MARK SEIDENBERG

LANGUAGE AT THE SPEED OF SIGHT

HOW WE READ, WHY SO MANY CAN'T, AND WHAT CAN BE DONE ABOUT IT

LANGUAGE AT THE SPEED OF SIGHT

How We Read, Why So Many Can't, and What Can Be Done About It

MARK SEIDENBERG

BASIC BOOKS
New York

Published in the United States by Basic Books, an imprint of Perseus Books, a division of PBG Publishing, LLC, a subsidiary of Hachette Book Group, Inc.

Text design by Jeff Williams

Library of Congress Cataloging-in-Publication Data

Names: Seidenberg, Mark S., author.
Title: Language at the speed of sight : how we read, why so many can't, and what can be done about it / Mark Seidenberg.
Description: New York : Basic Books, an imprint of Perseus Books, a division of PBG Publishing, LLC, a subsidiary of Hachette Book Group, Inc., [2017] | Includes bibliographical references and index.
Identifiers: LCCN 2016046731| ISBN 9780465019328 (hardcover) | ISBN 9780465080656 (ebook)
Subjects: LCSH: Reading (Higher education) | Language experience approach in education. | Cognition disorders. | Psycholinguistics.
Classification: LCC LB2395.3 .S44 2016 | DDC 428.4071/1—dc23 LC record available at https://lccn.loc.gov/2016046731

10 9 8 7 6 5 4 3 2 1

CONTENTS

Reading, Writing, and Speech

How We Read

The Educational Challenges

READING, WRITING, AND SPEECH

CHAPTER 1

The Problem and the Paradox

I DON'T THINK WE'VE MET, but I know two things about you. One is that you are reading these words. And because you're reading these words, I also know that you are an expert at what this book is about. Which is reading.

A friend once gave me a scholarly 380-page book about the history of the pencil. I like pencils. I'm good at using them. I like a fresh pink eraser tip. I just do not need to read about how they got that way. Perhaps reading is for you like the history of pencils is for me. Why read about reading?

Reading is one of the few activities you do every day whether you want to or not. Street signs, menus, e-mails, Facebook posts, novels, ingredients in Chex Mix. You read for work, for school, for pleasure; because you have to, because you want to, because you can't help it. That is a lot of practice over a long period. If it takes thousands of hours to become an expert at something like chess, we readers are in grandmaster territory.

One result is that everyone has an opinion about how they read. When you are introduced at the dinner table as someone who studies reading for a living, you hear all the stories. People tell me whether they read word by word or in big chunks. Whether they read visually or hear the sounds of words in their heads while they read. How they learned to read and how their kids learned. Many confessions from people who say they are slow readers, more often than not from accomplished people in fields that seemingly demand a high level of reading competence—lawyers or school superintendents, say. Most of you think that other people read faster than you. It's the opposite of Lake Wobegon: reading is the land where the folk are all below average.

But here's the rub: people manage to be good at reading without knowing much about how they do it. Most of what goes on in reading is subconscious:

we are aware of the result of having read something—that we understood it, that we found it funny, that it conveyed a fact, idea, or feeling—not the mental and neural operations that produced that outcome. People are unreliable narrators of their own cognitive lives. Trying to understand reading by observing our own reading is hopeless, like trying to understand how a television works by watching *Game of Thrones*. Being an expert reader doesn't make you an expert about reading. That is why there is a science of reading: to understand this complex skill at levels that intuition cannot easily penetrate.

I am a psychologist/psycholinguist/cognitive neuroscientist who has been studying reading since the disco era. I'm not alone: a huge community of scientists studies reading around the world. Many people are surprised to learn there is a science of reading. Really, what is there to study? Words on page; eyes scan words; words are comprehended. It's just like listening, only visual. Book ends here. Beneath this seemingly simple behavior, however, a vast, coordinated network of activities is occurring. A snapshot of a person's brain activity while reading shows that most of the brain is involved: areas involved in vision and language, of course, but also neural systems that control action, emotion, and decision making; several memory systems; and much of the rest. Forget the myth that people only use 10 percent of their brains: we use more than that merely reading. The relationship between the experience of reading and its underlying neurocognitive mechanisms is about as opaque as the relationship between behavior and its psychodynamic causes in Freudian theory. Fortunately we have better methods than Freud for exploring the subconscious basis of behavior. We'll have you lie down and talk to us, but in the barrel of a magnetic resonance imaging machine, not on a couch.

We've learned quite a lot actually.

We understand the basic mechanisms that support skilled reading, how reading skill is acquired, and the main causes of reading impairments.

We know which behaviors of three- and four-year-olds predict later reading ability.

We know how children become readers during the first years of schooling and the obstacles that many encounter.

We know what distinguishes good from poor readers, younger from older skilled readers, and typical readers from those who are atypical because of constitutional factors (such as a hearing or learning impairment) or environmental ones (such as inadequate instruction or poverty).

We understand what is universal about reading (things that all readers do the same way because their brains are essentially alike) and what is not (because writing systems differ).

We have identified the main neural circuits involved in reading and some of the anomalous ways they develop in children with reading impairments.

We even have computational models of learning to read, skilled reading, dyslexic reading, and the loss of reading ability due to brain injury. It takes a deep understanding of these phenomena to develop models that reproduce them.

This vast research base has led to the development of methods that can reliably help many children who struggle to read. Researchers disagree about many details—it's science, not the Ten Commandments—but there is remarkable consensus about the basic theory of how reading works and the causes of reading successes and failures.

Sputnik Lands on USA

The successes of reading science create a paradox: if we know so much about reading, why are literacy levels in the United States so low?

In 2011 America had what our president called a Sputnik moment, occasioned by the release of the latest results from the Programme for International Student Assessment (PISA), a massive appraisal of the reading abilities of fifteen-year-olds in seventy-four countries and municipalities from Kyrgyzstan to Canada. As in previous rounds and again in 2012, US performance was close to the average for the thirty-four member countries of the Organisation for Economic Co-operation and Development (OECD), which conducts the PISA exercise. However, the 2009 round was the first to include data from Shanghai and Singapore, which scored higher than the US, as did Asian neighbors Korea, Hong Kong, and Japan. The Shanghai students lapped everyone else, scoring highest in reading, math, and science by wide margins. These results received far more attention than the fact that the US always scores lower than countries like Australia, Canada, Finland, and New Zealand in such exercises. Government officials and commentators treated the results as a wake-up call about the uncompetitive state of American education, said to pose a threat to the country's future akin to that represented by Sputnik in 1957. The Soviet Union's stunning success spurred a rapid, comprehensive governmental response that seems as much a relic of a bygone era as the satellite itself. Within two years, legislators had created the National Aeronautics and Space Administration and the precursor of the Defense Advanced Research Projects Agency (which funded the development of breakthrough technologies including the Internet), tripled the National Science Foundation budget, and allocated hundreds of millions of dollars for student loans and scholarships under the National Defense Education Act. But 2011 was not 1957, and the second Sputnik moment passed quickly, rapidly dropping out of public discourse (Figure 1.1).

Although the PISA results made the news, there is plenty of in-house data about America's literacy issues. The country is a chronic underachiever. A 2003

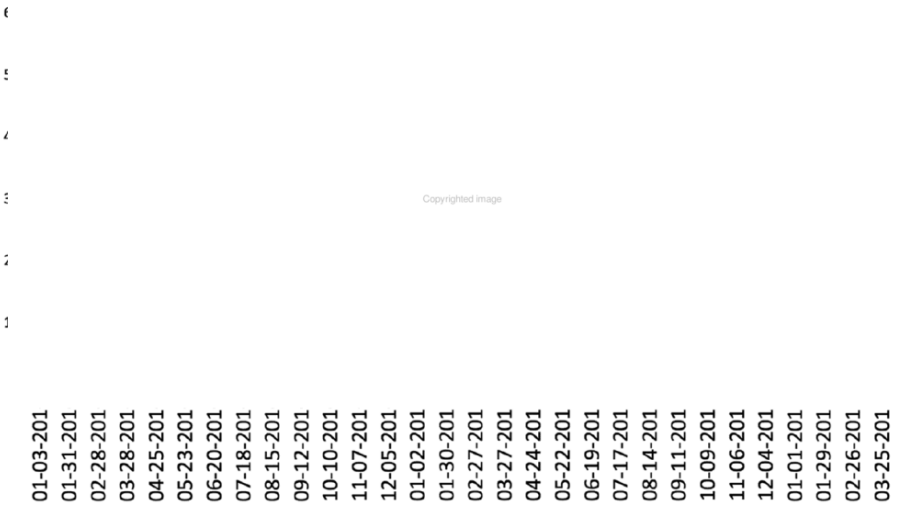

FIGURE 1.1. The number of times the phrase "Sputnik moment" was uttered on CNN, the American news network, over a two-year period. The large spike followed the release of the 2009 PISA results and coincided with the president's State of the Union address.

study found that about 93 million adults read at basic or below basic levels. At those levels, a person might be able to follow the instructions for mixing a batch of cake mix but not understand a fact sheet about high blood pressure. The emergence of our literacy problem is visible in the performance of fourth and eighth graders on the National Assessment of Educational Progress (NAEP), "the Nation's Report Card," an assessment administered by the US Department of Education. Over half the children have scored at basic or below basic levels *every time* it has been administered. At the upper end, we are turning out fewer highly proficient readers than expected given our economic resources. Like everything else about education in the US, the results of assessments like the PISA and NAEP have generated controversy. Are the tests too hard? Are they poorly matched to what American children are taught? If the scores are so low, why does the US continue to have the biggest national economy in the world? The American polity has been in a test-happy phase, and much other data about who can read and how well paints a consistent picture: large numbers of individuals in the US read poorly, as has been true for many years. Although I focus here on the US because it is home, the situation is similar in many other countries with advanced economies.

The consequences of marginal literacy for the affected individuals and for society are vast, as we all know. Reading is fundamental—so the slogan

goes—and children's failures to acquire reading skills have rapidly cascading effects on learning other subjects. Even math is implicated, given the heavy emphasis on language in math curricula and instruction (word problems, explaining your work, the practice of embedding math problems in real-world contexts). Having less reading experience makes it harder to learn how to learn and how to think critically and analytically. Reading failures place children at high risk of falling out of the educational system. As adults, poor readers cannot participate fully in the workforce, adequately manage their own health care, or do much to advance their own children's education. If that doesn't convey sufficient urgency, a 2012 report from the Council on Foreign Relations declared, "Large, undereducated swaths of the population damage the ability of the United States to physically defend itself, protect its secure information, conduct diplomacy, and grow its economy."

This information plagues me, especially because the situation is not new. Scores on standardized reading tests have been nearly flat for decades. Hopes that the *Harry Potter* craze would increase children's involvement in reading gave way to deep disappointment when it didn't. We hear the suspicion voiced everywhere—from Ivy League colleges to local middle schools—that important reading skills and habits not measured by standardized tests, such as the ability to engage in close reading of challenging texts, are in active decline. According to Philip Roth, that game is already over. "The evidence is everywhere that the literary era has come to an end. . . . The evidence is the culture, the evidence is the society, the evidence is the screen, the progression from the movie screen to the television screen to the computer. . . . Literature takes a habit of mind that has disappeared."

Anxiety about reading achievement underlies endless debates about how reading should be taught. Parents know it is essential that their children learn to read well, but lacking confidence that schools can do the job, they are driven to seek additional help from tutors or commercial learning centers, if they can afford it. Our knowledge about reading has grown enormously, but, as ever, many people cannot read, or can read only poorly, or are able to read but avoid it. And so I have asked myself whether our science has anything to contribute to improving literacy outcomes in this country and elsewhere.

It might not. Reading failures could simply be collateral damage caused by other, deeper problems, such as poverty. About 15 percent of the US population lives in poverty, as the US Census Bureau defines it, including over 16 million children. Poverty is associated with higher infant mortality, higher risk of atypical neurodevelopment, shorter life span, worse health and health care, higher crime and incarceration rates, lower educational achievement, higher dropout rates, poorer schools with weaker teachers, and, at the bottom of a

much longer list, poor reading. Surely addressing poverty would have a bigger impact on literacy than anything inspired by our research.

Poverty is a huge factor in America, as in other countries. Reading achievement is higher in Australia, Canada, and New Zealand than in the US, but not among their aboriginal peoples, for whom poverty rates are high. No discussion of what to do about reading can ignore the noxious conditions that govern many people's lives. However, 93 million is a lot of marginal readers, and they are not confined to the lowest economic stratum. There's Ron Gronkowski, a colorful professional football player, who said he hadn't read a book since *Mockingbird to Remember* in ninth grade. But look around: they walk among us. We could be talking about an airport security agent, an X-ray technician, police officers. High school graduates who cannot handle the literacy demands of community college. "Student athletes" who play big-time college football. "Low-information voters" whose primary news sources are talk radio and cable news. Or perhaps some readers of this book, highly motivated yet struggling through it. People with only basic reading skills can hold demanding jobs that affect health care, personal safety, law enforcement, business, politics, and every other element of society.

How much of the problem is due to cultural characteristics that conspire against our becoming readers, such as the profusion of screen-based activities: 24/7 movies and television, gaming, apps, cat and music videos, PewDiePie (45 million YouTube subscribers, 2,900 videos)? Are reading and writing gradually being reduced to tools in the service of reading-ish activities such as texting, tweeting, and Facebooking? What about overscheduling: the relentless demands of ballet, piano, violin, drama, tae kwon do, 6 a.m. slots at the hockey rink, soccer practice, matches, and travel? Middle schoolers cannot read while running dribbling and passing drills. Perhaps reading is simply less important than it used to be. Writing was the first information technology, but now there are others. We have screens, they have pictures, sound, video. Text carries less of the communicative load than before. I greatly prefer to read a recipe, but a video demonstration conveys additional information, for people who have the patience to sit through it. You can read about the phenomena I describe in this book but also see some of them in videos or try them out, which I highly recommend (and provide links to allow).

Something is happening here, but we don't know what it is, yet. If you position your ear close to the Internet, however, you can hear the steady thrum of reports and pronouncements about the state of reading: the good (easier access to a larger quantity and greater variety of texts than ever, integration of text with sound and image), the bad (easier access to a larger quantity and greater variety of nontextual media and undemanding text), and the ugly (a

long-term decline in the linguistic complexity of K–12 schoolbooks). The National Endowment for the Arts issues grim warnings about the decline of reading, while educators view reading as one of several "literacies" for the modern era.

The results of this real-world experiment in how information is transmitted will not be known for generations. I think it is safe to assume that reading is not going away any time soon, if ever, and that people who are skilled readers will continue to have advantages over people who are not. With this in mind, I focus on a factor that has direct, lasting effects on who reads and how well: education. With some exceptions, children learn to read in school. The only certain way to obviate low literacy is prevention: successfully teaching children to read in the first place. Would more people be better readers if they had been taught differently? How much does schooling affect how well children read and, with it, their engagement in reading?

Here we encounter a problem. There is a profound disconnection between the science of reading and educational practice. Very little of what we've learned about reading as scientists has had any impact on what happens in schools because the cultures of science and education are so different. These cross-cultural differences, like many others, are difficult to bridge.

The gulf between science and education has been harmful. A look at the science reveals that the methods commonly used to teach children are inconsistent with basic facts about human cognition and development and so make learning to read more difficult than it should be. They inadvertently place many children at risk for reading failure. They discriminate against poorer children. They discourage children who could have become more successful readers. Many children who manage to learn to read under these conditions wind up disinterested in the activity. In short, what happens in classrooms isn't adequate for many children, and this shows in the quality of this country's literacy achievement. Reading is under pressure for other reasons, but educational theories and practices may accelerate its marginalization.

Reading science cannot address all of the issues that affect reading achievement—all those sequelae of poverty, for example—but it does speak to serious issues about how reading is approached in schools and at home. A look at the basic science suggests specific ways to promote reading success. These do not require more testing or new federal laws; they do not require vast infusions of money; they are not based on classroom computers that treat learning like a video game or other faddish uses of technology. What they require is changing the culture of education from one based on beliefs to one based on facts.

The next several chapters provide an insider's guide to some beautiful science that looks at reading at levels ranging from the movement of one's eyes

across the page to the brain circuits that support reading to genes that affect how those circuits develop. Some of it is paper-clip-and-string research using the simplest of methods, such as recording the errors that children make when they read aloud or having the subjects in an experiment decide if SUTE is a word (it isn't, but it sounds like one, which creates momentary uncertainty even when it is read silently). Some of the research involves exotic methods such as transcranial magnetic stimulation, which briefly disrupts neural processing in parts of the brain relevant to reading. We will look at children who struggle to read and children who can read words but comprehend very little. Of course there is the terrible fascination of patients whose brain injuries impair reading in unusual ways. The late Oliver Sacks's famous agnosic patient mistook his wife for a hat; a patient with an analogous reading impairment, called deep dyslexia, read the word SYMPATHY as "orchestra." Along the way I cheerfully destroy a few myths about reading, including ones that have supported years of commercial schemes said to improve your child's reading in a few short weeks or to make you a super-mega-reader who can whip through books a page at a glance. We will look at educational neuroscience, the study of the brain bases of skills such as reading and math and how they are acquired. Fact: in neuroimaging studies, poor readers show atypically low activation in a part of the brain that processes the spellings of words. Can such findings inform how reading is taught, or do they merely provide neural window dressing for something teachers already know, that poor readers don't read or spell words very well?

With an understanding of the basic science of reading in hand, I take a close look at what happens in classrooms in America. If you are curious about the state of reading, these findings will, at the least, surprise you. If you are (or were) a parent of a school-age child, you might feel the shock of recognition and the dropping of scales from your eyes.

Understanding how the science relates to educational practices is an important step, but it is not enough to generate change. Educators are deeply immersed in their own worldview and well defended against incursions from outside. The education side focuses on "literacy" (literacy practices, cultural attitudes toward literacy, multiple types of literacy including ones that do not involve print), not reading. The scientific perspective is seen as sterile and reductive, incapable of capturing the ineffable character of the learning moment or the chemistry of a successful classroom. That people who enter the field of education do not gain exposure to modern research in cognition, child development, and cognitive neuroscience deprives them of the ability to evaluate what is studied, how it is studied, what is found, and what it means. Instead, they are exposed to the views of a few authorities, the most

influential being Lev Vygotsky, who lived in the Soviet Union, wrote in Russian, died in 1934, and never saw an American classroom or a television, calculator, computer, video game, or smartphone. We scientists naively take the importance of our findings and their implications for classroom practice to be obvious, overestimating how much people on the education side know about or value them.

Education as a discipline has placed much higher value on observation and hands-on experience, which brings us back to page one (of this book). Intuitions about reading do not penetrate very far, and observations are influenced by prior beliefs, including what one has been taught about how children learn and how reading works. Educators are people too, subject to the same cognitive limitations and biases as everyone else. The lack of scientific literacy, combined with deep faith in the validity of personal observation, creates vulnerability to claims that are intuitively appealing but unproved or untrue.

People complain about the state of education about as much as they complain about the government. There's an endless, hopeful, desperate search for a game-changing innovation that will make more people smarter, more quickly, with less effort and expense. Some people focus on eliminating poverty itself, which would obviously have enormous impact on education and much else. We can focus at the same time on how to modify more tractable conditions to produce better outcomes. Learning more about the values and beliefs of the educators who teach teachers, design curricula, and create instructional practices could be a powerful impetus to change. In the closing chapters I spell out these concerns as they have arisen in reading education. This analysis clarifies what needs to change, and I discuss some ways we could move forward.

Since change can't be achieved overnight, I've also written this book to give readers—a parent interacting with a teacher, a school principal, or the bureaucrats who run the local school system; a voter whose representatives have the power to affect educational goals and practices, how public education gets funded, and how tax dollars are spent—information they can use now. These are the tools you need: an explanation of what we know about skilled reading, learning to read, the causes of reading difficulties, and the brain bases of normal and impaired reading; a critical analysis of the ways that reading is taught; an account of the controversies in reading education; and an understanding of educational ideology and how it evolved. Like health care, education is a multi-billion-dollar industry involving multiple stakeholders—government, business, educators, parents, children, taxpayers, unions, interest groups, philanthropy—whose perspectives and interests often conflict. In education, as in health care, end users benefit if they are informed, proactive participants rather than passive recipients.

Of course plenty of information about reading is available online, but on the Internet nobody knows your website is a dog. Websites do not come labeled with independent scientific seals of approval; doctrinal biases are not labeled as such; a layperson has no way to evaluate contradictory claims and assertions. Everyone—ranging from the National Council of Teachers of English (not much science there) to the Scientologists (whose teaching method L. Ron Hubbard is said to have devised himself)—has an opinion about reading. School districts post jargon-laden statements of teaching philosophies and curricular goals on their websites. It's a jungle out there, and a person could use some help.

A Science of Reading

Reading is one of the oldest topics in experimental psychology—the first American professor of psychology, James McKeen Cattell, studied it in the 1890s—and yet it is also a favorite topic of researchers in cognitive neuroscience, a field that emerged about a hundred years later. Why this enduring interest in such a familiar activity?

Reading is unique. Reading is among the highest expressions of human intelligence. Although spoken language is usually taken as the capacity that distinguishes people from other species, researchers have debated the degree to which other species's communication systems resemble language. No other species has a linguistic capacity equal to ours, but animal communication systems share some properties with human language. The late African Grey parrot Alex clearly had communicative interactions with his longtime trainer Irene Pepperberg. Was his use of words very much like human speech or an oddly evocative simulation, the result of thousands of hours of intense training? Whatever the answer to that question, we know that no other species has an ability remotely like reading. Indeed, *Homo sapiens* didn't either until the invention of writing about five thousand years ago. Understanding this complex skill means understanding something essential about being human.

Reading is important. Human culture has evolved to the point where this skill is critical to our ability to thrive. For most of human history people were illiterate and yet functioned well enough. That is not the case in modern societies. In a 2007 article bemoaning Minnesota children's low levels of reading achievement, Garrison Keillor observed, "Teaching children to read is a fundamental moral obligation of the society." Given

its importance in modern culture, could anyone disagree? Unfortunately, reading is not always an easy skill to acquire; for most children it requires instruction, and for all of us it entails years on task. How children are taught matters a great deal: it can affect whether they become readers or not, their level of reading skill, and the extent to which they enjoy and seek out the experience. In order to teach children effectively and make this essential skill available to as many people as possible, we need to know how reading works. As a scientist I mainly study reading in order to understand the parts and how they fit together; it's a puzzle that interests me deeply. However, my own moral obligation is to pursue how the research can serve to help more children become better readers. This book is a part of that effort.

Reading is a tool for understanding human cognition. The capacity to use language evolved in humans over many thousands of years, the end result being that children acquire it easily and rapidly through interactions with other speakers. Reading is different: it is a technology, like radio, that came into existence because a person—or possibly several—had the insight to invent it. The advent of reading occurred relatively recently in human history, well after humans had evolved capacities to speak, think, perceive, reason, learn, and act. Reading was a new tool created out of existing parts. The fortuitous by-product of this history is that we can use reading to investigate all these capacities. A person doesn't have to be a reading scientist to study reading; they might study vision or memory, for example, using experimental methods that happen to involve having people read words and sentences. This bonus has resulted in the creation of a research literature of exceptional depth and quality.

In short, reading is interesting. It's complex, it's essential, and there is an urgent need to reduce the number of people who read little or not at all and to ensure that future generations will be sufficiently literate to thrive in the world they will inhabit. Out of such elements emerged a science of reading and this book.

CHAPTER 2

Visible Language

WE READ WITH OUR EYES, but the starting point for reading is speech. Informally we think of reading and speech as different ways of representing the same thing. Spoken language, visual language: one is a transposition of the other, just as numbers can be written in decimal or binary. In fact writing and speech are not interchangeable, but they are closely intertwined, each deeply affecting the other, like a couple of linguistic codependents with serious boundary issues.

Reading is secondary to spoken language: there are no natural languages (the term is used to distinguish human language from bee language, body language, markup language, and every other metaphoric extension of the term) that have a written form but not a spoken one. The capacity for spoken language evolved in humans well before writing was invented. For most of human history spoken languages existed without written forms, and many contemporary languages still do. Children's development recapitulates this pattern: their use of spoken language invariably precedes reading and writing.

For linguists who study the formal properties of languages, reading is about as interesting as Morse code. The creation of writing piggybacked on speech, which is where the core, defining characteristics of language are found, they say. Though a brilliant technological innovation of enormous practical importance, writing did not fundamentally change the nature of language. This view is simplistic, however. The origins of reading are one thing, but how it functions in modern literate cultures and in the brains of modern literate individuals is another. It is not merely that written language is supremely important, although of course it is—modern societies exist because of it and could not survive without it—but also that the technology afforded new ways

to use language. It matters little whether reading "fundamentally" changed the nature of language or was merely the most important aftermarket add-on in history. Any account of language that excludes reading ignores what language has become because of it.

The intimate but complex relationship between written and spoken language underlies many of the important scientific questions and educational controversies about reading and so will emerge as one of the major themes of this book. Consider:

- The beginning reader's initial task is to learn how the spoken language they know relates to the written code they are learning. This step can be difficult for many children precisely because the two codes are not simple transpositions of each other.
- Because writing systems represent spoken languages, many educators concluded that learning to read should proceed just like learning one's first language. In fact, no one learns to read the way they learned to talk, thus occasioning the question, What happens when several generations of children are taught to read under this mistaken assumption?
- Debates about how to teach reading have focused on whether to promote or discourage the use of sound-based information in silent reading. Phonics methods are controversial because they emphasize connections between print and speech.
- Dyslexia, the developmental reading disorder, involves impairments in relating the written and spoken forms of language.
- Children's progress in learning to read is greatly affected by their experience with spoken language: the quantity and variety of the speech to which they are exposed and how the spoken language or dialect used in the home and community compares to the one used in school.

Whereas learning to relate print to speech is the beginning reader's first challenge, describing how the two are related is mine.

Each of These Things Is Not Like the Other

Yes, spoken and written language are alternative ways to represent the same thing. This observation would be too obvious to mention were it not for the cases where it does not hold. A prominent example is hearing-impaired people born to hearing-impaired parents who learn to sign rather than speak. Sign

languages such as American Sign Language (ASL) exhibit the same types of grammatical structures and communicative functions as spoken languages, proving that the capacity for language can be realized using another modality. However, the grammar of ASL is radically different from the grammar of English. Because texts are not written in ASL, deaf individuals who learn ASL as their first, "spoken" language must learn a second language, English, for reading and writing. Using one language for conversation and a second language for reading? Not easy.

The other way to decouple spoken and written language is neuropathology, such as a stroke or Alzheimer's disease. In the absence of pathology, a person who can read and understand a sentence such as JOHN GAVE THE BOOK TO MARY can do the same when it is spoken. Under some deeply unfortunate conditions, comprehension is disrupted in one modality while leaving it relatively intact in the other. Some patients can understand words such as BOOK or GAVE when they are spoken but not written, and others show the opposite pattern. These individuals are of great interest to researchers precisely because such disruptions in the marriage of speech and reading are exceptional: it takes *brain damage* to cause them.

Spoken and written language share several other basic characteristics, two of which are of particular interest because they are true but didn't have to be. First, both print and speech are comprehended rapidly and automatically. You, reader, are understanding this sentence right here as you move along, and the same would be true if I were saying it to you. Language comprehension in both modalities is relentless and almost unstoppable, short of shutting one's eyes while reading or saying "lalalalala" while listening. It is not that comprehension works without error or that the result is the same for everyone. Rather, comprehension waits for no one, closely trailing the flow of words whether written or spoken. It is impressive that our brains work this rapidly and surprising that the two modalities behave so much alike, given their many intrinsic differences. Sign languages are also understood in this immediate, online manner.

The other important similarity is that people's reading and spoken-language skills tend to be very closely matched. For adults, the positive correlation between them is about 0.9 (the maximum is 1.0). Dyslexics are exceptional because their reading is much poorer than their spoken language. Some texts are tough going (I'm thinking of you, Henry James and Stephen Hawking), but the audiobooks aren't *Winnie the Pooh*.

There are plenty of books about the nature of spoken language, and so if reading and speech were merely very similar, this one could end right here. Once we look beyond these broad commonalities, however, they differ at every turn.

Speech evolved in the species.

Reading is a cultural artifact, like money.

Speech is universal: in the absence of pathology, everyone learns to talk.

Reading is like Wi-Fi: only some people have it.

Children learn a spoken language through interactions with other language users.

Reading is taught, beginning with alphabet songs and bedtime stories and continuing through several years of schooling.

Speech is fast fading: the signal is gone once it is produced.

Writing systems were created as a way to transcend the impermanence of speech. This text is not disappearing as you read it.

Speech is messy. Producing a coherent utterance is a complex action: deciding what to say, picking the words and grammatical structures that express the intended meaning, loading a program to articulate the sequence of words, and then running the program, all done on the fly. The process is so demanding that the end product is littered with dysfluencies and errors.

Writing is cleaner than speech because it isn't produced under the same time constraints and can be edited. Speech is more like a rough draft, text like the corrected version.

That one has some obvious exceptions. Some formal modes of speaking are grammatically immaculate, and informal writing can be messy. A television performer such as Stephen Colbert is a professional talker skilled at producing fluent, well-formed utterances most of the time. Conversely, there is Wikipedia, the Great Wall of Words, a system for creating texts that are full of grammatical, spelling, and usage errors.

Finally, speech and reading are opposites with respect to who controls the rate at which comprehension can occur. Readers decide how rapidly to proceed. Listeners are at the mercy of speakers performing the difficult task of producing coherent, mostly fluent utterances. In exchange for ceding control to the speaker, the listener gains one advantage: speech affords more opportunities for clarification. Circumstances permitting, speakers can be asked about what they have said, whereas readers cannot query authors. (We have book tours and Reddit Ask Me Anythings for that.)

The net result is that properties of spoken and written language diverge. The distributions of words in speech and text differ; many words occur more often in one than the other. For example, which of the following words occur more often in speech and which more often in text?

```
yeah such right actually obviously thus sh*t which
like diverge*
```

* The examples alternate between more common in speech (`yeah`) or in writing (`such`).

Texts for children and for adults also include more complex sentence structures than occur in their own speech. There are other differences, but let's forgo the exhaustive list and settle for the main point: because of the ways they are perceived and produced, spoken and written language are not simple variants of each other.

Mashup!

These dissimilarities arise from properties of the auditory and visual modalities and the corresponding sensory and motor systems. What happens if properties that are associated with one modality are ported to the other? Could a hybrid combine the best features of both? Or would it be a hideous mutant? The modern world has done a few such experiments for us. One is texting. Texting is a mashup of written and spoken language, created in response to the challenge of using language under the novel conditions created by smartphones. The result is a form of written language with several characteristics of speech:

- It's fast(er) fading. I type a text message; it's read by the other party, hopefully the intended one. The text remains on the screen for a while but then drifts off. For the user, text messages are more transient than traditional texts but more persistent than speech. Of course text messages are far more permanent than they appear, having left the screen but not necessarily the Internet's colossal long-term memory.
- As with speech, control over the rate at which information becomes available rests with the producer. Typing a text message is more like formulating a spoken utterance; if the text is dictated using your phone's amazing speech-recognition software, then it is based on actual speaking. Text messages are less amenable to editing than ordinary texts. I said that speech was more like a first draft and a standard text more like the final version? Texting is a rough draft with limited vocabulary and syntax, spelling errors, neologisms, abbreviations, and, for the exuberant, emoticons, emojis, and other sprinkles. These characteristics largely result from the constraints imposed by smartphone technology, combined with software limitations such as imperfect autocorrection. We use the shortcuts that are available in this medium to approximate a real-time conversation even though we're tapping virtual keyboards.

Texting is a case in which merging some characteristics of text and speech is successful, creating a novel hybrid that fills a new, if limited, communicative

niche. Early fears that texting's idiosyncrasies would corrupt grammar and spelling have turned out to be unfounded, although they have created new genres of humor. Internet amusements aside, the evidence to date suggests that people—even tweens who text a few thousand times a month—do not find it difficult to maintain a distinction between texting and more formal types of writing.

Other hybrids are disasters because the properties of the two codes are mismatched. Some kinds of texts can be understood when read but not when heard. For many years academic philosophers made a point of giving talks that consisted of reading a characteristically abstruse paper aloud. The text was written knowing that readers would be able to proceed at their own pace, rereading, pausing, and consulting other sources as necessary. The spoken form eliminates these options, making the same text much harder to understand—often unintelligible, I'm told.

The education of many generations of hearing-impaired people has been affected by the misuse of hybrid forms of English. Like other spoken languages, English was shaped by our hearing and speaking capacities. Signed languages such as ASL are shaped by the characteristics of the visual-gestural modality. Signed English is a code that educators of the hearing-impaired devised so that English grammar could be used in both signing and reading, a logical goal. However, taking a grammar adapted to the properties of one modality (speech/hearing) and porting it to a different modality (vision/gesture) creates an epic mismatch. Signed English can be easily comprehended and produced by hearing people but is tedious and awkward for the hearing impaired, the intended beneficiaries.

Hear My Words

Reading and speech share a common linguistic core, but the modality differences mean that some kinds of information are easily communicated in one but little or not at all in the other. This means that there is more to language than afforded by either code itself. That the codes are complementary rather than wholly redundant contributes to the symbiosis between them and to my rejection of the view that reading is merely the handmaiden of spoken language.

Some types of information are represented in the written code but not in speech. Spaces indicate the boundaries between words in alphabetic writing systems; there are no "spaces" between words in fluent speech. Spoken languages have homophones, words that are pronounced the same but are otherwise unrelated. They are ambiguous when spoken but often spelled differently

(e.g., BLUE/BLEW, PEAR/PARE/PAIR). Typographical conventions such as using capitalization to indicate proper nouns do not have spoken equivalents in English.

That is small stuff compared to what's left out. Writing systems represent spoken languages, but they are not close representations of speech itself. The speech signal contains clues to everything from the speaker's sex, age, race, and education to whether they are happy or anxious, alert or distracted, telling the truth or dissembling. We do not rely on writing systems to represent this information, even though it is often essential to comprehending a text. Consider the following monologue:

> Are you talking to me?
> Are you talking to me?
> Are you talking to me?
> Are you talking to me?
> Are you talking to me?
> Are you talk? . . .
> Are you talking to me?
> Are you talking to me?
> Are you talking to me?
> Well, you must be, because I'm the only one here.

A person might experience feelings of dread merely reading the transcript of Robert De Niro's scabrous monologue from the 1976 movie *Taxi Driver*. But the quote is not from that movie. It's from *Wise Guys*, a 1986 gangster comedy in which Harvey Keitel does a send-up of his friend De Niro. The example illustrates the fact that text conveys limited information about how an utterance is spoken, data that can signal the difference between psychopathy and parody. And speaking of humor, here is John Cleese's favorite joke:

> A grasshopper hops into a bar and onto a stool. The bartender says, "We've got a drink named after you." The grasshopper says, "What, Norman?"

The joke only succeeds if the reader assigns the appropriate intonation—a property of speech—to the punchline "What, Norman?" (roughly the same as in "Who, me?"). Punctuation provides helpful clues but does not fully specify the humor-intensive intonation. In written form the joke relies on the reader's ability to mentally supply this information. Thus a joke can be funny even if it is read silently rather than told by John Cleese. Successfully "hearing" the relevant intonation is part of the pleasure.

These examples illustrate an intrinsic property of writing: properties of speech that are relevant to comprehension are not systematically represented. Reading is the use of a written code (orthography) to represent language. As in speech, the meaning of a text is greatly affected by phonology—the sound patterns of language. Writing systems vary in how much sound-based information they represent, but all of them omit a great deal. Properties such as pitch (Elijah Woods's is higher than Seth Rogan's), timing (speech rate, the length and placement of pauses), and loudness convey important information but are represented little or not at all. We're left with two questions. How do we manage to avoid regularly misinterpreting what we read? And wouldn't it have been smarter to include this information in the written code? Our keyboards are packed with characters that could be commandeered for the purpose.

The basic answer to both is that reading works because we are able to supply this missing information, most of the time, without its being an explicit, codified part of the writing system. Although it could be fully represented in writing, the costs would greatly outweigh the intermittent benefits. The system is a compromise that works quite well but not perfectly.

The principal mechanism at work is that many kinds of phonological information can be safely omitted because they are predictable from other things we know—about the topic, about the context in which a sentence occurs, about spoken language. Explicit marking would be superfluous. The ability to use what we know to go beyond the information given is a fundamental property of human perception and cognition. We fill in missing information all the time. We see shapes and letters where the parts are only implied, as in visual illusions and "incomplete" type fonts. We "hear" words when we read. Permit me to demonstrate. You just read the word PERMIT in that last sentence, and you unconsciously gave it the iambic (weak-strong) stress pattern: perMIT. But if you read I JUST GOT MY NEW PARKING PERMIT, you give the same spelling pattern the trochaic (strong-weak) stress: PERmit. We "hear" the appropriate stress pattern in our mind's ear. This solution is apparently good enough, and we're smart enough for it to work, but do we have to like it? Wouldn't life be simpler—and reading a little easier—if the writing system deigned to indicate syllabic stress using, say, accent marks?

Actually, no. As it has evolved over many centuries, written English has come to respect a very deep principle about what makes a code (such as a writing system or encryption cipher) efficient. Informally it can be stated as, Formulate your message in a way that makes it likely to be understood but avoid including more information than that requires. Explicit marking of syllabic stress, as in PerMIT ME TO DEMONSTRATE HOW TO USE THE PARKING PERmit, would be overkill because we can determine stress patterns well enough without it. Using

capitalization, accent marks, or other typographic elements to explicitly indicate syllabic stress would only gild this particular linguistic lily.

Adding notational elements also comes with costs: more symbols to learn, more to write, more to read, more paper and ink or bytes. These far exceed the costs of occasional errors. Worse, providing too much information makes reading more difficult. Modern Hebrew provides a beautiful example of the trade-offs between explicitness and efficiency. Like English and other alphabetic writing systems, the modern Hebrew alphabet includes symbols for consonants and vowels. Unlike English, Hebrew can be written either with or without the vowels. The pointed version, in which small diacritic marks indicate the vowels, is used for beginning readers. Skilled readers transition to the unpointed version in which the vowels are omitted. What happens, then, when adults are asked to read texts in which the vowels are put back in? They read more slowly. It's orthographic TMI.

For both English stress patterns and Hebrew vowels, the moral is that readers are better off with texts that omit some information because we can reliably fill it in based on our vast experience, which tells us what is likely to be correct. The operative phrase here is "vast experience," which is what it takes to sort out the complex details of systems that have statistical tendencies but not inviolable rules. For example, many noun-verb pairs are of the PERmit/perMIT type, with strong-weak stress for the noun and weak-strong stress for the verb (others are consort, detail, protest). The language allows many exceptions, including ones where both the noun and the verb have the strong-weak pattern (use the anchor to anchor the ship) and ones where both have the weak-strong pattern (release the press release). This characteristic is not limited to syllabic stress; it is a fundamental characteristic of language. The term for a system, like language, that exhibits rule-like regularities but also admits instances that deviate from these central tendencies in varying degrees is "quasiregular." Languages are quasiregular, but chess is not, because the pieces are not allowed to deviate intermittently from their designated movements.

We are able to speak and comprehend language with great skill despite its quasiregularity—indeed, because of it. Communication requires shared knowledge, and so languages must be systematic rather than arbitrary. However, the demands of comprehending and producing language require additional flexibility because speakers produce forms that deviate from standard patterns and listeners must be able to comprehend them. Many shortcuts that promote fluent speech eventually enter the language, such as "gonna," "hafta," and "tryna," which partially overlap with the source words. The product of these conflicting pressures is quasiregularity. These patterns can be mastered with extensive practice, which is easy to obtain if you've grown up speaking a language and

become a fluent reader. Mastering stress patterns is much harder for people learning English as a second language, who often exhibit "stress deafness."

In short, writing lets readers make use of the knowledge that allows spoken language to work so well, without getting overly specific. This solution does not work perfectly, but aiming for perfection would make it worse.

Just Cheat

And then there's cheating. The degree to which writing underrepresents speech creates room for misinterpretation. The risk can be reduced by taking advantage of ways the basic alphabetic code can be embellished, which range from punctuation (the Oxford comma, which shoots and leaves) to fonts and styles to ornaments such as emoticons and emojis. Users of English take these options for granted, but many writing systems admit far less of this stuff: there is little underlining in Chinese or italic in Hebrew, for example. In the writing systems that allow such embellishments, they are often used to fill phonological gaps, adding missing information on an as-needed basis. You may be constitutionally opposed to emojis and the effusive use of italics and exclamation points, but these devices are weapons in a revanchist struggle against phonological underspecification in the written code.

Clive James, a prolific writer of Australian origin and a witty presence on British television for many years, is possessed of an enviable ability to uncover many of the elegant sentences that are buried in the combinatorial explosion of possibilities afforded by the English language. A man who writes "Everyone has a right to a university degree in America, even if it's in Hamburger Technology": *that* is a humorist. A writer who has translated Dante's *Inferno* has no use for typographical embellishments. The closest James will come to an emoticon is a discussion of John le Carré's Smiley. But what if failing to use typography in a familiar way misrepresents the meaning of an utterance? Here is James reviewing a British reality TV program:

> Nobody should ever watch *Classic Car Rescue* for any reason, but if you did happen to see it you might find one of its unprepossessing male presenters pointing at a car and saying: "That is a car!" The information content of such a statement is zero. . . . The statement "That is a car!" can tell you nothing, unless the car is disguised as, say, a heap of cardboard boxes.

The program sounds dreary, but the utterance "That is a car!" does not lack content. James's print version conveys that the fellow was stating the obvious with mindless enthusiasm. It seems far more likely that he was expressing an

opinion about the object, which could be conveyed by simply using italics in the familiar way to indicate heavy stress on the initial word: "*That* is a car." The phrase exemplifies a familiar phrasal template used to identify an entity as a special example of its kind. Of Dr. Martin Luther King Jr.'s "I Have a Dream": *that* was a speech. Of a special dish: *that* is food. Of an automobile exhibiting qualities that distinguish an object of automotive envy from a Pontiac Aztek: *that* is a car. "Nobody should ever watch *Classic Car Rescue* for any reason" is hyperbole, but "the information content of 'that is a car' is zero" is fallacious because of James's disdain for using a trivial typographical convention to accurately represent the form and meaning of the man's utterance. *That* was disingenuous.

It may not please the National Council of Teachers of English (or concern Clive James in the slightest), but typographical variation is a fine way to represent heavy word stress, a property that can greatly change meaning:

Fred didn't take the test yesterday. (Somebody else did.)

Fred *didn't* take the test yesterday. (Although he could have.)

Fred didn't *take* the test yesterday. (He posted it on the Internet.)

Fred didn't take *the* test yesterday. (He took a different—lesser?—one.)

Fred didn't take the *test* yesterday. (He took something else.)

Fred didn't take the test *yesterday*. (He took it some other day.)

These differences in emphasis are a property of spoken language that is easy to represent in writing using italics, boldface, caps, or color.

These options are fine as far as they go, which unfortunately isn't very. Typography works pretty well for individual words but not for prosody—phonological patterns that extend over sequences of words. Prosody can make "Are you talking to me?" sound threatening and "What, Norman?" funny. Like other writing systems, written English represents very little prosodic information despite its impact on meaning. A few simple patterns can be represented fairly reliably, such as uptalk, in which a statement is spoken with question intonation—as in "And so, one time? I was at band camp? And we weren't supposed to have pillow fights? But we had a pillow fight! And it was so much fun!" This one is easy because the question marks successfully evoke question intonation even when the sentence is an assertion. The technique is not a precision instrument, however, and it helps if you've seen the movie.

Mischief can ensue. Figurative language is a potential weapon of self-destruction in public settings such as politics and popular culture. Verbal irony works by creating a contradiction between the intended meaning of an utterance and its literal meaning. The speaker's ironic intent is often conveyed by

prosody, which is not represented in the written code. The transcription of an ironic remark will better represent the unintended meaning, a formula for spectacular miscommunication. The utterance "The Beatles are more popular than Jesus" could be said in a manner that conveyed astonishment, regret, or glee, distinctions that are lost in a literal transcription. Journalists are ostensibly required to report what was said, not what they judge the speaker to have meant, but quoting only the exact words is not neutral or "objective." By omitting information about how the utterance was said, a quotation can convey a different meaning than the intended one.

It is a short step from The Beatles to Ronald Reagan. His Russian bombing remark ("My fellow Americans, I'm pleased to tell you today that I've signed legislation that will outlaw Russia forever. We begin bombing in five minutes") was audacious humor as spoken by a president of the United States, but the utterance can read quite differently, as it did to the Soviets ("The USSR condemns this unprecedented and hostile attack by the US president"). Nonliteral expressions are high-risk behavior for public figures because of their susceptibility to misinterpretation, especially in print, which is why an ironist like David Letterman did not grow up to be president.

Typographic conventions can reliably convey some kinds of useful information but are utterly ineffective with others. They are objects of prejudice because what they do convey can usually be expressed more clearly using language itself and because their genuine communicative utility is undermined by other factors: their slackish use in place of language, their decorative functions, their appropriation by middle school children. The net result is that they act as useful adjuncts to a writing system rather than being systematically incorporated. We want these gimmicks around because they are helpful and entertaining, but they are not sufficiently well mannered to gain full membership in the orthographic club.

What Is It Like to Be the Word "Bat"?

Writing does not specify features of speech that carry important information, which readers can usually fill in. Making a habit of this could do things to a person's brain. For literate individuals, what we know about the written code deeply penetrates our knowledge of spoken language, changing how we think about speech, how we understand it, and how it is represented in the brain.

Say the word BAT out loud. What is that sound made of? We think of "bat" as comprising three sounds, called phonemes, which I'll write as /b/, /a/, and /t/, spoken in that order. TAB consists of the same phonemes spoken in the opposite order. Everyone knows this. Right?

FIGURE 2.1. Waveform for the spoken word "bat."

Figure 2.1 shows the sound wave produced when I recorded the word "bat" on my laptop. Time runs from left to right, and the height of the waves indicates their intensities (roughly, loudness). Where are the three phonemes in this waveform? The boundaries between the letters in BAT are clear, but the boundaries between the sounds in "bat" are not. (The same is true if the spoken word is viewed as a spectrogram, a more detailed image that also represents sound frequency). It isn't possible to select parts of the waveform that correspond to each of the phonemes because the word is pronounced as a continuous articulatory gesture, not three discrete ones. Parts corresponding to the /b/, for example, also include parts of sounds that come later. The letters of a word are like beads on a string, but the sounds are more like a cascading waterfall.

Why do we have the overpowering sense that "bat" consists of three phonemes? Because we read and spell. English is written with an alphabet. We learn to treat the spoken word *as if* it consisted of three discrete sounds because it is written with three discrete letters. This abstraction allows units in the written code (graphemes, which are single letters or combinations, such as SH) to correspond to units in the spoken code (phonemes). A beginning reader can then learn systematic mappings between the two.

Although we think of speech as consisting of discrete phonemes, it is easy to demonstrate that it doesn't. Think of the common activity in which a Muppet (or person) models sounding out a simple word. Letters appear on the screen and the Muppet says the sounds associated with them, one at a time, "b" . . . "a" . . . "t" . . . , gradually decreasing the pauses between them. Sometimes the letters are displayed far apart on the screen and gradually brought closer together as a visual cue. The sounds do not meld into "bat" no matter how rapidly in succession they are spoken because it does not consist of three discrete segments. A discontinuity always occurs at the very end when the rapidly but

discretely enunciated phonemes are followed by the word pronounced as a whole. How to get from one to the other, the Muppet does not say. The activity is useful because the child learns about letters and their sounds. It encourages the fiction that words consist of discrete segments even as it demonstrates that they do not.

Learning to treat spoken language as if it were composed of phonemes is an important step in learning to read an alphabetic writing system. Spoken-language games such as rhyming take readers part of the way there. Hearing that "cat" and "hat" have the same ending "at" makes it easier to isolate the differing initial sounds. However, it takes exposure to the spellings of words—learning to read and write—to complete the full phonemic illusion.

If the sense that spoken words consist of a series of letter-like sounds arises from learning an alphabet, do people who cannot read perceive speech differently than people who can? Definitely. It is trivial for you, a literate speaker of English, to decide if two spoken words end with the same sound (e.g., "bat," "sit") or not ("bat," "sick"). Preschool children who are ready for reading can too. In a classic study, illiterate Portuguese adults (tested in their language, of course) found the task extremely difficult, and their accuracy was poor. A person who does not read treats a word like "bat" as a pattern without discrete parts, more like the speech wave pictured above. Learning the alphabetic code reveals components hidden within spoken words.

This fact is crucial. Using spoken language does not require knowledge of phonemes: the Portuguese adults could speak the language; they just couldn't read. Learning to read changes the representation of speech, promoting the emergence of an abstract unit, the phoneme. Representing spoken words this way makes it easier to read the alphabetic code, which in turn solidifies representing speech as phonemes. This feedback loop is a critical mechanism in the development of reading skill. Learning to read is more difficult when the development of phonemic representations is impaired or discouraged by educational practices.

There are easier ways to observe the impact of alphabetic knowledge on spoken language than studying Portuguese illiterates. We'll do an experiment. On every trial the subject (a reader like you or me) will hear a pair of words and decide if they rhyme. Some pairs will be rhymes, such as "bank"/"sank," and some will be nonrhymes, such as "beer"/"sank." The task is easy. The question is, What happens if the rhyming words are spelled similarly (e.g., GOAL/COAL) or dissimilarly (BOWL/COAL)? Keep in mind that the experiment only involves rhyming: the subjects hear the words, with no reading or writing involved. The results come out very clearly: it is harder to decide if two words rhyme if they are spelled differently. Subjects do not make many mistakes—they know

that "bowl" and "coal" rhyme—but it takes measurably longer to decide. Apparently the spellings of words affect how we hear them. The effect is so strong you can replicate it with an obliging friend and a stopwatch app.

Spelling's impact on the brain's representation of speech can also be seen using neuroimaging and brain-stimulation methods. For people who can read, there are no pure representations of the sounds of words in the brain because they've been contaminated by spelling. Language is a virus, as the musician-artist Laurie Anderson said, but orthography is the virus that infiltrates language, as observed by the noted British cognitive neuroscientist Uta Frith: "Learning an alphabetic code is like acquiring a virus [that] infects all speech processing, as now whole word sounds are automatically broken up into sound constituents. Language is never the same again."

Reading is powerful. It took most of our history as a species to evolve the capacity for speech, yet it is readily modified by a few years of exposure to print. That is a very good thing, because bringing print and speech into alignment makes reading feasible.

Reading and speech are different because of the modalities they involve, the ways they are acquired, the information they convey, and the conditions under which they are used. Different, yes, but closely bound and mutually dependent, each one changed by knowledge of the other, from behavior to brain.

CHAPTER 3

Writing: It's All Mesopotamian Cuneiform to Me

WHAT IS THE MOST IMPORTANT invention in history? In the era of the Internet and the listicle (and who invented that?), the question is easy to ask and has been many times. The answers are boringly predictable: wheel, printing press, lightbulb, computer, Internet, and so on. The bar code and laser scanner are huge for people in retailing; according to one poll, the invention that MIT undergraduates could least do without was the toothbrush (followed by car, computer, mobile phone, microwave oven). These lists are more interesting for what they consistently omit: writing.

The first thing that happened to reading is writing. For most of our history, humans have been able to speak but not read. Writing is a human creation, the first information technology, as much an invention as the telephone or computer. Had the US Patent and Trademark Office been there, writing could have been patented—as "a system for inscribing messages on semipermanent media" perhaps—whereupon a Sumerian patent troll could have sued the Phoenicians, Egyptians, and Chinese for infringement and demanded licensing fees. People probably don't think of writing as an invention because they do not clearly distinguish writing from speech. Writing's origins are also obscure: if there was an ancient Thomas Edison, that person neglected to use their astonishing invention to record their own name. And yet the creation of writing was one of the greatest achievements in human history. The development of modern civilization could not have occurred without the massive increase in the creation, retention, and transmission of information that

writing afforded. Without writing there would *be* no printing press, lightbulb, computer, or Internet.

Writing systems are fascinating but also exhausting, like the Egyptian rooms in the British Museum. The sheer number of writing systems and range of elements they have been fashioned from are a vivid testament to human creativity. Writing seems to come in about as many styles as beer. Orthography mavens are like birders, tracking exotic species identified by visual features, sound, and habitat. They have Pinterest boards and websites. Projects on cuneiform and Egyptian hieroglyphic are middle school standards, occasions for retelling stories about the ingenious clay tablets and inscrutable pictographs. In some cultures the young find it pleasing to have symbols from exotic writing systems inscribed on their bodies, often to comic effect though they may not know it.

Writing systems are immensely interesting in their own right (or, with a nod to John Lennon, in their own write), but they are not just for hobbyists or bad tattoos. The properties of writing systems are important because of their influence on how we read. Their apparent diversity raises a fundamental question: How many ways *are* there to read? Does reading work in essentially one way—because human brains and other relevant body parts are alike—or do the many writing systems indicate that reading is accomplished in multiple ways, a case of different strokes (of the pen) for different folks?

The surprising finding is that what is most salient about writing systems—their visual heterogeneity—camouflages what is most important about them: that they are essentially alike. The great insight was that spoken language could be represented in the medium we call writing. Several thousand years of experimentation with such systems ensued. The development of writing systems was shaped by human capacities, cultural factors, and some accidents of geography and history, eventually resulting in a high degree of convergence in how they work. They are similar because we are similar, as are the functions and uses of writing and the ways we read.

Interestingly, these same considerations militate against there ever being a single, universal writing system. The pairings of writing systems and languages are not random. Like writing systems, spoken languages are also "essentially alike," with some variation in how they are organized. Some types of writing systems work better with some types of spoken languages. Sufficient congruence between the properties of writing systems and the spoken languages they represent is critical to their functionality. The misalignments that periodically occurred during that long period of experimentation were fatal for some writing systems but also the source of breakthroughs in the advancement of this technology, the one we truly cannot live without.

The primal question about writing is how it originated. How did writing come into being, when, and how many times? Was it devised by a genius inventor? Or is writing, like the computer, a technology that has many sources but no originator? The similarities to that other primal question, about the origin of human life, are obvious. We have a deep-seated urge to understand how not only we but also our greatest artifacts, institutions, and ideas came into existence. Just as our existence is commonly attributed to a supernatural being, writing has been regarded with such awe that it too has been thought to have divine origins. The Babylonians believed that the god Oannes, part fish, part man, emerged from the sea to impart language, science, and writing; in Egypt, it was Thoth, also the god of wisdom, magic, language, and arithmetic.

Writing can also be seen as the result of either intelligent design or evolution. For some scholars, writing was such a profound advance, so unlike what came before and so fully realized from the start, that it could only have been designed by a supreme intellect, an ancient Newton or Einstein. Others interpret the evidence as showing that writing systems resulted from an evolutionary process that proceeded in fits and starts and punctuated equilibria on a very long time scale, shaped by adaptations and mutations that produced intermediate forms that eventually gave way to contemporary ones. The controversies are similar as well, as in the interpretation of gaps in the fossil record for evolution and in the archaeological record for writing. Writing systems are not biological organisms—they don't have genes to pass on to their offspring, for one thing—which is where the close analogies end. Writing systems are the product of cultural evolution, which is loosely related to the biological kind.

Unlike the origin of the human species or the invention of the wheel, tangible evidence about the origins and development of writing exists because it provided a record of itself. The clay tablets turned out to communicate far more than their creators intended and to be more durable too. We do not know whether the inscriptions on the Pharaohs' tombs accomplished their intended eschatological functions, but they succeeded in preserving information about the Pharaohs' lives, the lives of their people, and their culture, including the ingenious hieroglyphic writing. The drama of the discovery and deciphering of these artifacts is another source of writing's fascination, the stuff of symbology and *Indiana Jones* movies: ancient objects carrying faint messages from deep in the human past, the literal beginnings of recorded history. From the troves of writing samples that have been recovered, anthropological archaeologists have managed to establish a remarkable amount about early writing and the peoples who produced it. For example, much is known about the domestication of cattle in Sumeria in 3000 BC because scholars were able to identify signs for cattle and dairy products and ones for quantities signifying the number

of cattle in a herd and amounts of milk and cheese. It was a scholarly tour de force to determine that the first person whose name is known to us from a written record is Enmebaragesi of Kish, a Sumerian king who lived around 2600 BC. The impact of this research runs much deeper than the accumulation of such facts, however. Viewing a tablet that recorded some sort of transaction involving sheep and oil in an ancient village whose location can be pinpointed on Google Maps, we experience a connection to those people that is intimate yet farther removed in time than even the most elaborate genealogy or oral history.

What the archaeological record does not offer is definitive answers to the primal question. The available evidence is frustratingly incomplete and about as jumbled as a pathological hoarder's stash. It reveals a great deal about what was created: various writing systems, their properties, how they changed over time. But specific claims about where, when, and by whom they were invented are (literally) educated guesses. Consider cuneiform, often celebrated as the first writing system, said to have originated in Sumeria in the village of Uruk, where the crucial tablets, about 4,000 of them, were found. Every element of this account is uncertain. Thousands of inscribed tablets have been found across a wide swath of the modern Middle East; many predate cuneiform and may have influenced its development, but few have been deciphered, and most are in storage, unexamined. The cuneiform tablets may have been brought to Uruk rather than originating there. The language may not be Sumerian but that of a neighboring people who created the tablets. Some scholars argue that the cuneiform system was invented out of whole clay, so to speak, by a single Sumerian genius, others that it developed gradually over a wide geographical area. Enmebaragesi is the name, but artifacts discovered in the 1980s suggest that this king might have been a queen or a fictional composite.

Although the archaeological record is indeterminate, the primal question is too interesting to forgo, leaving scholars to vigorously debate various scenarios in accordance with Benford's law of controversy: "Passion is inversely proportional to the amount of real information available." For example, most observers have concluded that writing systems were independently invented three times: in Sumeria, China, and Mesoamerica. Working from the same data as everyone else, anthropologist Jared Diamond proposed that the Chinese picked up the concept from their Egyptian trading partners, who adopted it from the Sumerians. Diamond's version adds a novel twist to an old story, it keeps the discussion alive, it could be true—and if more of those unanalyzed clay tablets are deciphered, some day it may be possible to tell.

The incompleteness of the archaeological record is unfortunate because the experiment called the invention of writing is over and cannot be replicated

with a better data-collection plan. The combination of deep-seated questions and a rich but incomplete historical record is toxic stuff for scholars, whose essays exhibit the "false and damaging certainty" that writer Janet Malcolm finds in reporting about contemporary events that, like the origins of writing, have emotional resonance but inherent ambiguity, such as hideous crimes committed under murky circumstances. We demand clear narratives in such cases, even if the truth is hopelessly underdetermined by the evidence. The stories told about where and when writing was invented are what I would call honest confabulations: serious attempts to bring coherence to an immense yet inadequate body of evidence, the docudramas of the academic world, "based on a true story."

What the History of Writing Has to Say About Reading

The primal question concerns the activities of individuals and peoples at particular times and locations. Whatever the interest of that question and prospects for answering it, we can also look at writing as a product of the human species. The properties of writing systems, their similarities and differences, and the ways they changed over time are data that can help to address the hardest question about writing systems, which is not who developed them but how *anyone* could have. How did humans manage to progress from drawing *pictures* of horses to writing *about* them? The invention of writing forces us to confront a general puzzle about human creativity: Where do new ideas come from? Writing seems to be a prime example of a cultural artifact that is sui generis—lacks precedent. Writing did not exist for all of human history until it did. Something happened. But where did the very idea of writing come from if it did not already exist? This puzzle also arises in connection with the several great innovative leaps that occurred during writing's long, slow elaboration and refinement. For example, alphabets represent phonemes, but the phonemic abstraction depends on exposure to an alphabet. How could phonemes have been discovered without alphabets that represent phonemes?

These issues are more complex than whether Enmebaragesi was king or queen, but they may be more answerable because the evidence is not limited to the archaeological record. What we now know about human psychology, biology, and culture is relevant to interpreting what happened to people very much like us several thousand years ago. We can also use what has been learned about how we read to identify crucial properties of writing and track how they developed. The answers to these questions do not turn on information at the level of names, dates, and places and so the archaeological gaps are of less concern. There is insight to be gained about the mechanisms by which writing

systems developed from looking at the capacities of the people, the properties of the languages they spoke, and the conditions under which they lived and by asking, What are the problems for which writing systems are the solution?

The standard story about the origin of writing goes like this. Humans have been creating representational images for more than 30,000 years, the approximate date of the oldest known cave paintings. The paintings are depictions of things their creators saw, mainly animals, objects, and body parts. Early writing is said to have built on this capacity, using simplified depictions of objects, called pictographs. The use of pictographs limited communication to what can be rendered in this manner. Illustrations similar to Figure 3.1, found on countless websites, show how signs in Sumerian cuneiform, the best documented of the earliest writing systems, changed over the 3,000 years it was in use. The illustration conveys that the writing system mostly began with pictographs, which gradually became more abstract, greatly expanding what could be represented and communicated. On this view, modern forms of writing came about when ancient peoples developed notations that overcame the limits of depiction, analogous to the transition from realism to abstraction in the history of Western art.

This story has great intuitive appeal, but it is the science project version. It is true that overcoming the limits of depiction was an enormous advance that took eons to achieve. The surprising twist, however, is that writing did not originate with pictures that gradually became more abstract. The monumental development was using graphical elements, some of which were pictographic, in a radically new way, to represent language. It was this innovation that took so long to achieve, and how it could have occurred is the great question. The gradual loss of the pictographic element in some signs was a minor development in comparison; rather than leading to writing, it occurred *because* writing was successful and came into wide use.

Of all the events that occurred in the long development of writing, four are crucial to understanding the essential properties of writing systems and their relation to how we read:

Moving from depiction to symbol: Writing systems emerged when pictographs and other graphical elements were used to symbolize language rather than signify things. A picture of a bird, for example, could be used to represent the spoken word for bird. It is then a symbol for a sound pattern. Using a picture for a purpose other than signifying the pictured entity was not merely counterintuitive, it went unintuited for eons. Once the trick of using graphical elements as symbols was discovered, they could be pictographic or abstract, and they could be used to represent

FIGURE 3.1. This illustration (and variants found on numerous websites) conveys that writing began as pictures but evolved into more abstract symbols. The illustrations are based on one in Kramer (1963), popularized by DeFrancis (1989). It isn't quite right because almost all of the earliest cuneiform signs were abstract. From Kramer (1963), 304–305. Copyright © 2015 by University of Chicago Press. Reprinted by permission..

many types of information (e.g., words, initial sounds of words, concepts, categories of objects, grammatical elements).

Representing entire languages: The proto-writing systems only represented some elements of a language, mainly words for objects and quantities. It took another couple of thousand years to develop workable writing systems that fully represented language. The major advance was determining how a relatively small set of symbols could represent a

much larger set of words. The general solution, which every successful writing system employs, is using combinations of symbols that represent clues about sound and meaning.

Discovering phonology: Writing systems require treating spoken words as consisting of parts, which can then be represented by a limited set of graphical elements. We take it as obvious that speech consists of units such as words, syllables, and phonemes, but these units are phonological abstractions that had to be discovered. Writing and the phonological way of thinking coevolved over a long period.

Establishing congruence: The properties of writing systems need to align with properties of the spoken languages they represent. Writing systems only converged on this crucial feature over a long period of trial and error.

The questions then are why each of these advances was so important, how they could have been achieved, and why they took so remarkably long. Reading is indeed an "unnatural act" compared to speech, as Philip Gough, a distinguished reading researcher, put it, but the events that were crucial to writing's development were supremely unnatural, distributed over many years, regions, cultures, languages, and individuals.

What Ought to Be in Pictures?

Having developed the necessary capacities and tools, *Homo sapiens* began to produce representational drawings and has done so ever since, a testament to their power, interest, and apparently inexhaustible variety. Prehistoric cave paintings have been discovered at numerous sites around the world, suggesting that creating such images is a very basic human impulse. Why the pictures were created and the functions they served we don't know, of course, but they tell us that the people who produced them possessed the confluence of perceptual, cognitive, and motoric skills that are required to merely depict an object. If the painted image was produced to be communicative rather than decorative or as a way to kill Pleistocene time, other sophisticated capacities would have been involved, such as formulating a communicative intention and understanding how the image would be experienced by someone else. Humans have a "theory of mind" that allows us to make such attributions about the mental states of others.

The cave paintings are remarkably advanced. They are not children's drawings or line drawings. To the modern eye they read as artful representations

LOOK	one morpheme
LOOKED	two morphemes (the-ED morpheme indicates the past tense of the verb)
LOOKOUT, OUTLOOK	two morphemes (when each is a word, the combination is a compound)
DIET	one morpheme (but two syllables)
DIES	two morphemes (but one syllable)

The association between the sound pattern "cat" and its meaning is arbitrary (as is the spelling-meaning association); however, what is learned about "cat" is relevant to other words in which it occurs: cats, catty, catnip, bobcat, and so forth. Morphemes such as cat make similar phonological and semantic contributions to many words. Because the spelling is alphabetic, CAT represents both phonology and semantics. Thus, alphabetic writing systems also conform to the phonology + semantics principle: they represent the sounds of words as sequences of phonemes, which also correspond to meaning-bearing morphemes.

Finding Phonemo

Everyone seems to have figured out that writing systems must represent sound and meaning, but it took a very long time to develop fully functional systems. The main obstacles were on the phonology side. Spoken words had to be treated as consisting of component parts, which could then be represented by a much smaller number of graphical symbols. The would-be architects of writing systems had to develop something that we now consider an ordinary, teachable aspect of learning to read: phonological awareness. Children demonstrate their understanding that words consist of parts when they can tell if two spoken words rhyme or begin or end with the same sound, or count the number of syllables in a word and identify component phonemes. The question is, How did our ancestors start treating spoken words this way? And how did they converge on the units that were right for their language in accordance with the Goldilocks principle: not too big, not too small, just right. The two developments were closely linked via the successes and failures that occurred during the long evolution of writing.

Although the exact events are unknown, the fact that different types of spoken languages require different types of writing was clearly crucial. The languages spoken in the greater Mesopotamian area where the landmark