The Impact of Language Experience on Language and Reading
A Statistical Learning Approach

Mark S. Seidenberg and Maryellen C. MacDonald

This article reviews the important role of statistical learning for language and reading development. Although statistical learning—the unconscious encoding of patterns in language input—has become widely known as a force in infants' early interpretation of speech, the role of this kind of learning for language and reading comprehension in children has received less attention. In fact, the implicit learning of co-occurrences of words, sentence structures, and other components of language forms a critical part of children's language comprehension and fluent reading. Beyond introducing basic information about language statistics, the article offers a discussion of how variability in the amount and nature of language experience can affect language development and literacy, including variation owing to the amount of language input in the child's linguistic environment and the variable nature of input for children who are exposed to multiple languages or multiple dialects. **Key words:** *bilingualism*, *dialect*, *individual differences*, *language variation*, *statistical learning*, *vocabulary*

THIS ARTICLE provides an overview of a theoretical framework for examining the impact of experiential variability on children's spoken language comprehension and learning to read. Researchers now know that children's early language experience is more variable than had previously been recognized, and that these differences have substantial effects on children's progress in learning to read, findings that have also become widely known to the general public. Other factors aside,

children whose spoken language is more advanced learn to read more easily (McCardle, Scarborough, & Catts, 2001); children with lesser skills are at greater risk for reading difficulties; and children who do not attain sufficient reading skill have difficulty acquiring the many other types of knowledge that depend on print. These include math, given the heavy linguistic emphasis in modern curricula (e.g., explaining your work, story problems, math problems involving real-world situations described in words). They also include language itself, because reading is the primary means by which linguistic knowledge expands through the life span.

In this article, we examine variability in language experience from the perspective of statistical learning, an important type of learning that people engage in from birth. Statistical learning is the largely unconscious process of learning the patterns of one's environment— the probabilities that events will occur, or occur together, and in which sequences (Lany & Saffran, 2013; Seidenberg, 2017). The range of "events" that can be learned is broad, such

as the sequence of green, yellow, and red traffic lights; the pairing of the word "tiger" and a picture of a tiger in a child's book; and the fact that IN is a more common letter sequence in English than IE, even though the letter E is overall more common than N in written English. Because statistical learning occurs over the entirety of an individual's experience, the amount and variety of that experience greatly affects what is learned. Statistical learning provides a broad theoretical framework for evaluating how differences in spoken language and reading experience affect children's development.

The structure of the article is as follows: We first provide an overview of the statistical learning approach to language, including speech and reading. Statistical patterns exist within and between multiple levels of linguistic structure. This information plays a crucial role in learning and using spoken language. It is implicated in reading in three respects. First, variability in the child's exposure to spoken language, associated with socioeconomic status and other variables, has downstream effects on linking print and speech. Second, learning about the written code—orthography—is itself a statistical learning problem. Children learn about the frequencies and distributions of letters, which patterns form words, and how those patterns relate to spoken language. This topic has been discussed extensively elsewhere (Castles & Nation, 2006; Grainger, 2008; Seidenberg, 2011). Third, the statistical properties of texts and speech differ. Texts employ vocabulary and sentence structures that rarely occur in speech. These characteristics have been termed "academic" or "school" language, and they are often seen as a major hurdle for many children (Bunch, Abram, Lotan, & Valdés, 2001).

With this overview in hand, we then consider the impact of individual differences in experience: the amount and variety of the child's spoken language experience and variability in the type or types of linguistic code(s) to which the child is exposed. Statistical learning provides a perspective on how these var-

ied experiences affect language development and learning to read. The approach is also relevant to the important question of how experience can be structured to promote success in learning to read and educational achievement in general, especially for children whose language experience is bilingual or bidialectal.

STATISTICAL LEARNING

Until recently, research on how languages are acquired and used was conducted within the framework established by Noam Chomsky. This framework emphasized the structure of language (grammatical competence) rather than facts about how language is used (performance). Researchers sought to identify the universal and language-specific properties of language, focusing on grammaticality (the well-formedness of utterances) rather than their use in the varied circumstances of people's lives. The apparent fact that children converge on knowledge of a language despite substantial differences in experience was taken as evidence that essential aspects of grammar are innate rather than learned. With this understanding of grammar in hand, one could then ask how children acquire knowledge of a particular language, how this knowledge is used in comprehending and producing language, and how language is represented in the brain.

Although there had previously been objections to this view (e.g., Bates & MacWhinney, 1987), in the 1990s some researchers began to advance an alternative. The competence approach excluded information about performance; however, the explanations for how language is acquired and used may have been thrown out with the performance bathwater (Seidenberg, 1997). With the availability of tools for collecting and analyzing large language corpora and the creation of archives such as the Child Language Data Exchange System (CHILDES, MacWhinney, 2000) and the Penn Treebank (Marcus, Marcinkiewicz, & Santorini, 1993), it became apparent that language as it is used exhibits myriad statistical regularities: patterns in the use of elements

within and between levels of linguistic structure, including phonology, morphology, vocabulary, and grammar, as well as in the relations between language and the contexts in which utterances occur. These statistical regularities turn out to be crucial: Languages can be learned and used because they exhibit these patterns and humans are capable of learning and using them.

Children engage in statistical learning from birth, acquiring knowledge about the patterns of common events and their co-occurrences. For example, when toddlers knock a spoon off their high chair tray or drop a piece of banana over the side, they are learning about gravity—that unsupported objects will fall—and they also may be learning that dropping objects gets the attention of adults in the room. As we will show, much of language is learned in a similar manner. Statistical learning is implicit, meaning that it occurs without conscious intention to learn and without awareness that learning is taking place. It occurs in the background as we engage in activities such as reading, speaking, and acting. It is sensitive to the amount and variety of experience, and thus to individual differences in language background.

Learning to read differs from learning a first language (or two) in many respects, but it is similar in an important one: it too depends on statistical learning. The beginning reader's task is to learn how orthography—the written code—relates to the spoken language they already know. The spellings of words exhibit complex statistical patterning, as do the mappings between spelling to sound. Because the writing system is alphabetic, spelling and sound are more highly correlated than spelling and meaning (see Seidenberg, 2017, for detailed discussion). The same implicit learning mechanisms are employed in acquiring these types of knowledge as in acquiring a first language.

Statistical learning is not the only way that people acquire knowledge; we also learn via explicit experience—instruction and feedback, exploration and discovery—with intention and awareness. These types of learning are closely intertwined, each affecting the other. Reading depends more on explicit instruction than does spoken language, but the difference is again a matter of degree. Learning the inflectional systems in languages such as Finnish and Hungarian, which are much more complex than in English, requires extended explicit instruction in school, including the use of textbooks and other written materials. In contrast, learning to read those languages requires less instruction because their writing systems have few inconsistencies of the sort that are so prominent in English.

Statistical properties of language

Spoken and written language exhibit innumerable statistical patterns—inhomogeneities in the frequencies and co-occurrence of elements. Table 1 lists a number of statistical patterns with references to relevant research. The table is not an exhaustive list, and it does not do justice to the patterns that exist between levels. It also ignores how statistical patterns are altered by differences in language experience. However, the examples convey that statistical patterns exist at all levels of linguistic structure, ranging from the spoken and written forms of words, to combinations of words, to entire sentences. The table provides a snapshot of what is now a vast body of research conveying that from infancy onward, people are implicitly learning and using statistical knowledge to understand spoken and written language and their world around them (e.g., Conway & Christiansen, 2005; Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). The table stops at the level of comprehending sentences, but it also could have included statistical properties of entire texts. Such properties have been used to identify the authors of texts where the authorship is disputed (Jockers & Witten, 2010), and they are the basis for automatic procedures for grading documents such as freshman English essays (Shermis & Burstein, 2013).

Table 1. Several types of language statistics in English that are crucial for oral and written language use

Language Level(s)	Example	Relevance
Phoneme positions	/h/ does not occur at ends of words, /ŋ/ does not occur at word beginnings	The statistics of phoneme locations in words support word recognition and identifying word boundaries in the speech stream (Vitevitch, Luce, Pisoni, & Auer, 1999).
Phoneme and letter transition probabilities	The phoneme sequence /nv/ is rare compared with /nt/; the same holds for the corresponding letter sequences NV and NT.	Sequences with low-transition probability are likely candidates for word and syllable boundaries in speech and syllable boundaries in reading (e.g., CANVAS, CAN VOICE; (Seidenberg, 1987).
Syllable stress assignment in reading	Pronunciation of RECORD as REcord vs. reCORD varies with part of speech	Readers use sentence context to identify stress and part of speech even in silent reading. Patterns are only probabilistic: ANCHOR, e.g., has the same stress pattern in both noun and verb forms (Seidenberg, 2017).
Word meaning	GROUND (n) floor vs. GROUND (n) background vs. GROUND (v) past tense of grind vs. GROUND (v) conduct electricity GROUND (adj.) pulverized	Most common content words in English are ambiguous; meanings often belong to different parts of speech. Comprehenders must use semantic and syntactic context to identify the intended interpretation. Context is statistical—the floor sense of GROUND co-occurs with *fell on the*, and the adjective sense co-occurs with foods such as *meat* and *coffee* (MacDonald, 1993; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982).
Conceptual combination	BIRD STATUE = statue depicting a bird; MARBLE STATUE = statue made of marble (not depicting it), MARBLE CATALOG = catalog describing marble, not made of it.	Compound nouns, such as BIRD STATUE, MOUNTAIN MAGAZINE, ANCHOR LOCK, have many possible meanings but also probabilistic regularities that help comprehenders settle on the most likely interpretation (Murphy, 1990).
Collocation	Collocations in American English include: BACK IN THE DAY, NO WORRIES, GOOD TO GO, BRUSH YOUR TEETH, ON THE OTHER HAND	Collocations are high-frequency word sequences. These include but are not limited to idioms. Whereas frequency effects of individual words are well known to language researchers, there is growing recognition that comprehenders also track collocation frequencies and use them in speech and reading (Arnon & Snider, 2010).
Pronoun ambiguity	Maria told Sue that she . . . SHE can refer to either Maria or Sue	Pronouns are extremely common in speech and texts but create ambiguities. Statistical regularities, such as pronouns more often referring to grammatical subjects or other prominent nouns, guide children's interpretation (Arnold, Brown-Schmidt, & Trueswell, 2007).

Although the cases listed in Table 1 involve different types of information, they are similar in this respect: at each level, a finite number of elements (say, letters) are combined to form a much larger number of units (e.g., words). The combination of elements is highly constrained: only a subset of the possible combinations is allowed. Millions more words can be formed by combining the letters of the alphabet than the 30,000 or so in a person's vocabulary, for example. The same holds for combining words to form multiword sequences. For example, the words WE, NEED, and TO form one very common trigram WE NEED TO, whereas other combinations (WE TO NEED, etc.) are rare (though they may well occur, e.g., WHO ARE WE TO NEED HIM?). In all of these cases, the frequency distributions are highly skewed: a small proportion of the patterns (letters, phonemes, words, sequences of words) are used with high frequency, and then there is a long tail of combinations that are used much less frequently. Statistical patterns arise from constraints on combining elements that arise from a variety of sources, both endogenous (e.g., human information processing and learning capacities) and exogenous (e.g., systematic characteristics of the world in which language is used).

Table 1 lists many statistical regularities separately, but a critical feature of language statistics is that they are correlated rather than independent, so that regularities at one level are informative at many others. For example, knowing that a word contains the morpheme BELL provides information about other words in which it occurs: BELLS, BELLED, COWBELL, BELLBOTTOMS. Similarly, BLUE occurs in BLUEBIRD, BLUEGILL, and BLUEBERRY. These patterns are helpful in learning and using new words—BLUEBELL, say. The word BELL is correlated with other linguistic and extra-linguistic information as well: bells are objects that are usually rung; bell ringing usually will involve an overt or implicit agent to make it happen (THE CUSTOMER RANG THE BELL; THE BELL RANG); bells are used for some purposes (as alarms; to make music) but not for many others (cracking nuts, recording a lecture). All these

patterns are probabilistic, meaning that they afford a range of possibilities that differ in their probability of occurring, and violations of the dominant patterns abound (dumbbells are neither dumb nor bells; a bell might be used to crack a nut in a crunch; a bellboy is a person who carries your luggage, not your bells, and probably is not a boy, either). Language, like the world, is quasiregular (Seidenberg & McClelland, 1989): mostly predictable, though not entirely, and thus probabilistic patterns are helpful.

This partial analysis of a single word, BELL, illustrates an important point about what it means to know a word. Researchers tend to think of words as countable entries in a mental dictionary. The contents of that mental dictionary are assessed using tests such as hearing a word like "bell" and picking out the corresponding picture, as on the Peabody Picture Vocabulary Test (PPVT, Dunn & Dunn, 2007). The entry for a word specifies its form (the pronunciation of BELL, and later, its spelling), grammatical category (noun), and meaning (instrument that produces sounds by being struck). However, our knowledge of a word is not like a dictionary entry. A word is more like a hook on which to hang many types of information: multiple meanings, senses, and grammatical functions of the word (e.g., verb senses of BELL); encyclopedic knowledge (facts about where bells are made or found); the kinds of words that are likely or necessary to occur with a word in different contexts; and other things we know as well. Children who know the word BELL can differ greatly in how much information they associate with it, a property that Perfetti (2007) termed "lexical quality." Moreover, much of our grammatical knowledge is associated with individual words (MacDonald, Pearlmutter, & Seidenberg, 1994), specifying the kinds of sentence structures they occur in and their grammatical and thematic functions. Words are not just discrete building blocks out of which sentence meanings are formed, nor are they processed as such. In reading or hearing each word, we are generating, confirming, and revising

expectations about other words. Sequences of words are comprehended by converging on the interpretation that best satisfies the expectations words carry about each other. The problem is greatly simplified by knowing the statistics: that is, which patterns are likely to occur and, just as important, which of the many possible patterns can be ignored because they occur very rarely or not at all.

LEARNING AND USING LANGUAGE STATISTICS

People are good at using language statistics and yet largely unaware of the mechanisms involved, because learning and using language statistics are unconscious and automatic. To get a sense of what is going on under the hood, it is helpful to walk through a simple example. Consider the following sentence, which is from *Lost in the Tunnel of Time* by Sharon Draper, a chapter book for 8–12-year-old readers: "The early bell seemed to hear him, for the signal to go into the building sounded just as he spoke" (Draper, 2011a, p. 4).

Even the first word, THE, carries several probabilistic cues to what is coming: first, that a noun will occur within the next few words; second, that the very next word is likely to be a noun or part of a noun phrase, such as an adjective; third, that the noun is likely to be one describing something alive such as GIRL or DOG, because that is the most common pattern for the subject of a sentence.

The second word, EARLY, confirms the existence of a noun phrase but also reduces the expectation that the upcoming noun will describe a living thing. EARLY is an adjective that usually modifies inanimate things (such as MORNINGS, BELLS, and RESULTS) more often than it modifies living things (such as BIRDS). EARLY also creates a new ambiguity because it can refer to a recent time (this morning's early bell) or the distant past (early Stone Age tools).

The next word, BELL, reduces the time ambiguity because the text is more likely to be about a recent event, such as the ringing of a school bell, rather than a bell of historical interest. The probabilities would change, of course, if the book were "Bronze Age Chinese Bells" rather than a novel.

At this point, the reader has found a likely, interpretable noun phrase, THE EARLY BELL, although it takes an additional word, the verb SEEMED, to confirm that the noun phrase ends there (it could have continued with another noun, as in THE EARLY BELL SCHEDULE).

This process of generating expectations and picking the most probable option continues: the verb SEEMED suggests an upcoming adjective (as in THE EARLY BELL SEEMED LOUD), but the sequence SEEMED TO rapidly shifts expectations to an upcoming verb.

The verb expectation is then confirmed but with an unexpected word, HEAR. The author confounds her readers because their knowledge of the world tells them that a bell, an inanimate object, can be heard but cannot hear. Readers or listeners may experience momentary uncertainty about how to fit the words together, an effect the author likely would have anticipated.

The final word of the first clause, HIM, indicates the object of the hearing (English uses HE for a male subject and HIM for an object). This unambiguous object marking supports the surprising interpretation that the bell is doing the hearing. To make sense of this, the reader/listener must then infer that the event has been described figuratively rather than literally. The phrase has now been assigned a plausible interpretation. Yet with all that progress, HIM creates a new ambiguity because the sentence occurs in scene in which two male characters are conversing. Additional work is required to determine the likely referent of HIM, and so it goes, as comprehension of the sentence and text proceeds.

This analysis illustrates some of the probabilistic constraints that contribute to comprehending just a seven-word phrase. Comprehenders cannot predict exactly what will occur following THE EARLY BELL but they can assign probabilities to a limited range of likely possibilities based on prior experience. The example also illustrates the important

fact that statistical information constrains interpretation in the forward direction, so information from one word yields predictions about aspects of upcoming words, but also in the backward direction, where that same word may also be refining the interpretation of the words that were previously read. Comprehension proceeds in this way because language is deeply and pervasively ambiguous. Ambiguity is not limited to words with two obviously different meanings, such as PITCHER meaning a container or an athlete. Instead, *every* word contains some amount of ambiguity, and the perceiver must continuously use prior knowledge and the preceding context to predict and evaluate what is coming and also revise the interpretation of preceding material as necessary. In this way, an understanding of a sentence emerges from the unconscious process of using the statistics of language and the world to settle into the most likely sentence meaning.

The key to the use of such statistics is how they are combined. The process can be illustrated by a simple game. What word am I thinking of?

Something yellow: Not very informative. Many yellow things. Low probability of guessing correctly.

Something kind of round: same thing. Lots of round things in the world.

A kind of fruit: again, not a very informative clue taken in isolation. Lots of fruits; low probability of guessing correctly.

All of these clues are weak taken independently, but the combination of yellow + round + fruit yields a very likely answer: *lemon*. The conjunction of cues is highly constraining because the facts are not independent: fruits are objects that have colors and sizes.

Language is similar but on a much larger scale. Each level of linguistic structure exhibits statistical regularities, but the levels are not independent: each constrains what can occur at other levels. The ability to derive a high-probability event (e.g., the meaning of a phrase) from a combination of much lower probability ones occurs in language and in many other domains: it is a fundamental characteristic of human behavior. We make intelligible cognitive and linguistic mountains out of miserable statistical molehills, all the time, and effortlessly.

The evidence that language users combine probabilistic constrains in this way derives from several sources. There is abundant research on children and adults' use of statistical information in comprehending speech. This research has documented the use of simple statistics in very young children and the developmental progression that occurs as they learn more about their language and their world, which is marked by changes in the statistics that are used and increases in proficiency in using them (Lew-Williams & Fernald, 2007; Saffran et al., 1996). A few studies have tied children's abilities to use spoken language statistics to language experience: children who hear more language, measured as number of words of child-directed speech, complexity of that speech, and amount of conversational turn taking, perceive speech more rapidly and accurately and combine statistical cues faster as well (Fernald, Marchman, & Weisleder, 2013).

The bulk of the evidence for complex cue integration during sentence comprehension comes from studies of reading. We know that skilled adult readers use all the types of statistics illustrated in the EARLY BELL example and in Table 1. Psycholinguistic research has yielded a huge body of work on how adult readers integrate probabilities like these in comprehending sentences, and how variations in predictability affect reading patterns, as measured using eye-tracking and related methods (for review, see MacDonald & Hsiao, in press). There are few comparable studies of children's reading, however, leaving a gap in what is known about how these skills develop and change over time.

Evidence of adults' exquisite use of statistics leads to the question of how we got here: how does a language learner or user know which of the myriad statistical properties of a language are relevant? This could be called the "richness of the stimulus" problem:

languages afford many generalizations that children never make. In an earlier era, the explanation was that innate grammatical knowledge restricts the range of hypotheses about language, obviating the problem (Cook, 1991). The alternative is that the problem is solved by the statistical learning procedure itself. The use of statistical information in learning, perception, and cognition can be formalized in several related ways; see McClelland et al. (2010) and Griffiths, Chater, Kemp, Perfors, and Tenenbaum (2010) for comparisons between the Bayesian and connectionist approaches that are widely used in studies of language and cognition. The choice between approaches often depends on the type of question being asked and how much is known about the problem domain. In general, Bayesian models start with strong theories about which types of information (e.g., language statistics) are relevant to a behavior such as categorizing objects or predicting grammatical categories of words in sentences. They then specify the way to combine existing knowledge with new data (e.g., about a sentence being read) to reach optimal inferences. In language, however, we do not know in advance which statistics are relevant or their numerical values. Connectionist (aka "deep learning") networks are particularly useful tools for examining how this knowledge develops incrementally with experience. To examine this, a computer simulation model is given a task such as recognizing a letter string as a specific word. Given the task, the architecture of the model, the corpus of training examples, and the learning algorithm, the model acts as a discovery procedure, determining which properties allow the task to be performed quickly and accurately (Seidenberg & MacDonald, 1999). The statistics that are relevant do not have to be specified in advance; the model solves the problem itself, such that over time, the model's behavior is determined by statistical properties that allow the task to be performed. The same principles apply in networks that process sequences of words that form meaningful sentences (Elman, 1990).

Something similar happens in people. People are statistical learners. We cannot prevent ourselves from responding to patterns in the environment, registering repetitions and novelties, similarities and differences, the way things vary and covary, and then how the things that covary actually covary. Learning a language, followed by learning to read is a Big Data problem for humans. For language learners, the data are the millions of utterances to which they are exposed and their relations to situations and events in the world. Statistical learning is the process of distilling regularities from this mass of data—patterns, largely shared by speakers of a language, that enable communication to occur (Seidenberg, 2017).

STATISTICAL LEARNING AND LANGUAGE VARIATION

Statistical patterns become effective guides for comprehension only when humans are exposed to large numbers of examples, initially spoken and later written. The evidence that these patterns are critical to speaking and reading suggests that differences in the amount and variety of language experience should greatly affect performance. Of course, we know this to be true: characteristics of children's early spoken language experience vary in ways that affect what they learn, with downstream effects on educational progress, especially in reading. In the following sections, we examine several important sources of individual differences in language and reading from a statistical learning perspective.

Vocabulary learning is statistical learning

The Hart and Risley (1995) study is the best known of what are now many studies to have examined variability in the quantity and variety of language in the learner's environment (earlier such studies include Snow et al., 1976). There is an extensive literature linking language experience to socioeconomic status (Hoff, 2003) but also important studies demonstrating variability within lower socioeconomic status families (Pan, Rowe,

Singer, & Snow, 2005; Weisleder & Fernald, 2013). Although these studies involved characterizing multiple properties of the language environment, vocabulary gets the most attention, perhaps because it is easy to quantify: as the number of different words a child produces, or as performance on a simple word–picture matching test such as the PPVT. Vocabulary measured in these ways is an indicator of some characteristics of children's linguistic knowledge, and the substantial individual differences that are apparent at the start of kindergarten are related to progress in learning to read (Cunningham & Stanovich, 1991; Walker, Greenwood, Hart, & Carta, 1994). It is also important to note that these assessments index children's knowledge of the standard dialect of the language used in school but not their knowledge of another dialect or language and thus may underestimate many children's linguistic knowledge.

The emphasis on vocabulary might seem narrow, given that using language communicatively involves many other types of information. As we have observed, however, vocabulary assumes even greater importance when it is recognized that words are statistically linked to other words and to other levels of linguistic representation and thus carry information about the sentences in which they occur. Children with smaller vocabularies—assuming their vocabularies have been adequately assessed—do not simply know fewer words; they also know less about language and the world. Children who experience higher rates of child-directed speech not only have greater vocabulary knowledge but also comprehend speech more rapidly (Weisleder & Fernald, 2013).

Vocabulary knowledge that is below age or grade-level expectations is troubling because it is hard to ameliorate. There simply is not sufficient time to explicitly teach hundreds of words. However, statistical learning provides a potential mechanism for building vocabulary more efficiently. Few of the words we know were learned through explicit instruction. We infer the meanings of words from the linguistic and extra-linguistic contexts in which they occur. Words with similar meanings tend to occur in similar contexts (Firth, 1957). Much can be inferred about the meaning of the word LYNX, for example, because it appears in the same statistical contexts as the related words LION and TIGER (even better if there is also a picture). Lila Gleitman and colleagues famously demonstrated that the meanings of verbs can be "bootstrapped" to a great extent from the syntactic contexts in which they occur (Fisher, Hall, Rakowitz, & Gleitman, 1994). Landauer and Dumais (1997) described how knowledge of language statistics prepares people to learn new words before they have been experienced. Even a limited amount of vocabulary instruction is helpful because learning a new word also facilitates the learning of other, related words (Beck, Perfetti, & McKeown, 1982).

Impoverished language experience is a source of deficits in vocabulary and other areas, but enriching that experience can potentially accelerate learning without heavy dependence on instruction. We do not yet know how to do this effectively, though researchers are exploring methods that include promoting reading to children, providing training to parents, and monitoring rates of child-directed speech, as well as prompts for child-directed conversation in public spaces such as grocery stores (Hirsch-Pasek et al., 2015; Ridge, Weisberg, Ilgaz, Hirsh-Pasek, & Golinkoff, 2015; Suskind et al., 2016). New approaches that have different or broader impacts are likely to emerge from these efforts.

Statistics of speech and reading

Language statistics undergo a substantial change as the child transitions from prereader to independent reader because the statistics of written language differ markedly from those of spoken language. Words such as LIKE and YEAH occur more often in speech; THUS and WHICH occur more often in print. ACTUALLY is more common in speech; OBVIOUSLY in print. Lower frequency words such as DIVERGE and CONCLUDE are unlikely to be learned except through reading.

Differences between the modalities extend beyond the distributions of words. Speech, especially to children, is often about the here and now, where objects are visible and recent events are part of the shared conversational context. A sentence such as DON'T DO THAT AGAIN! is short and simple, and it gains its full meaning via its relationship to the environment, namely, whatever it was that the child was doing. Texts are decontextualized, and the "world" is established with words and sentences, creating substantial differences compared with spoken language. One example is relative clauses, which give extra context about the people and objects mentioned in a text. They are commonly used in setting the scene in stories for children, such as the classic openings of fairy tales: ONCE UPON A TIME, THERE WAS AN OLD WOMAN [WHO LIVED IN A HUT [THAT WAS COVERED WITH FLOWERS]]. Here the relative clauses, shown in brackets, give extra information about the main character, the old woman, and also about her hut. Relative clauses exist in child-directed speech but they are quite rare (Montag & MacDonald, 2015), largely because the environmental context often carries some of the meaning—we do not have to say the relative clause in "Pick up the toys [that you got out]" if we can simply point while saying "Pick up the toys." Montag and MacDonald (2015) found that compared with child-directed speech, texts for young children contain far more relative clauses—a four-fold increase over speech for some types of relative clauses, and a 200-fold increase for other types. Some examples of the sorts of relative clauses that preschool children and elementary school readers encounter are shown in Table 2.

The examples in Table 2 do not have particularly difficult vocabulary, but they are nonetheless likely to be challenging for young readers. The reason lies in language statistics: whereas relative clauses are rare in child-directed speech, they are much more common in texts for children. Sentence structures, including relative clauses, that occur mainly in texts are a type of "academic language." Children must master these unfamiliar forms as they are learning to read.

Although "academic language" is typically treated as something to be learned in school,

Table 2. Relative clauses in picture books and texts for young independent readers

Source and Reading Level	Examples, With Relative Clauses [in Brackets]
Asim (2006). *Whose Knees Are These?* (Picture book for reading to 1–3-year-olds).	Knees [like these] don't grow on trees.
McQuinn (2006). *Lola at the Library* (picture book for reading to 2–5-year-olds).	She put all the books [she borrowed last week] in her backpack.
Thomson (2010). *Keena Ford and the Field Trip Mix-up*, an early reader chapter book for 6–8-year-olds.	But it turns out a thesaurus is just a book of words [that mean the same thing as other words] . . . . I know about student council because this fifth grader [named Lamont] [who walks me home] was on student council last year.
Draper (2011b). *Shadows of Caesar's Creek* (a chapter book for 7–12-year-old readers)	Using the remains of an old fence [that the boys had found in Ziggy's backyard], they had built the clubhouse themselves the previous summer. They had cut holes [that looked a lot like windows] in two side walls . . . .

Note. Reading level retrieved from Amazon.com Web site.

a surprising finding from recent research is that children are often introduced to it much earlier via shared reading with caregivers. The fact that shared reading supports reading development is widely recognized (Payne, Whitehurst, & Angell, 1994) and promoted (see, e.g., "Reading is Fundamental" and other such public service announcements). The activity introduces children to literacy, familiarizing them with the format and content of books and demonstrating that print represents language. The activity is highly valued as a way to generate interest in reading and motivation to learn how to read.

It is less widely recognized that books for reading to and with young children provide exposure to language that rarely occurs in child-directed speech (Montag, Jones, & Smith, 2015). Such texts can enrich children's vocabulary because they contain a wider range of words than in speech, and the words occur in more variable contexts. Moreover, as Table 2 illustrates, picture books for adults to read to children can have an abundance of complex relative clauses and other constructions that are thought of as "academic language" precisely because they rarely occur in everyday child-directed speech. The same is true for many other sentence types that are not themselves syntactically complex but are nonetheless much more common in text than speech. In *Peter's Chair* by Ezra Jack Keats (1998), the protagonist, a boy named Peter, says, "We'll take my blue chair, my toy crocodile, and the picture of me when I was a baby," a construction a young child would rarely produce or hear. While definitive data are not yet available, these observations suggest that, in addition to its other functions, shared reading of picture books is important because it is a mechanism for familiarizing children with complex sentence structures they would not otherwise hear. As with child-directed speech, adults' shared book reading with children is highly variable, both in lower SES families (Payne et al., 1994) and in higher SES families with high levels of parent education (Montag & Smith, 2017). An obvious path for future research is to investigate the shifts in vocabulary and sentence structure statistics that come from typical picture books intended for shared book reading and to examine the effects of children's exposure to these books on their early reading.

Variability in type of language experience

Most studies of language and reading in the United States have focused on mainly White, middle-class individuals who are monolingual speakers of the mainstream ("standard") dialect (i.e., people very much like the researchers who usually conduct the studies). Language experiences frequently differ greatly from this case, however. There is a large literature on bilingualism, referring to children exposed to more than one language, dating back many decades, and a much smaller though growing literature on children who speak a minority dialect. In the United Sates, this dialect is most often African American English (AAE). Minority and "standard" dialects and their educational implications also have been studied in many other countries and languages (Latomaa & Nuolijärvi, 2002; Levin, Saiegh-Haddad, Hende, & Ziv, 2008). Dialects spoken by mainly low-income minority groups have been studied in Australia and Canada (Siegel, 2010). Here we consider how the statistical learning framework, which was also developed with the monolingual/mainstream dialect situation in mind, applies to more complex language learning environments.

It is difficult to generalize about "bilingualism" because it covers such an enormous range of circumstances. Consider the following examples:

a. Typological distinctiveness of the languages. English overlaps more with Spanish than with Mandarin Chinese. Learning English and either of these languages is more difficult than learning two Romance languages such as Spanish, Italian, and French, which overlap much more.

b. Experiential variables such as timing of exposure to the two languages; amount of exposure; conditions under which the

languages are used including who speaks the language(s) in the home; the school language environment and curriculum; incentives and disincentives to learn a second language; the socioeconomic status of the speakers (e.g., becoming a Spanish–English bilingual is different for middle-income English speakers learning Spanish vs. lower income Spanish speakers learning English); and many others.

The combinations of these (and other) factors create a huge, varied space of bilingual learning conditions. Disagreements about best practices in bilingual education (Hakuta, 1999; Kim, Hutchison, & Winsler, 2015) may reflect the fact that no procedure works equally well under all these conditions.

What can be said is that the learning environment is more complex for children who are exposed to two languages. The child (or adult) is learning two systems for expressing the same things; a monolingual learns one. Acquiring a language depends on the amount and variety of linguistic experience. For bilinguals, that experience is split, in varying proportions, between the languages. Bilingual children's knowledge of each language typically lags behind monolingual children's knowledge of their single language; this finding follows naturally from the experience-dependent character of learning. Over the longer term, these initial costs may be superseded by the benefits of being bilingual (Bialystok, Craik, & Luk, 2012).

Among other things (such as learning about the conditions under which different languages are used), the bilingual individual is learning two sets of statistics. Very little is known about this process. Consider, again, vocabulary. There is a long history of research on whether bilinguals represent the words in the two languages in separate "coordinate" lexicons or in an integrated "compound" lexicon. For example, CAMPANA is a Spanish word for BELL. A Spanish–English bilingual might develop an integrated lexicon in which BELL and CAMPANA are linked to their shared, musical instrument meaning, or they might develop separate lexicons for the two languages. Words such as BELL and CAMPANA are sometimes called translational equivalents: words in different languages with the same meaning. Although the meanings of the words are similar, they differ in most respects. For example, CAMPANA also refers to a hood for covering food (e.g., to keep flies away) but BELL does not. BELL can refer to both church bell and school bell; Spanish uses different words (church: CAMPANA; school: TIMBRE). All of the statistics governing the structures associated with each word and its co-occurrences with other words are radically different. However, the words are not completely unrelated. The properties of bells (the musical instruments), their purposes, the conditions under which they are used, and the fact that they are inanimate and, therefore, heard rather than hear are much the same. Acquiring these two sets of statistics is far more complex than memorizing translational equivalents or developing a compound versus coordinate lexicon. Encoding the similarities and maintaining the differences are hard learning problems and not well understood.

Although there is little research on statistical language learning in bilinguals (for review, see Weiss, Poepsel, & Gerfen, 2015), one gains a sense of the complexity of the task by comparing monolingual speakers of the languages. For English-speaking children, encountering an article such as A or THE provides evidence that a noun is upcoming in speech or text, and children as young as 2 years of age can use their knowledge of the statistics of article usage to predict upcoming nouns, even though they do not yet produce articles in their own speech (Zangl & Fernald, 2007). Article usage statistics are different for Spanish-speaking children because the language uses grammatical gender to distinguish masculine and feminine nouns. Both adjectives and articles are also marked for gender in Spanish, so that the equivalent of English THE in Spanish is EL for masculine nouns and LA for feminine nouns. On the one hand, this system is more complex than using THE for every noun, and there is the additional burden of learning the gender for

every noun; on the other hand, encountering a gender-specific article can narrow the range of possible nouns in a sentence, especially in combination with other contextual cues. Spanish-monolingual toddlers can use the statistics of gendered article–noun pairings to anticipate a masculine noun when hearing EL but a feminine noun when hearing LA (Lew-Williams & Fernald, 2007).

By contrast, adults who speak Spanish as a second language, typically having first been exposed to it in middle school, do not show evidence of taking advantage of the gender marking in this way (Lew-Williams & Fernald, 2010). A bilingual child is learning both sets of conditions, of course. Basic, though fascinating, questions remain to be addressed, such as whether bilingual children exhibit similar effects in a given language as are seen in studies of monolinguals. It seems clear that people have difficulty learning properties of a second language that are unattested or strikingly different than in their first language (e.g., Mandarin speakers' difficulties with articles such as THE; English speakers' difficulties with tone in learning Mandarin). Less is known about whether degree of exposure to properties of one language, such as the predictive utility of gendered articles in Spanish, affects speech or reading in a language that lacks that property.

Similarly, Spanish-English and Mandarin-English bilingual children who are learning to read relative clauses in English may face additional burdens compared with their monolingual English peers, because the properties of relative clauses in Spanish and Mandarin differ from those in English (and from each other). Spanish has relative clause options that do not exist in English, including alternative "impersonal" forms with different word orders, as well as the common use of relative clause markers that are similar to WHO versus WHOM, which can help disambiguate subject and object roles in relative clauses (Gennari, Mirković, & MacDonald, 2012), whereas WHOM is largely absent from English-speaking children's spoken language input. Mandarin is radically different: Whereas English and Spanish place the relative clause after the noun it modifies, as in CATS [THAT SLEEP ON THE BENCH], Mandarin places the relative clause before the noun it modifies, something like [SLEEP ON THE BENCH THAT] CATS. Again, although there is some evidence about how the language-specific conditions are learned by native speakers, little is known about learning them in two such languages.

Dialect variation presents many of the same issues. It is a sociolinguistic truism that languages and dialects exist on a continuum without a discrete boundary between them. We should then expect issues related to bilingual experience to be relevant to experience with two dialects. Just as languages vary in degree of overlap, so do the dialects of a language. Globally, English has numerous major dialects (Szmrecsanyi & Kortmann, 2009). Different dialects function as the "standard" in different regions, with other dialects diverging from the standard in varying degrees. As in the bilingual case, the bidialectal experience is highly variable, involving many of the same factors such as age and amount of exposure, the contexts in which the dialects are used, socioeconomic status of the speakers, and others. Here too the range of circumstances discourages broad generalizations.

Although bilingual and bidialectal environments differ in many respects, they share the fact that the child has more to learn than a monolingual/monodialectal child, and that the child is learning alternatives to things he or she can already say. Learning two dialects clearly represents another complex task. In the United States, the "nonstandard" dialect that has been studied most is AAE. The linguistic properties of the dialect (and its regional variants) have been described with numerous examples of their use, and there are many studies of children's AAE usage (e.g., in relation to their progress in reading; see Washington, Terry, & Seidenberg, 2013, for a review). Characterizations of the statistical properties of a language such as English usually look at general patterns seen in large corpora, ignoring dialectal variation. There is, as yet, almost no research applying statistical learning concepts to either AAE itself or to

the bidialectal learning environment, such as when AAE is mixed with the more "standard" dialect, often called Mainstream American English (MAE). Nonetheless, descriptions of linguistic features of AAE make it clear that they entail language statistics that differ from the mainstream dialect.

Consider, for example, a sentence that begins "Why did the boy . . . .." An MAE speaker learns a probability distribution for possible continuations, which includes likely ones such as SMILE, GO, ASK, and so on. Although this distribution is likely to be similar for AAE speakers, it will also differ because it includes AAE-specific constructions such as "Why did the boy didn't stop?" (Washington, 2001). How much these probability distributions differ between dialects and how much they vary across speakers of the dialects are not known because the research has not been conducted. From a statistical learning perspective, using both AAE and MAE requires two sets of overlapping but nonidentical statistics. Predictions about the word that follows "Why did the boy . . . ." depend on dialect knowledge but also sociolinguistic factors such as who is speaking and in what context (e.g., home vs. school). The learning task may be more difficult in some respects than in the case of two languages because the dialects, which are variants of the same language, differ but nonetheless overlap a great deal. Roughly speaking, the minimal overlap between the grammars of, say, English and Spanish means that there is relatively little facilitation or interference in learning one compared with the other (even less for English and Mandarin). The partial overlap between dialects of English facilitates learning of transparently similar forms, but it is also a source of interference when something that applies in one dialect is disallowed or disfavored in the other. In short, the dual-dialect language environment introduces additional ambiguities at many levels of linguistic structure, the resolution of which depends on knowledge of the relevant dialect-specific statistics.

Many children successfully manage the demands of learning and using two languages or dialects but outcomes vary greatly. The statistical learning framework could potentially provide important evidence about the conditions that promote or interfere with managing dialect (and language) differences. We have conducted one demonstration (Brown et al., 2015). The study focused on one of the simple differences between AAE and MAE, the statistics regarding alternative pronunciations of many words. Pronunciations can differ because, as one example, AAE permits optional reduction of consonant clusters. Thus, TEST can be pronounced "tes," omitting the final /t/ in the MAE pronunciation; COLD can be pronounced "cole" which is homophonic with COAL. The conditions governing the deletion of phonemes are complex and speakers differ in the extent to which they employ this option (Washington, 2001).

The existence of alternative pronunciations complicates learning in two respects. First, the AAE speaker has to learn alternative pronunciations of the same word, including an increased number of homophones such as "cole." Whether this presents a significant challenge compared with accommodating pronunciations that differ in pitch, rate, or other speech qualities is not known. Differences in the number of phonemes in two pronunciations definitely become prominent when the child learns to read. Beginning readers learn how spellings represent spoken words. Acquiring these spelling–sound correspondences is a classic statistical learning problem (Seidenberg & McClelland, 1989), already complex because of the properties of written English and even more complex for AAE speakers because spellings such as COLD map onto two pronunciations.

Brown et al. (2015) developed a simple computational model to compare the single versus dual dialect conditions. The main finding was that models that had learned the AAE pronunciations of words for speech had more difficulty learning the MAE pronunciations for reading. Learning proceeded more slowly than in the single-dialect case because there was more to learn. The model yielded two other findings: first, that performance of

the dual-dialect model eventually caught up to the single-dialect model with sufficient experience; second, that learning occurred more rapidly when the model included reliable contextual cues indicating whether to use MAE or AAE.

Such effects are not enormously surprising but they are not widely recognized either. One of the major differences between bilingual and bidialectal situations is in how behavior is interpreted. African American English and MAE overlap more than two languages, but the learning demands are similar in many respects. These similarities are not reflected in educational research, practices, or policies. The concept of bilingualism is clear to most people from examples such as individuals who can fluently speak both English and another language (such as Spanish). There is less familiarity with the bidialectal situation. People's lay concepts of dialect frequently fail to distinguish the linguistic properties of dialects (e.g., that they are comparable in complexity and expressive power) from the sociolinguistic status of dialects (e.g., that one may be institutionalized as the high prestige "standard," compared with lower status variants). The educational needs of bilingual children have been the focus of extensive research, resulting in programs that utilize a variety of strategies to promote acquisition of the school language, but the corresponding dialect issues remain politicized, understudied, and underrecognized in the United States.

When children who speak a different language in the home lag behind monolingual peers in speech and reading, the behavior can be attributed to the fact that they are English learners. African American English speakers may lag for similar reasons but that is rarely taken into account. Many speakers of AAE, educators, and observers still view AAE as a deficient version of English, despite decades of research showing this to be false. In contrast, a child who speaks a language such as Spanish in the home is not said to have learned "bad English." Whereas the linguistic distinction between language and dialect is a matter of degree,

educational policies and practices treat them dichotomously.

Both bilingual and bidialectal learning are areas where the statistical learning approach could help identify how to structure children's experience to promote successful learning. It could be determined, for example, how to take advantage of the child's existing knowledge of language to promote learning a second language or dialect. It could also be used to avoid practices that make learning the second code more difficult. For example, a high level of skill with one code may result in "entrenchment" that makes it difficult to accommodate a second code, which Seidenberg and Zevin (2006) termed "the paradox of success." Their simulation models suggested that this negative by-product of language proficiency can be avoided by interleaving experience with a second language relatively early, before the first language becomes highly overlearned.

CONCLUSIONS

The statistical learning approach to language and reading was developed in the context of research on monolingual speakers of "standard" language but is ripe for extension to variable learning environments. A theoretical approach that views language learning and use as the result of a large number of interacting statistical constraints can take into account a broader range of language experiences, as well as the impact of extra-linguistic factors (such as differences in background knowledge). Several approaches to statistical learning and probabilistic decision making exist and are being applied to a wide range of cognitive and linguistic phenomena (Clark, 2013; McClelland et al., 2010; Perfors, Tenenbaum, Griffiths, & Xu, 2011). These approaches have not yet had much influence on the study of variability in language experience and its impact on learning to read, but the field is wide-open and the potential payoffs for both theory and practice are huge.

Aside from the need for more researchers with the relevant skills to examine these

issues, the greatest need is for additional facts about the variability of language experience and use. The statistical learning approach resulted from advances in understanding human learning and in the development of computational and quantitative methods for analyzing large language corpora. Studies such as Hart and Risley (1995) look antiquated by modern standards: what was a heroic effort for the era now looks like a relatively small data set from an idiosyncratic set of families. Bringing statistical learning together with language variation requires gathering much larger samples of language behavior from a much wider range of individuals.

REFERENCES

Arnold, J. E., Brown-Schmidt, S., & Trueswell, J. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes*, *22*(4), 527–565. doi:10.1080/01690960600845950

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82. doi:10.1016/j.jml.2009.09.005

Asim, J. (2006). *Whose knees are these?* New York, NY: LB Kids.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157–193). Hillsdale, NJ: Earlbaum.

Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, *74*(4), 506–521. doi:10.1037/0022-0663.74.4.506

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240–250. doi:10.1016/j.tics.2012.03.001

Brown, M. C., Sibley, D. E., Washington, J. A., Rogers, T. T., Edwards, J. R., MacDonald, M. C., et al. (2015). Impact of dialect use on a basic component of learning to read. *Frontiers in Psychology*, *6*, 196. doi:10.3389/fpsyg.2015.00196

Bunch, G. C., Abram, P. L., Lotan, R. A., & Valdés, G. (2001). Beyond sheltered instruction: Rethinking conditions for academic language development. *TESOL Journal*, *10*(2–3), 28–33. doi:10.1002/j.1949-3533.2001.tb00031.x

Castles, A., & Nation, K. (2006). How does orthographic learning happen. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing*. Hove, Sussex, UK: Psychology Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences*, *36*(3), 181–204. doi:10.1017/S0140525X12000477

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology Learning, Memory, and Cognition*, *31*(1), 24–39. doi:10.1037/0278-7393.31.1.24

Cook, V. J. (1991). The poverty-of-the-stimulus argument and multicompetence. *Interlanguage Studies Bulletin (Utrecht)*, *7*(2), 103–117. doi:10.1177/026765839100700203

Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge, and spelling. *Journal of Educational Psychology*, *83*(2), 264–274. doi:10.1037/0022-0663.83.2.264

Draper, S. M. (2011a). *Lost in the tunnel of time* (Reissue edition). New York, NY: Aladdin.

Draper, S. M. (2011b). *Shadows of Caesar's creek* (Reissue edition). New York, NY: Aladdin.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Minneapolis, MN: Pearson Assessments.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, *16*(2), 234–248. doi:10.1111/desc.12019

Firth, J. R. (1957). A synopsis of linguistic theory, 1930–1955. In Philological Society (Ed.), *Studies in linguistic analysis*. Oxford: Blackwell.

Fisher, C., Hall, D. G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua*, *92*(Suppl. C), 333–375. doi:10.1016/0024-3841(94)90346-8

Gennari, S. P., Mirković, J., & MacDonald, M. C. (2012). Animacy and competition in relative clause production: A cross-linguistic investigation. *Cognitive Psychology*, *65*(2), 141–176. doi:10.1016/j.cogpsych.2012.03.002

Grainger, J. (2008). Cracking the orthographic code: An introduction. *Language and Cognitive Processes*, *23*(1), 1–35. doi:10.1080/01690960701578013

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of

cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357–364.

Hakuta, K. (1999). The debate on bilingual education [Editorial]. *Journal of Developmental*, *20*(1), 36–37.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes Publishing.

Hirsch-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., et al. (2015). The contribution of early communication quality to low-income children's language success. *Psychological Science*, *26*(7), 1071–1083. doi:10.1177/0956797615581493

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1368–1378.

Jockers, M. L., & Witten, D. M. (2010). A comparative study of machine learning methods for authorship attribution. *Literary and Linguistic Computing*, *25*(2), 215–223. doi:10.1093/llc/fqq001

Keats, E. J. (1998). *Peter's chair* (Reprint edition). New York, NY: Puffin Books.

Kim, Y. K., Hutchison, L. A., & Winsler, A. (2015). Bilingual education in the United States: An historical overview and examination of two-way immersion. *Educational Review*, *67*(2), 236–252. doi:10.1080/00131911.2013.865593

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. doi:10.1037/0033-295X.104.2.211

Lany, J., & Saffran, J. R. (2013). Statistical learning mechanisms in infancy. In J. Rubenstein & P. Rakic (Eds.), *Comprehensive developmental neuroscience: Neural circuit development and function in the brain* (Vol. *3*, pp. 231–248). Cambridge, MA: Academic Press.

Latomaa, S., & Nuolijärvi, P. (2002). The language situation in Finland. *Current Issues in Language Planning*, *3*(2), 95–202. doi:10.1080/1466420020 8668040

Levin, I., Saiegh-Haddad, E., Hende, N., & Ziv, M. (2008). Early literacy in Arabic: An intervention study among Israeli Palestinian kindergartners. *Applied Psycholinguistics*, *29*(3), 413–436. doi:10.1017/S0142716408080193

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193–198. doi:10.1111/j.1467-9280.2007.01871.x

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464. doi:10.1016/j.jml.2010.07.003

MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language*, *32*(5), 692–715. doi:10.1006/jmla.1993.1035

MacDonald, M. C., & Hsiao, Y. (in press). Sentence comprehension. In *Oxford Handbook of Psycholinguistics*. Oxford, UK: Oxford University Press.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703. doi:10.1037/0033-295X.101.4.676

MacWhinney, B. (2000). *The childes project* (3rd ed.). Mahwah, NJ: Psychology Press.

Marcus, M. P., Marcinkiewicz, M. A., & Santorini, B. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, *19*(2), 313–330.

McCardle, P., Scarborough, H. S., & Catts, H. W. (2001). Predicting, explaining, and preventing children's reading difficulties. *Learning Disabilities Research & Practice*, *16*(4), 230–239.

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., et al. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, *14*(8), 348–356. doi:10.1016/j.tics.2010.06.002

McQuinn, A. (2006). *Lola at the library* (1st ed.). Watertown, MA: Charlesbridge.

Montag, J. L., Jones, M. N., & Smith, L. B. (2015). The words children hear: Picture books and the statistics for language learning. *Psychological Science*, *26*(9), 1489–1496. doi:10.1177/0956797615594361

Montag, J. L., & MacDonald, M. C. (2015). Text exposure predicts spoken production of complex sentences in 8- and 12-year-old children and adults. *Journal of Experimental Psychology. General*, *144*(2), 447–468. doi:10.1037/xge0000054

Montag, J. L., & Smith, L. B. (2017). *Picture book reading in the lives of 18-30 month old children: A diary study*. Paper presented at the 39th Annual Cognitive Science Society Meeting, London, England.

Murphy, G. L. (1990). Noun phrase interpretation and conceptual combination. *Journal of Memory and Language*, *29*(3), 259–288. doi:10.1016/0749-596X(90)90001-G

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development*, *76*(4), 763–782.

Payne, A. C., Whitehurst, G. J., & Angell, A. L. (1994). The role of home literacy environment in the development of language ability in preschool children from low-income families. *Early Childhood Research Quarterly*, *9*(3), 427–440. doi:10.1016/0885-2006(94)90018-3

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension, *Scientific Studies of Reading*, *11*(4), 357–383. https://doi.org/10.1080/10888430701530730

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, *120*(3), 302–321. doi:10.1016/j.cognition.2010.11.015

Ridge, K. E., Weisberg, D. S., Ilgaz, H., Hirsh-Pasek, K. A., & Golinkoff, R. M. (2015). Supermarket speak: Increasing talk among low-socioeconomic status families. *Mind, Brain, and Education*, *9*(3), 127–135. doi:10.1111/mbe.12081

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. Retrieved from https://books-google-com.ezproxy.library.wisc.edu/books?hl=en&lr=&id=n1cd70T4WxIC&oi=fnd&pg=PA90&dq=elissa+newport&ots=1XlC2tMNSu&sig=BF7rz_5i82_rUrTYBDcPYqQaOGk

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

Seidenberg, M. S. (2017). *Language at the speed of sight: How we read, why so many can't, and what can be done about it*. New York, NY: Basic Books.

Seidenberg, M. S. (1987). Sublexical structures in visual word recognition: Access units or orthographic redundancy? In M. Coltheart (Ed.), *Attention & performance XII: Reading*. London, England: Earlbaum.

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. (Cover story). *Science*, *275*(5306), 1599.

Seidenberg, M. S. (2011). Reading in different writing systems: One architecture, multiple solutions. In P. McArdle, B. Miller, J. R. Lee, & O. Tzeng (Eds.), *Dyslexia across languages: Orthography and the brain–gene–behavior link* (pp. 146–168). Baltimore, MD: Paul H. Brookes Publishing.

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, *23*(4), 569–588. https://doi.org/10.1016/S0364-0213(99)00016-6

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, *14*(4), 489–537.

Seidenberg, M. S., & Zevin, J. D. (2006). Connectionist models in developmental cognitive neuroscience: Critical periods and the paradox of success. In Y. Munakata & M. Johnson (Eds.), *Attention & performance XXI: Processes of change in brain and cognitive development* (pp. 585–612). Oxford, England: Oxford University Press.

Shermis, M. D., & Burstein, J. (Eds.). (2013). *Handbook of automated essay evaluation: Current applications and new directions*. New York, NY: Routledge/Taylor & Francis Group.

Siegel, J. (2010). *Second dialect acquisition*. Cambridge, New York: Cambridge University Press.

Snow, C. E., Arlman-Rupp, A., Hassing, Y., Jobse, J., Joosten, J., & Vorster, J. (1976). Mothers' speech in three social classes. *Journal of Psycholinguistic Research*, *5*, 1–20.

Suskind, D. L., Leffel, K. R., Graf, E., Hernandez, M. W., Gunderson, E. A., Sapolich, S. G., et al. (2016). A parent-directed language intervention for children of low socioeconomic status: a randomized controlled pilot study. *Journal of Child Language*, *43*(2), 366–406. doi:10.1017/S0305000915000033

Szmrecsanyi, B., & Kortmann, B. (2009). The morphosyntax of varieties of English worldwide: A quantitative perspective. *Lingua*, *119*(11), 1643–1663.

Thomson, M. (2010). *Keena Ford and the field trip mix-up*. New York, NY: Puffin Books.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language*, *68*(1–2), 306–311. doi:10.1006/brln.1999.2116

Walker, D., Greenwood, C., Hart, B., & Carta, J. (1994). Prediction of school outcomes based on early language production and socioeconomic factors. *Child Development*, *65*(2), 606–621. doi:10.2307/1131404

Washington, J. A. (2001). Early literacy skills in African-American children: Research considerations. *Learning Disabilities Research & Practice*, *16*(4), 213–221. doi:10.1111/0938-8982.00021

Washington, J. A., Terry, N. P., & Seidenberg, M. S. (2013). Language variation and literacy learning: The case of African American English. In C. A. Stone, R. Silliman, B. J. Ehren, & K. Apel (Eds.), *Handbook of language and literacy: Development and disorders* (2nd ed., pp. 204–221). New York, NY: Guilford Press.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152. https://doi.org/10.1177/0956797613488145

Weiss, D. J., Poepsel, T. J., & Gerfen, C. (2015). Tracking multiple inputs: The challenge of bilingual statistical learning. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 167–190). Amsterdam, Netherlands: John Benjamins.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*(2), 250–271. doi:10.1016/j.cogpsych.2008.08.002

Zangl, R., & Fernald, A. (2007). Increasing flexibility in children's online processing of grammatical and nonce determiners in fluent speech. *Language Learning and Development: The Official Journal of the Society for Language Development*, *3*(3), 199–231. doi:10.1080/15475440701360564