chance that this book can help.