Contents

Chapter 1

Introduction

It is easy to take our knowledge of language for granted. We learn language before we carry our first backpack to school, and we use it almost every waking hour of every day. Although we may not have studied quantum theory, or read Homer or James Joyce, we are each expert at using our own native language. The challenge that all learners face becomes more apparent when we try to learn a second language in school or as adults.

There are many utterances that are perfectly understandable, but which nonetheless tend to be avoided by native speakers of English. If asked, speakers will agree that there is something mildly "off" about them, even though they may have difficulty articulating exactly why they don't sound quite right. For example, we might confess that someone is driving us *crazy* (or *bananas* or *insane*), but we know that it would sound odd to complain that someone is driving us *angry*. We know that *tall* bushes are *high* bushes, but a *high* teenager is not necessarily *tall.* We can be creative in how language is used, but our creativity is constrained in ways that can be hard to articulate. For example, someone can *tell me something* or *tell something to me*, but they can only *explain this to me*; that is, it sounds somewhat unconventional to native speakers of English to say, *explain me this*. That is what this book aims to explain: when, why, and how native speakers are sometimes creative with language and yet at other times much more conservative.

Speakers avoid saying certain things, of course, simply because they want to avoid overtly negative reactions. The following are examples of such ill-advised utterances:

Sorry Mom, I didn't mean to get caught.
I only care about my grade in this course.
Your nose is too big for your face.

But children are not systematically corrected for the types of utterances this volume aims to address, which will hereafter be indicated by a preceding "?" (?*explain me this*, ?*drive him angry*, etc.). Caregivers are much more focused on the content of children's speech than on its form, as long as the message is clear enough. For example, a child who says *Me loves you, mommy* is more likely to get a hug than a grammar lesson, and a young child who utters an impressively grammatical utterance such as *I have just completed a mural on the living-room wall with indelible markers* is unlikely to get positive feedback from most parents. The sorts of formulations that native speakers recognize as odd are also not the sorts of formulations that grammar teachers warn against, since they are so rarely uttered by native speakers that no admonishment is needed.

To be clear, it is not that one *never* hears expressions such as ?*explain me this* (or ?*drive him angry*), or that all speakers judge them to be equally odd. In fact, speakers' judgments are gradient and dependent on a number of interrelated factors that are the focus of this book. But corpus and experimental studies confirm that certain types of utterances are avoided by native speakers much more than would be expected by chance. In order to think about how these aspects of language are learned, it's worth thinking about what speakers and language learners are trying to do.

1.1 The Puzzle

The learner's goal is to comprehend messages, given the forms she witnesses, and to produce forms, given the messages she wants to convey. Therefore, speakers must learn the ways in which forms and functions are paired in the language(s) they speak. These learned pairings of forms and functions are referred to here as grammatical CONSTRUCTIONS. Speakers also aim to express their intended messages efficiently and effectively while respecting the conventions of their speech communities, as discussed more below.

Constructions generally allow us to apply our linguistic knowledge to new situations and experiences. English tends to be particularly flexible in the ways in which constructions are PRODUCTIVE. A few examples of productive uses of familiar constructions are provided in table 1.1, with labels for each grammatical construction provided on the right.

TABLE 1.1. Novel linguistic exemplars that demonstrate the productivity of various constructions

"Hey man, *bust me some fries*."	Double-object construction
"Can we *vulture your table*?"	Transitive causative construction
"Vernon *tweeted to say she doesn't like us*."	*To* infinitive construction
"What a *bodacious thing* to say."	Attributive modification construction

Attested examples are cited in quotation marks. Here and below unless otherwise noted, attested examples come from Google.

TABLE 1.2. Novel formulations that are judged odd by native speakers

?She explained him the story. (cf. She told/guaranteed him the story.)	Double-object construction
?He vanished the rabbit. (cf. He hid/banished the rabbit.)	Transitive causative construction
?She considered to say something. (cf. She hoped/planned to say something.)	*To* infinitive construction
?The asleep boy (cf. The astute/sleepy boy)	Attributive modification construction

At the same time, the constructions exemplified in table 1.1 resist being used productively with certain verbs or adjectives, even when the intended meaning is perfectly clear. Examples that illustrate the lack of full productivity are provided in table 1.2. Under each ill-formed example is a closely related fully acceptable example, in parentheses. The latter are provided to indicate that there are no simple, system-wide explanations for why the odd sentences strike native speakers of English as odd. Thus, constructions can be extended for use with some words (table 1.1), but they are rarely completely productive (table 1.2), even when no general constraints are violated. How is it that native speakers know to avoid certain expressions while nonetheless using language in creative ways? It is no exaggeration to say that this basic question has bedeviled linguists and psychologists for the past four decades.

1.2 The Roadmap

The paradox of PARTIAL PRODUCTIVITY of constructions is what this book aims to address. We will also address several issues that have not widely been viewed as directly related. In particular, chapter 2 includes a discussion of how we learn to circumscribe the meanings of words. Close attention to word meanings reveals that speakers possess a vast amount of rich contextual knowledge about what each word means, and about which other words it tends to co-occur with. But, initially, young children make certain errors. They may call the moon a *ball*, or the mailman *Daddy*, before they learn and become fluent with other words (specifically, *moon* and *mailman*). That is, children need to learn to restrict their use of individual words by witnessing how those words *and other words* are used in particular contexts. The rest of the book argues that the same mechanisms involved in learning and restricting word meanings are used when learning and restricting grammatical constructions, and that this process explains how we come to avoid formulations such as ?*explain me this*. By beginning with word meanings, I hope to make the discussion of our primary target—the partial productivity of grammatical constructions—more accessible. That is, once we have a better understanding of word meanings, we can tackle grammatical constructions by essentially asking: What would words do?

Chapter 3 outlines the various factors that are relevant to our knowledge of how grammatical constructions are used within a given speech community. These include formal properties (*syntax*), words and partially filled words (*morphology*), meaning (*semantics*), discourse function (*information structure*), and social context. An appreciation of these factors is a prerequisite for solving the *explain-me-this* puzzle. This chapter also highlights the remarkable degree of cross-linguistic variation that exists in how simple clauses are expressed in the world's languages, in an effort to emphasize just how much people must learn in order to use the constructions in their language appropriately.

The proposed solution to the partial productivity puzzle allows both generalizations (table 1.1) and exceptions (table 1.2) to be learned via the same mechanisms. In particular, in chapters 4 and 5, two key factors—COVERAGE and COMPETITION—are discussed. Chapter 4 explains how constraints on meaning and use *emerge*, as witnessed exemplars cluster within the high-dimensional conceptual space in which our representations for language exist. This chapter outlines how clustering licenses creative uses of constructions. In particular, a single factor, COVERAGE, combines variability, type frequency, and similarity; specifically, a new instance is licensed to the extent that the ad hoc category required to contain it has been well attested (has been sufficiently "covered"). Also outlined in this chapter is a useful model for formalizing the required mechanism; namely, an incremental Bayesian clustering algorithm (Barak et al., 2014, 2016; see also Alishahi and Stevenson, 2008; Matusevych et al., 2017).

In chapter 5, the critical role of competition is detailed. As we comprehend utterances, we attempt to anticipate what the speaker will say next, and we are able to use what the speaker actually says to improve future predictions through a process of error-driven learning. Repeatedly witnessing certain formulations in certain types of contexts strengthens the connections between those grammatical constructions and the intended messages-in-context expressed; this results in conventional formulations becoming more accessible for expressing the types of messages that have been previously witnessed. When there exists a readily available formulation that expresses the intended message in the given context, it usually wins out over potential novel formulations. A special effort is required to buck conventional formulations, although this is possible, for the sake of memorability or playfulness (as in the title of this book). But when there is no readily accessible combination of constructions available to express a speaker's intended message-in-context, she needs to extend language creatively.

The proposal is situated in a larger context in chapter 6. Many studies have demonstrated that children are initially less creative than adults: children behave "conservatively" in that they generalize constructions less freely than adults do. Yet other studies have found that children generalize *more* broadly than adults. This apparent paradox is reconciled by recognizing that children are less adept at aligning bits of knowledge within their high-dimensional conceptual space: sometimes they fail to recognize relevant parallels across exemplars, at least with sufficient confidence (and so they behave conservatively); other times they fail to recognize or retain relevant distinctions (and so they generalize or simplify). Appropriate use of grammatical constructions emerges once the relevant conditioning factors for each construction are learned, and the language user becomes more fluent at accessing the appropriate constructions from memory.

Chapter 6 also outlines why adult learners of a second language tend to have particular difficulty avoiding the types of odd formulations this book addresses (including ?*explain me this*). The suggested reasons go beyond the fact that adults receive less input overall, and that the input they do receive is less well suited to learning. In particular, adult learners need to inhibit their well-practiced native language in order to process a new language, and this appears to lead to a reduced ability to take full advantage of the competition among constructions within the new language. Since competition is argued to be key to constraining generalizations via statistical preemption (chapter 5), second-language learners tend to be more vulnerable to producing certain types of formulations that make sense but which native speakers systematically avoid. Additionally, while adults are generally quicker to discern which dimensions of similarity and dissimilarity are relevant to clustering linguistic representations within their hyper-dimensional conceptual space, they are at the same time prone to miss very subtle similarities and distinctions that are not relevant in their first language.

1.3 The CENCE ME Principles

The basic understanding of language that this book outlines is based on the key ideas listed in table 1.3, which are discussed in detail in the following chapters. An acronym of the key words in these principles is EEMCNCE, but EEMCNCE would be impossible to pronounce. So, let us instead use an anagram of EEMCNCE: **CENCE ME**. "CENCE ME," pronounced "sense me," is intended to emphasize the importance of sensible communication. CENCE ME also usefully illustrates productivity, since the phrase itself is a novel use of the transitive construction. The CENCE ME principles spell out some key assumptions of the more general

TABLE 1.3. The CENCE ME principles

A. Speakers balance the need to be **E**xpressive and **E**fficient while conforming to the conventions of their speech communities.
B. Our **M**emory is vast but imperfect: memory traces are retained but partially abstract ("lossy").[a]
C. Lossy memories are aligned when they share relevant aspects of form and function, resulting in overlapping, emergent clusters of representations: **C**onstructions.
D. **N**ew information is related to old information, resulting in a rich network of constructions.
E. During production, multiple constructions are activated and **C**ompete with one another to express our intended message.
F. During comprehension, mismatches between what is expected and what is witnessed fine-tune our network of learned constructions via **E**rror-driven learning.

[a] Our representations are "lossy," a term from computer science, in the sense that they are not fully specified in all detail.

USAGE-BASED CONSTRUCTIONIST APPROACH to language that are widely shared (see, e.g., Bybee, 2010; Christiansen and Chater, 2016; Goldberg, 2006; Kapatsinski, 2018; Langacker, 1988; Tomasello, 2003; Traugott and Trousdale, 2013). The approach also shares much with memory-based EXEMPLAR-BASED MODELS (Aha et al., 1991; Bod, 2009; Bybee, 2002; Daelemans and van den Bosch, 2005; Gahl and Yu, 2006; Kruschke, 1992; Nosofsky, 1986). The CENCE ME approach emphasizes that exemplars—structured representations—cluster within a hyper-dimensional conceptual space giving rise to emergent constructions, which are then extendable as needed for the purpose of communication.

Individual languages can and do vary in striking ways, as will be emphasized, but the usage-based constructionist approach adopted here suggests that the CENCE ME principles are at work in every natural language, serving to constrain and shape the range of possible human languages. The present book emphasizes examples in English because the majority of the experimental and modeling work to be described has been done on English, and because English is the language I know best.

My understanding of what a construction is has evolved. Early on, I adopted the following definition:

> C is a CONSTRUCTION if and only if C is a form-meaning pair $<F_i, S_i>$ such that some aspect of F_i or some aspect of S_i is not strictly predictable from

> C's component parts or from other previously established constructions. (Goldberg, 1995, 4)

Later, I recognized that this definition was too narrow. Our knowledge of language comprises a network of constructions, and we clearly know and remember conventional expressions even if they are in no way idiosyncratic. So I broadened my definition of constructions as follows:

> Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency. (Goldberg, 2006, 5).

The present volume offers a still more inclusive understanding of what constructions are, motivated by a better appreciation of human memory, learning, and categorization. Here, as explained in the following chapters, constructions are understood to be emergent clusters of lossy memory traces that are aligned within our high- (hyper!) dimensional conceptual space on the basis of shared form, function, and contextual dimensions.

Proponents of alternative perspectives or readers who wish to compare the present proposal with other proposals in more detail may find chapter 7 particularly relevant. There, several recent alternative proposals that aim to account for the partial productivity of constructions are discussed. These include, for example, the idea that speakers avoid straying from what they have witnessed ("conservatism via entrenchment"), that it is useful to posit invisible syntactic diacritics or underlying structures without specifying how these are to be identified by learners, that putting a cap on the number of exceptions and a floor on the number of instances that follow a generalization will ensure how and when generalizations are productive (the Tolerance and Sufficiency principles of Yang [2016]), or that incorporating degrees of uncertainty into formal rules is predictive (O'Donnell, 2015). While aspects of each of these proposals have merit, we will see that the usage-based constructionist approach, described by the CENCE ME principles, explains the facts more fully. The final chapter stands back and puts the discussion in a broader context, while raising several outstanding issues that remain to be addressed.

1.4 Speakers Are Efficient and Expressive and also Conform

Before leaving this introductory chapter, let's go over the first of the CENCE ME principles: We aim to express our messages effectively and efficiently while obeying the conventions of our speech communities. To clarify the key terms involved:

1. **Expressiveness**: Linguistic options must be sufficient for conveying speaker's thoughts, beliefs, and attitudes in ways that listeners are able to understand.
2. **Efficiency**: Fewer and shorter constructions are easier to learn and produce than more or longer constructions.
3. **Obeying conventions**: Learners attempt to use language in the ways that others in their language communities do.

A language is only sufficiently expressive if it has the means to adequately convey a speaker's thoughts, beliefs, and attitudes in ways that avoid failures of communication. A *maximally expressive* language might have an ever-increasing number of words and constructions, with every potential distinction indicated by a unique form. On the other hand, a *maximally efficient* language would have a single, easy to learn and use form (perhaps the form, *ah*). The fact that language users need to be both effective *and* efficient requires natural languages to find a balance between these two opposing factors, as has been long discussed by functional linguists (Briscoe, 1998; Bybee, 1985, 2003; Givón, 1979; Goldberg, 1995; Grice, 1975; Haiman, 1985; Levinson, 1983; Paul, 1888; Slobin, 1977; von Humboldt, [1832] 1999).

The recognition that languages must be efficient and expressive, and that these pressures are mutually constraining, has gained new traction within the "noisy-channel" approach to language processing, which recognizes that speakers are attempting to express information as efficiently as possible, under imperfect or noisy conditions (Gibson et al., 2013; Jaeger and Levy, 2006). The noisy-channel approach has emphasized the dynamic nature of the balance between efficient and expressive communication. When a speaker is reasonably certain that an intended message will be successfully conveyed, the balance is tipped toward efficiency, with forms being reduced and distinctions being underspecified (Hopper and Traugott, 2003; Jaeger, 2010; Levy, 2008; Levy et al., 2009; Piantadosi et al., 2011, 2012). For example, when a verb appears with the construction it most commonly appears in, the verb form itself tends to be reduced (Gahl and Garnsey, 2004). Similarly, the complementizer *that* is more likely to be omitted when it is predictable in context (Wasow et al., 2011). On the other hand, under noisy or uncertain conditions, distinctions may be exaggerated, and language may be made less ambiguous in various ways (Bradlow and Bent, 2002; Buz et al., 2014; Gibson et al., 2013; Hall et al., 2013). Thus, efficiency and expressiveness balance each other and lead languages to vacillate between using shorter or fewer forms to express a given message on the one hand, and adopting new, longer, or additional forms in order to ensure that messages are understood as intended, on the other.

The last idea, that speakers tend to obey the conventions of their language community, captures the fact that humans treat language as a normative enterprise. For example, people within a given community tend to believe that there is a "right" way to pronounce words, even if other communities are recognized to pronounce them differently. The Gershwins' famous lyric, "You like tomato and I like tomahto, . . . Let's call the whole thing off" epitomizes this idea. Humans are a rarity within the animal kingdom in using arbitrary communicative symbols that are shared within a community and distinct from those used in other communities (Tomasello, 2016). In fact, humans quite generally appreciate that there are culture-specific "right" and "wrong" ways to do a great many things, and we learn to obey these conventions in a way that other species do not (Boyd and Richerson, 1988; Horner and Whiten, 2005). Many normative conventions are, at least initially, self-conscious; for example, our knowledge about how to eat food politely, whether or how much to tip at restaurants, or whether it is polite to sneeze or burp in public. Other social norms may be obeyed without self-conscious awareness, including how close to stand to each other while speaking, or what sort of foods are appropriate for breakfast.

The importance of cultural norms for human behavior has enjoyed a long and rich appreciation within philosophy (e.g., Korsgaard et al., 1996). Our respect for normative patterns of behavior is what allows us to create complex cultural practices. For example, dollar bills would be meaningless were it not for the social agreement that imbues them with value. Driving would be terrifying if we couldn't rely on other drivers to (generally) obey standard driving practices. Again, work that has compared humans with other primates has emphasized that cultural norms may be uniquely human (Tomasello, 2009, 2016), particularly when they serve no clear purpose (e.g., Horner and Whiten, 2005; McGuigan et al., 2011).

Normativity is critical to the *explain-me-this* problem in that generations of learners obey restrictions that do not serve any clear communicative function: we respect the patterns that are evident in the input. For example, even though saying *she made it vanish* is somewhat less efficient than *she vanished it*, and even though the latter formulation is readily interpretable, native English speakers avoid the shortcut and normatively obey a shared preference for the periphrastic form. A mechanistic explanation supports our tendency to produce forms that our community deems acceptable, insofar as partially familiar formulations are easier to access than wholly novel formulations. That is, more conventional forms are more efficient to access from memory even if they are less efficient to produce (see section 4.5). But the reason native speakers judge the longer phrase (*explain this to me*) to be the "right" way to express the intended meaning and the shorter phrase (?*explain me this*) to be "incorrect," and the reason that familiarity tends to be more important than ease of articulation, is because we desire to speak like others in our community—language is a social and normative enterprise.

Chapter 2 briefly reviews the nature of word meanings and asks how we learn to use words appropriately. We will then see in the following chapters that many of the lessons learned from an appreciation of word meaning extend naturally to our primary question: How do we learn to use basic clause types—ARGUMENT STRUCTURE CONSTRUCTIONS (ASCs)—in creative but constrained ways?

Chapter 2

Word Meanings

As described in chapter 1, the primary goal of this book is to provide an explanation of how native speakers of a language know that they can combine words in certain new ways but not others. I refer to this as the *explain-me-this* puzzle, since this phrase provides a good example of the phenomenon: the intended meaning of ?*explain me this* is perfectly clear, but it nevertheless does not sound natural to native speakers of English (hence it is preceded by "?"). Before addressing this issue, we first focus on the simpler case of how speakers learn to use individual words appropriately, because there are several key parallels between the two problems, insofar as both cases require learning how to use language in ways that are creative and yet constrained.

2.1 Meanings Are Rich, Structured, and Partially Abstracted

Many of us don't reflect on word meanings very often, and so we may fail to appreciate just how much intricate knowledge is required for us to use words appropriately. Quite specific knowledge is needed, for example, to distinguish the more than a dozen verbs that imply that a person is forced to leave a situation or place (see table 2.1). Each of these words is appropriate in a distinct range of contexts, even though they all designate the same abstract type of event represented in figure 2.1.

Several of the words in table 2.1 imply that the removed person has somehow transgressed; they differ in terms of which organization or place the person is removed from. *To banish* is to remove a person from civilized society; *to expel* (in one sense of the word) is to remove from a school; *to deport* is to remove

TABLE 2.1. Partial list of verbs that designate the removal of a person from a situation or place

Banish	*Dismiss*	*Extradite*
Blackball	*Evacuate*	*Fire*
Blacklist	*Exile*	*Kick out*
Cast out	*Expatriate*	*Lay off*
Deport	*Expel*	*Oust*

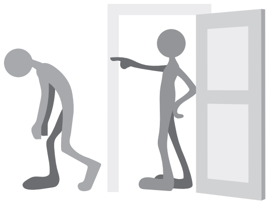


FIGURE 2.1. Concrete depiction of the general type of abstract event profiled by the verbs in table 2.1.

from a country; *to fire* is to remove a person from his job. *Lay off* differs from *fire* in that there's no stigma attached to the person who has to leave his job. A person who is *blacklisted* is barred from some type of work, often unfairly and for political reasons. People can be *ousted* only if they are political figures. This sort of contextual information or world knowledge is part of our knowledge of language, an observation that will be relevant as we approach the *explain-me-this* puzzle in the following chapters.

Different types of world knowledge are needed for different types of word meanings. Knowledge of quite specific motor sequences is required to distinguish the meanings of *walk*, *skip*, *hop*, *jog*, *power walk*, *run*, *race*, *sprint*, and *waltz*. While it would be difficult to try to explicitly define these terms, we can identify the actions at a glance. Moreover, we have contingent associations with many of these terms because of how they have been witnessed in the past. We know that *skipping* is most often done by children and is associated with being carefree, *power walking* is more commonly done in malls and by mothers or retirees, and *waltzing* is associated with formal events such as weddings.

In fact, once you start to think of the actual range of words that we know, it becomes clear that word meanings are not readily reducible to any finite list of recognizable features or attributes (Fodor et al., 1980) and word meanings are not easily disentangled from knowledge about real-world contexts in which these words tend to occur (Palmer, 1996; Willits et al., 2015). For example, if asked to try to define the word *bachelor*, many people will suggest "unmarried man" (Fillmore, 1975), a meaning that is often cited in philosophy classes because it seems to be definable in terms of simple features: unmarried and male. But this definition would seem to classify as bachelors the Pope, Tarzan, unmarried men with long-term live-in partners, and unmarried men on life-support. As Fillmore noted long ago, something clearly seems to be missing from this definition. He pointed out that the meaning of *bachelor* actually seems to be defined against a background of stereotypical world knowledge. This stereotypical knowledge takes for granted that boys grow up, date multiple people for a period of time, get married and stay married. Within this stereotype, there is clearly a period of time when an unmarried man is a bachelor. Fillmore suggested that each word meaning evokes a conventional SEMANTIC FRAME, where a frame is a structured abstraction or idealization over a set of coherent contexts (Fillmore, 1975, 1977, 1982, 1984; see also closely related ideas by Austin, 1962; Bartlett, 1932; Lakoff, 1987; Langacker, 1987). The word *bachelor* profiles an unmarried man, but it includes the relevant background frame of the stereotypical series of events just mentioned. When the background frame does not apply, neither does the word.

Semantic frames capture rich aspects of our world knowledge as is required for adequate characterizations of word meanings. Experimental work has found support for the psychological relevance of contextually evoked semantic frames of knowledge. For example, Bar (2004) has demonstrated that people perceive the same visually filtered shape to unambiguously be a hairdryer, if placed in the context of a bathroom sink, but to be a drill if placed in the context of a workbench (see also Beck and Kastner, 2009; Biederman et al., 1982; Murphy and Wisniewski, 1989; Walther et al., 2009; Zettersten et al., 2018). When we interpret individual words, neural areas that relate those words to various actions or perceptions may also be reliably activated, suggesting that our sensory knowledge, whether directly experienced or imagined, is linked to word meanings. More specifically, words referring to actions performed by the leg, arm, or face (e.g., *kick*, *pick*, or *lick*) have been found to evoke activation at or near the areas in the primary motor cortex that are active when we move our feet, hands, or tongue (Hauk et al., 2004). Activation of our motor cortex is also evoked when certain words are used metaphorically—e.g., *to grasp the idea*, *to kick an idea around* (Cacciari et al. 2011; Desai et al. 2011; Pulvermüller et al., 2005). Texture-selective regions in the somatosensory cortex are likewise activated by words related to texture, whether used literally or metaphorically—e.g., *She had a rough day* (Lacey et al., 2012)—and cortical areas that are evoked by tasting food are activated when words related to taste are used in isolation or in conventional metaphors—e.g., *That was a bitter breakup* (Citron and Goldberg, 2014). In fact, different words reliably evoke activation that spans almost the entire cortex, suggesting that word meaning can involve a wide range of associations (Bergen, 2012; Huth et al., 2016). Thus, words evoke rich conceptual and perceptual information gleaned from the contexts in which the words have been witnessed.

2.2 Vast Implicit Memory

How do we *learn* the rich structured representations required for word meaning? One prominent philosopher found the question so challenging that he concluded that we must actually be born with all possible current, past, and future word meanings, so that our job as learners is merely to attach a label to an already existing concept (Fodor, 1975). But this idea leads to the uncomfortable stipulation that Abraham Lincoln was born with meanings (if not labels) for concepts such as *computer*, *fax*, and *drone*, and that your grandmother was born with knowledge of the meanings of *frack*, *twerk*, and *rap*. This seems unlikely.

It is clear that even a single initial encounter with a word can potentially leave a memory trace of its use. Why is this clear? Consider the counterfactual: if no memory trace could exist after a single exposure, then the second time the word was encountered would be exactly the same as the first time. But then no memory trace of the word would be left upon this second encounter. This situation could be repeated ad infinitum without any memory trace of the meaning being retained. If this were the case, we would be utterly unable to learn any words. Therefore, it must be possible for an initial memory trace to exist in order for it to be strengthened upon subsequent exposure. Fortunately, we know that human brains have a vast capacity for implicit memory, even though memories may not readily be brought to consciousness (they are not always easy to *recall* or make *explicit*). In the domain of vision, for example, Brady and colleagues (2008) showed a group of participants 2500 images of distinct objects for three seconds each; remarkably, participants were able to recognize the pictures they had seen at well above chance rates, in that they were successfully able to select the picture they had seen from another picture of the same category (e.g., another dinner bell, star fish, or wooden desk). Participants were even able to distinguish the pictures they had seen from another picture of the *same object* positioned differently (e.g., a cabinet with doors closed or with one door ajar) (see also Standing, 1973). It seems that at least some of these fleeting memory traces become consolidated into long-term memory while we sleep (Marshall and Born, 2007; Stickgold, 2005).

An impressive capacity for memory for words has been found as well. Whenever we hear "the same" word form, it is pronounced slightly differently, as it may be spoken by a different speaker, or at a different speed, or in the context of different other words. While we readily generalize across these different tokens in order to appreciate that they involve the same word, it turns out we are faster to identify a repetition of a word if it is spoken by the same speaker, suggesting that aspects of that speaker's voice can be retained with the memory, at least for some period of time (Kleinschmidt and Jaeger, 2015; Palmeri et al., 1993). In fact, people find it easier to recall words from a list if they are tested in the same room in which they had originally read the list of words; this advantage can also be achieved if they are asked to imagine the original room just before recalling the list of words, as if the imagining is sufficient to reinstate the original context (Smith, 1979; see also Godden and Baddeley, 1975). In another experiment, people were shown a series of words (e.g., *spoon*, *table*, *house*) and pictures (of a fork, chair, shed) and subsequently queried, as they lay in a functional magnetic resonance imaging (fMRI) scanner to record their brain activity, whether or not they had seen a particular word or picture. In answer to a query about either *spoon* or *fork*, for example, the correct answer would be "yes." The queries were all made verbally; that is, the pictures seen initially were not displayed a second time. And yet the brain's *visual* cortex was found to be more active for those concepts encoded by pictures during the experiment (Vaidya et al., 2002). That is, it appears that the word label used during the query (e.g., *fork*) evoked the visual memory of the picture (of a fork) if a visual image was witnessed in the context of the experiment. Since very familiar words were used in the study, the findings demonstrate that the contexts associated with a word are continuously updated in a context-sensitive way (see also Goldinger, 1998; Hintzman, 1988).

Another way to make clear that we retain rich contextual information about how words are used stems simply from the fact that our vocabulary is rich and highly nuanced as already alluded to. This in itself provides evidence that we *must* retain quite specific contextual information about how words are used as we learn the meanings of those words (see also Borovsky et al., 2010; Johns et al., 2016; Nelson and Shiffrin, 2013; Walker and Hay, 2011). For instance, imagine that a child hears the word *write* for the first time in a sentence like *She's writing it now.* Does the child record the speed and care that is used to write? Exactly what was written? Whether the writing was done on paper or on a wall? Whether or not the writer was famous? Although it may seem unlikely that all of these aspects of the nonlinguistic context would be recorded, we *have to be able* to record detailed aspects of contexts as we learn the meanings of words, or else we would not be able to learn, without explicit instruction, the correct meanings of words such as *scribble* (write quickly), *scrawl* (write carelessly), *sign* (write one's name in cursive), *endorse* (write one's name in cursive on a check), and *autograph* (write one's name when one is famous). Thus, our memory for words is not only vast but also linked to the contexts in which those words are experienced.

At the same time, it is reasonable to assume that aspects of contexts that are perceived to be more relevant or that are more uniquely predictive of a given word are more likely to be retained (or are weighted more heavily). That is, surprising or unusual aspects of contexts are more likely to be encoded in memory than aspects that are ever present. Aspects of contexts and words that are more highly correlated (have higher MUTUAL INFORMATION) are also more likely to be retained. Of particular importance to children are the speakers' perceived intentions, as the child attempts to interpret what speakers mean by what they say (Austin, 1962; Carpenter et al., 1998; Clark, 1996; Levinson, 1983; Tomasello, 2001; Tomasello and Barton, 1994). In short, the representation of each encounter prioritizes information and predictive cues that are perceived to be relevant.

Moreover, memory for any experience is necessarily partially abstract insofar as experiences are not represented completely veridically. We can describe the representations of events as involving LOSSY COMPRESSION, by which we mean simply that not *all* information is retained. For example, we might have a memory trace of witnessing a kumquat that is abstracted away from the color of the kitchen table upon which it sat, the tiny scratch in its surface, and the length of its stem. As Christiansen and Chater (2016) emphasize, the rapid time scale of

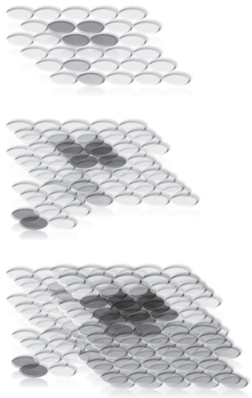


FIGURE 2.2. *A*, an abstract, structured, distributed representation of a single memory trace of a word witnessed in context; *B*, strengthened representation of the word as aspects are repeated over time (represented by darker nodes); *C*, representation of the word as it continues to be strengthened through experience.

language processing *requires* that our brains recode and compress incoming information.[1] Thus memory traces of experiences, no matter how vivid, are partially abstracted from our experience.

To summarize, an initial encounter with a word can form a lossy structured representation that prioritizes what the word designates and includes various contextual aspects of the encounter that are perceived to be informative or relevant to the use of the word, which may include quite detailed information about form, meaning, and context. Illustrated schematically in figure 2.2A is a single such representation, with more important aspects indicatd by darker ovals. As suggested by figure 2.2B, additional encounters with the same word typically overlap in some ways with the earlier representation, strengthening those shared aspects, while also potentially adding contextual information that is unique to that particular experience (Atkinson and Shiffrin, 1968; Light and Carter-Sobell, 1970). The result of multiple encounters of a word is a dynamic cluster of overlapping structured representations (figure 2.2C) situated within our HYPERDIMENSIONAL CONCEPTUAL SPACE. As indicated in the changes from figure 2.2A to 2.2C, the representation of a word becomes broader as aspects of each context are added. At the same time, those aspects of memory representations that overlap across multiple encounters of a word become strengthened over time, thus becoming more central to a word's meaning. In figure 2.2, overlapping aspects are indicated by the progressively darker parts of the representation. Figure 2.2 is also intended to convey that as memory traces accrue, their status as unique encounters gives way to an EMERGENT CLUSTER (or "cloud"), which constitutes what we think of as a single coherent word meaning (or *lemma*).[2]

1 Christiansen and Chater seem to imply that the "continual deluge of linguistic input" acts to "obliterate" previous linguistic material due to interference (2016, 1), but it is clear from conversations with both Morten Christiansen and Nick Chater that they fully recognize that memory traces of language experiences at different levels of generalization are retained in long-term memory (as long as they are consolidated, which often requires a night's sleep), as just emphasized.

2 Although the representations in figure 2.2 portray tight clusters in a small space, the clusters of neurons that are evoked by individual words typically span multiple brain areas.

The required representational space for language is hyper-dimensional and it must be part of our conceptual system because such an open-ended array of contextual factors is required for our representations of word meanings. That is, as described above, word meanings involve complex relational structure and do not correspond to lists of recognizable features (try to decompose the word *extradite* into a list of features and you'll appreciate this fact).

The formal domain (e.g., of sounds, word forms, word order) is quite a bit less variable than the open-ended domain of nonlinguistic context. Because of this, the formal similarities among representations play a special role in determining how our representations of experiences with words cluster together. Individual words are represented by a cluster of abstracted sequences of sounds and structured context-based semantic representations. The cluster emerges on the basis of similarities and parallels within representations, and from differences between other existing clusters. Thus, a "word" is in fact a cluster of partially overlapping structured representations within our hyper-dimensional conceptual space (see also Elman, 2009; Kapatsinski, 2018; Pierrehumbert, 2002).

At the same time, the conceptual space that is used for language does not cover our *entire* perceptual and conceptual knowledge or capacity: we have a myriad of nuanced experiences of smells, feelings, visual knowledge, and autobiographical memories that are not necessarily captured by any conventional words or constructions. In addition, as we will see, there exist dimensions that may be highly relevant to linguistic representations in just a few languages. Therefore, the high-dimensional conceptual space used for language is a subset of our even more complex perceptual and conceptual system.

The representations in figure 2.2 are intended to capture the fact that each word's presentation is affected by how often it is encountered (its TOKEN FREQUENCY) and the range of contexts in which it has been encountered. And, indeed, it is well established that words that have been encountered more frequently are reliably faster and easier to access in a range of tasks (Broadbent, 1967; Jescheniak and Levelt, 1994; King and Kutas, 1995; Meyer and Schvaneveldt, 1971; Oldfield and Wingfield, 1965; Rayner and Duffy, 1986). We even have explicit intuitions about the relative frequencies of words: we recognize that *elephant* is more frequent than *pachyderm*, and that *dog* is more frequent than *elephant*. People are also able to judge quite accurately the relative frequencies of words encountered within experimental settings (e.g., Balota et al., 2001; Brysbaert and New, 2009; Hintzman, 1988).

The variety of contexts in which a word has been witnessed also appears to play a key role in how quickly and accurately that word is accessed (Adelman et al., 2006; McDonald and Shillcock, 2001). Words that have only occurred in a narrow range of contexts may be restricted to those contexts. For example, no matter how many times we may have read Dr. Seuss's *Yertle the Turtle*, none of us consider *Yertle* to be a popular name. Instead, it immediately evokes the Dr. Seuss book. When a British friend says she is *chuffed*, I may understand its intended meaning in the context ("pleased"), but I am unlikely to use the word myself because I will associate the word with a different dialect of English.

This perspective predicts that knowledge of a word is not an all-or-nothing affair. The ability to correctly identify the meaning of a novel word in context (e.g., Carey and Bartlett, 1978; Woodward et al., 1994), does not entail the ability to successfully recall the meaning or even recognize the word after a delay. In one study, Horst and Samuelson (2008) demonstrated, for example, that two-year-old children were able to assign four new labels to novel objects when each new word was presented alongside two objects with familiar names (e.g., a pair of glasses, a toy dog, a "cheem"). But the children showed no evidence of retaining the word-object mappings after a five-minute delay when the new object (e.g., the "cheem") was presented along with two other novel objects. Therefore, although we retain a great deal of implicit information upon initial encounter, our ability to actively make use of the word-level abstraction over contexts develops through time and experience with the word (see also Fernald et al., 1998).

2.3 Clusters of Conventional, Related Senses

Most of the time, new contexts are relevantly similar to familiar contexts and so we can simply use our learned words to convey familiar meanings. But sometimes we encounter importantly new contexts, in part because the world itself is constantly changing, and in part because our relation to the world is changing. When faced with especially new contexts, we may need to use old words in creative ways. For example, the terms *file*, *folder*, *paper*, and *trash* all took on new meanings when they were applied to computer interfaces; they were immediately interpretable because they relate in clear ways to their original senses. A few other words that have acquired new conventional senses in addition to their original senses are provided in table 2.2. The employment of an old word for a new purpose is efficient for the speaker, who does not need to coin a new term and risk communication failure; it is also effective for the listener, who is able to use knowledge of an existing sense of a word as scaffolding to a new sense. The familiar meaning provides an indication of what is intended. Initially, extended senses are created on the fly, but if a new use of a familiar word catches on in a wider community, the new sense becomes part of the conventional meaning of the word. This gives rise to conventional POLYSEMY: a network of related senses for individual words (Brocher et al., 2017; Copestake and Briscoe, 1995; Floyd and Goldberg, forthcoming; Geeraerts, 1993; Klein and Murphy, 2001; Lakoff, 1987; Tuggy, 1993).

TABLE 2.2. Familiar words that have acquired new senses relatively recently

Word	Prototypical original sense (still in use)	Newer sense
Bandwidth	The maximum amount of information that can be transmitted along a channel	The mental capacity of a person, as in, *I don't have the bandwidth to set up a carpool*
Lit	Verb: past tense of *to light*	Adjective: drunk or exciting (*I was lit*; *The party was lit*)
Friend	A close intimate that isn't family	Anyone given explicit access to a Facebook feed
Optics	A field of study concerning vision	How a public event is perceived
Spam	Brand name of mass-produced inexpensive and unappealing mystery meat	Mass-produced unsolicited and unappealing e-mail
Literally	True without exaggeration or metaphorical interpretation	Emphasis without being true, as in *We were literally killing ourselves laughing*[a]

[a] From the *Oxford English Dictionary*, 3rd edition.

For example, a prototypical sense of the verb *to fire* involves a very quick, violent triggering of a bullet from a gun that can lead to injury or death and is often performed out of anger. When someone is *fired* from a job, the quick and brutal implications of the prototypical sense of firing a gun are retained, but other aspects are not. We can also *fire off* a letter, which shares the sense of a quick action, performed with a sense of urgency or out of anger. When neurons *fire* we understand there to be a quick and directed event, but of course no volition, addressee, anger, or harm. If we were to try to distill the meaning of the verb *to fire* into what all senses have in common, we might come up with a meaning such as "any quick, directed action." But this meaning would be too general, since it would seem to imply that chopping a tree or breaking an egg into a bowl could be instances of firing, and it would also fail to capture the implication of most of the senses (but not the firing of neurons) that harm is intended. Rather, multiple senses often cluster around a rich, prototypical (often stereotypical) semantic frame, with extensions based on some but not all of the attributes of this frame (Lakoff, 1987). The senses just mentioned are described

TABLE 2.3. Related senses of the verb *to fire*

Original prototype on which other conventional senses arose historically: *to fire* a gun	The quick action of pulling a trigger on a gun in a directed way, with the intention of causing a bullet to strike something; repeated rapidly; often done out of anger and with the intention of causing physical harm or death
To fire someone	To lay off someone from employment, typically resulting in financial and emotional harm to the employee; done quickly (without warning); sometimes done out of anger
To fire questions or insults at someone	To communicate comments in a quick, repeated rapidly and harsh way
Neurons *fire*	The action potential of a cell sends an electrical signal down the axon (quick, directed)
To fire off a letter	To quickly direct a communicative act toward someone, typically out of anger

in table 2.3 and represented in figure 2.3.[3] The observation that words are commonly associated with a "radial category" of senses radiating out from a rich prototypical semantic frame is due to Lakoff (1987), who argues that clusters of such senses exist for most nouns and verbs, except those that are used infrequently or in restricted technical contexts (Dautriche, 2015; Durkin and Manning, 1989; François, 2008; Heylen et al., 2015; Tuggy, 1993). To summarize, it is clear that words convey rich frame-semantic meanings, typically associated with a network of related meanings because new senses arise from the need for words to apply to new contexts.[4]

3 The prototypical meaning of the verb *fire* is also related to the noun *fire* (not pictured), since the firing of weapons commonly involved a flame. The noun *fire* is directly related to additional verbal senses including *to fire clay* and *to fire up a conversation*.

4 Over time, an original meaning may die out, leaving only the newer sense(s). For example, *divest* used to mean *undress* or *deprive others of rights or possessions*. These days, of course, it is restricted to the selling off of investments of a certain type. The word *clue* used to refer to a ball of yarn; since we can talk about "following a thread" to figure something out, the new and old senses are likely related but the original meaning of *clue* has been lost. More

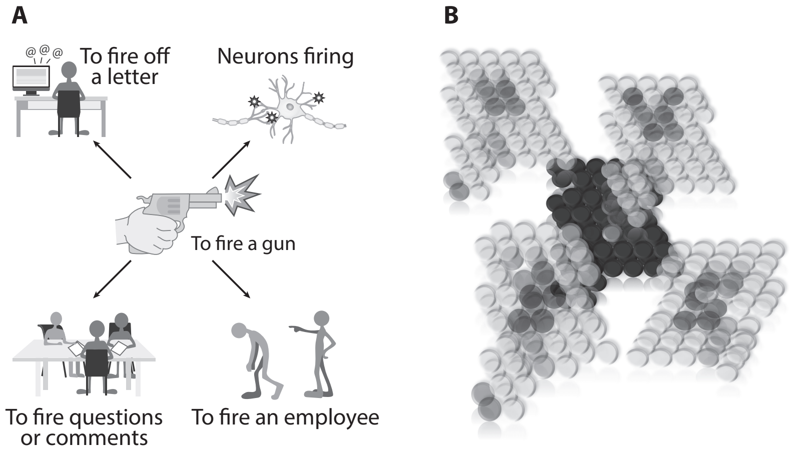


FIGURE 2.3. *A*, symbolic, and *B*, abstract, structured, distributed representation of related senses of the verb *to fire*.

2.4 Creativity

Assigning meaning to words is recognized to be a difficult and complex task (Gleitman, 1990; Quine, 1960). This has led to the idea that meanings that are initially associated with a word are routinely incorrect, and in need of correction or elimination. One popular idea is that learners form a tentative hypothesis about what a new word means, and then either confirm or reject that meaning when they encounter the word in a new context. If the new context is inconsistent with the hypothesized sense, this thinking went, the hypothesized sense would be dismissed, with learners essentially deleting the initially proposed association between the word and the original sense (Aravind et al., 2018; Berwick and Weinberg, 1986; Siskind, 1996; Trueswell et al., 2013; Woodard et al., 2016).

But armed with a fuller appreciation of word meanings, we can see that this "Propose but Verify" model, as it was called, cannot possibly work, because words typically have multiple distinct senses. Encountering a second sense of a word cannot eliminate the first sense. For instance, hearing *fire* used in the context of employment cannot lead the learner to eliminate the sense of *fire* associated with guns. It cannot even lead the learner to delete particular aspects of the hypothesized senses, because multiple senses of a word do not necessarily share any particular attribute among them. For example, consider the word *breakfast* (Fillmore, 1976). If a child initially encounters the word *breakfast* in the

commonly, old and new senses coexist as we've seen in the case of the multiple senses of *fire*, and all of the examples in table 2.2.

context of being told to *eat your breakfast*, the child may (correctly) hypothesize that *breakfast* refers to "food, like this cereal, that constitutes the first meal of the day." But what if a child later hears that she can order breakfast in the evening at a diner? Should she jettison her original hypothesis? Clearly not. Should she delete the restriction on the time of day and retain "food such as cereal"? The latter interpretation would be sufficient for these first two contexts, but what is the child to think when she subsequently hears that someone had pizza for breakfast? Adding this third context would require eliminating the hypothesis that *breakfast* refers to either "food such as cereal" *or* "food that constitutes the first meal of the day." But then the child is left with no meaning for *breakfast* at all!

The Propose but Verify model has since been updated by a "Pursuit" model, to allow for hypothesized meanings to be incrementally suppressed, rather than automatically eliminated, whenever a hypothesized meaning is incompatible with the current context (Stevens et al., 2017). Yet, oddly enough, this model stipulates that only one potential sense for a word is considered:

> If the most probable hypothesis fails to be confirmed, the model does not test out the second most probable hypothesis, but rather chooses a meaning from the current context instead. Furthermore, the selection of the new meaning also follows a minimalist strategy: If there are multiple meanings available, the model does not favor meanings it has seen before but chooses completely randomly. (Stevens et al., 2017, 12)

For this reason, the Pursuit model, like the Propose but Verify model before it, fails to allow words to be systematically associated with multiple related meanings (see Floyd and Goldberg, forthcoming).

Instead of deleting or suppressing a hypothesis that *breakfast* means "stuff like this cereal that is eaten in the morning," a child would be better off retaining this sense and *adding* additional senses, including "food, like this cereal, regardless of when it is eaten" and "food, regardless of what it is, that constitutes the first meal of the day." The first context instantiates the most common or prototypical sense of the word, after all, and the latter two senses are conventional extensions of that sense. The suggestion then is that memory traces of word forms and contexts are only added; an association between a word and meaning cannot be eliminated by additional encounters with the same word.

For the most part, past the age of a year and a half, learners are remarkably accurate at assigning the correct meanings to novel words, as they make use of a variety of cues to determine intended meanings (Akhtar et al., 1996; Baldwin and Tomasello, 1998; Childers and Tomasello, 2002; Imai et al., 1994; Landau et al., 1988; Markman, 1989; Mervis et al., 1994; Soja et al., 1991). In assigning an interpretation, even young children dynamically take into account not only their own knowledge of the discourse context, but also what they take to be the speaker's current state of knowledge. For example, Tomasello and Haberl (2003) found that 12- and 18-month-old infants, who were equally familiar with and interested in three toys, tended to offer the one toy that was new to an adult, when she asked, "Oh, wow! That's so cool! Can you give it to me?" (see also Gweon et al., 2014). That is, even very young children recognized that the adult was likely interested in the toy that was novel to that adult. Still, of course, learners can and do make errors of varying types. Can the idea that contexts are only added actually work?

2.5 Competition Constrains Word Meanings

How can a child ever *unlearn* a sense that is incorrectly associated with a word? For example, what if she incorrectly assumes that *breakfast* means "cereal"? Witnessing *breakfast* in additional contexts of cereal would not dissuade her of this sense, and we have just argued that witnessing *breakfast* in contexts *without* cereal should also not dissuade her of this sense. How, then, can the child come to dissociate the meaning of *breakfast* from the more specific meaning "cereal"?

In fact, the word *cereal* must come to the rescue. In contexts in which cereal but not breakfast is intended, the child will reliably hear the word *cereal*. The new association between *cereal* and "cereal" will eventually become stronger than the previous association between *breakfast* and "cereal." That is, each word fills its own semantic and distributional niche (Aronoff and Lindsay, 2016; Clark, 1987; Gauger, 1973). Words are not interchangeable and true synonyms are vanishingly rare. Each word is distinguishable from other words because each is associated with a rich network of contexts or semantic frames, as already discussed. Near synonyms reliably have different contextual restrictions or background frames; they may differ, for example, in terms of formality (*dog* vs. *pooch*), perspective (*ceiling* vs. *roof*), or attitude (*skinny* vs. *slim*).

Since we have seen that our memory for words includes aspects of their contexts of use, it is reasonable to assume that intending the same type of message in a relevantly similar context will serve to activate words used previously in those types of contexts to express that type of message. Whenever multiple words are activated and cannot be combined, they compete.

Thus, if children initially incorrectly assign a meaning to a word, they will eventually learn a different term that is more appropriate for the intended meaning. In this way, word forms are in competition with one another for meaning: *cereal* will eventually STATISTICALLY PREEMPT the incorrect association between *breakfast* and "cereal." If there is no more appropriate term, then there is no need for children to modify their sense of the word, and there is no evidence that they do. Instead, appropriate interpretations of a word will be strengthened over time as the word is witnessed in additional contexts.

The fact that words influence the potential meanings of other words is clear if we consider examples from other languages. For example, in German, the word *Blase* means "bubble" and is used to refer to blisters as well as soap bubbles. But among English speakers, only children or second-language users are likely to refer to a blister as a *bubble* because the word *blister* exists and is a more appropriate alternative. Ancient Japanese used a single word, *ao*, to describe either blue or green color, but English speakers cannot call the sky *green* or the grass *blue* because the other term is better suited to convey the intended meaning. Conversely, we rather indiscriminately speak of *knowing* how and *knowing* something, whereas French distinguishes the two types of knowledge with distinct verbs, *savoir* and *connaître.*

The hypothetical *breakfast* for "cereal" example discussed above is a case of a child initially *under*generalizing the meaning of a word; but what if she initially *over*generalizes the word's meaning? In the early stages of word learning, young children do overuse words in their small vocabularies in ways that are not conventional in the language (Bloom, 1975; Clark, 1973). For example, children may use the word *ball* to refer to a button or the moon. They may use the word *dog* to label all animals, and they may say *hi* when they should say *bye*. These overextensions are most common when children's vocabularies consist of only 50–150 words (Gershkoff-Stowe, 2001), an indication that they simply do not have a better word for the meaning that they wish to express at that particular moment. In an important study, Kuczaj (1982) found that when asked to "show me the dog" or "show me the ball," the same children who overgeneralized these words in their own utterances reliably chose a picture of a dog over pictures of other animals and chose a picture of a ball over pictures of other round things (see also Gelman and Naigles, 1995; Huttenlocker, 1974). That is, children seem to know that, at least, the best examples of the category *dog* are dogs, not other animals, and the best examples of *ball* are actual balls, not other round things. This is expected, given that that memory traces of encounters with words are retained: children may initially use words in an overly broad way, but they recognize that situations which better match their representations are more appropriate. Thus, children produce overgeneralizations when no other word has been learned or is sufficiently accessible at the moment of speaking (see also Harmon and Kapatsinski, 2017).

Overgeneralizations, just like undergeneralizations, are eventually avoided as new words enter the child's vocabulary. The child stops calling the moon *ball*, not because she witnesses *ball* in additional contexts, but because she learns and accesses the more appropriate word, *moon*: *moon* statistically preempts the use of *ball* to refer to the moon. What we learn are correlations between word forms and a range of rich representations that capture senses in contexts. The suggestion then is that different word forms are in competition with one another for a particular construal: if the word *moon* is used to refer to the moon in certain types of contexts, then *ball* is not used in to express that meaning in those types of contexts.

In the following section, we review the evidence in support of this idea. The notion of competition via statistical preemption will also play an important role in our solution to the *explain-me-this* puzzle.

2.6 Learning and Fluency Reduce Overgeneralizations

Distinct words intended to express the same meaning-in-context compete to be used. There are almost always contexts in which one word is preferred over the other. For example, we may believe that *quick* and *fast* are interchangeable, but in fact it is much more natural to talk about a *quick shower* and a *fast car* than a *fast shower* or a *quick car*. This is because *quick* tends to be used when there is a clear end point, whereas *fast* readily refers to speed without invoking an end point. Since, as discussed above, the relevant contexts that become associated with particular words are quite rich, distinct labels are associated with distinguishable CONSTRUALS or MESSAGES-IN-CONTEXT, even when they are in certain ways quite close in meaning (e.g., compare *jock* vs. *athlete*; *thrifty* vs. *stingy*; *banish* vs. *deport*) (Langacker, 1987).

There is ample evidence for competition between word *forms* as words are being learned. By about one and a half years of age, if children are asked to find the *moop* among a group of objects, all but one of which already have a familiar label, toddlers will assume *moop* names the unfamiliar object. This tendency has been described as a bias to assume that a new word refers to a new object ("mutual exclusivity") (Markman, 1989; Markman and Hutchinson, 1984; Markman and Wachtel, 1988; Xu, 2002). But the same object in the world can often be labeled in multiple ways ([tele]*phone*, *cell*, *device*, *thingy*). The bias is more accurately described as resistance to using two distinct words to designate the same *construal* or *message-in-context*.

The idea that distinct words compete to describe the same message-in-context is sometimes confused with a quite different claim; namely, that we have a bias against one label being associated with more than one message-in-context. This latter claim does not appear to be valid, at least not when the messages-in-context are semantically related to one another (recall, again, the multiple distinct senses of the one word form *fire*) (see Floyd and Goldberg, forthcoming).

In the case of word learning, competition between word forms explains why we generally assume that a new word refers to a new distinguishable concept.

Various models have been proposed that take advantage of competition between word choices in both production (Bates and MacWhinney, 1987; Horst and Samuelson, 2008; Rayner and Springer, 1986; Yurovsky et al., 2013) and comprehension (Gaskell and Dumay, 2003; McClelland and Elman, 1986). The idea that words within a given dialect compete for distributional niches has a clear analogy in biology. Two species that share the same ecological niches cannot coexist in a long-term equilibrium; one or the other will gain a slight advantage, and this advantage will snowball over time, ultimately driving the other species either to extinction or to a newly distinctive ecological niche (Darwin, [1859] 1993; Grant and Grant, 2002). Darwin in fact long ago drew the analogy to language, noting that two words cannot remain in a long-term equilibrium if they are both associated with the same meaning. Aronoff and Lindsay (2016) suggest that the winner-takes-all competitive phenomenon extends beyond biology to all complex systems. They appeal to Gause's general law of competitive exclusion (Gause, 1934) as an explanation for the fact that languages strongly disprefer true synonyms. But words are not species and meanings are not food. So it is fair to ask *why* don't words exist that are in free variation, usable in identical contexts?

There is a functional advantage to assigning each label a distinct meaning-in-context, even if the meanings differ only in terms of register (*buy* vs. *purchase*), dialect (*pop* vs. *soda*), connotation (*stingy* vs. *thrifty*), background frame (*land* vs. *ground*), or distribution (*sofa* bed vs. therapist's *couch*). The fact that distinctions exist allows speakers to more quickly access the best match for their intended message-in-context when they speak. If two words were truly interchangeable, speakers would be forced to make a totally random decision each time either word was used. This would violate the efficiency aspect of the CENCE ME approach without contributing to expressiveness, since unbiased decisions take longer to make (Ratcliff et al., 2004) and yet contribute no additional information.

The idea that each word has its own distributional niche has been thought to be based on a pragmatic inference, in that the following reasoning is attributed to the learner: "if the speaker had intended to refer to the 'spoon' she would have used the word *spoon*; therefore, this word must not mean 'spoon'" (Clark, 1987; Diesendruck and Markson, 2001; Goldberg, 1995). While this type of inference may be made by older children and adults, there is no need to posit a high-level inference, or even one that is specific to humans, to account for the effect of competition between word forms. It turns out that well-trained *dogs* have been shown to prefer that a novel word be assigned a novel meaning; in particular, a border collie, Rico, was found to select the one new object from among several familiar objects that already had familiar labels, when asked to bring the *sirikid* (a novel term) (Kaminski et al., 2004). This effect can be attributed to classic interference: once an A→B association has been learned, it is more difficult to learn a C→B association (Ellis, 2006). That is, once a word form (A) becomes associated with a particular meaning (B), it is more difficult to interpret an unrelated word (C) as having the same meaning. The ubiquity of polysemy predicts that learning an A→B association, however, does *not* make it more difficult to learn an A→B′ association, where B and B′ are distinct meanings that are recognizably related to one another.

2.7 Summary

There is much more to say about word meanings and how they are learned and used. The general topic deserves several monograph-length works of its own (see, e.g., Bloom, 2000; Bowerman and Levinson, 2001; Clark, 1995; Fellbaum, 1998; Hart and Risley, 1999; Lakoff, 1987; Murphy, 2002; Pustejovsky, 2012; Tomasello, 2003). However, our primary focus in this book is not on word meaning or word learning, but on the fact that learners are able to avoid using certain phrasal formulations even though those formulations are easily interpretable and are not syntactically ill formed in any obvious way. That is, our quarry is the *explain-me-this* puzzle. The key take-away lessons from the present chapter are the following:

- Words evoke semantically rich, structured meanings, partially abstracted from contexts of use.
- Memory for how words are used is vast.
- We regularly employ old words for new uses, so that common words come to evoke a cluster of conventional, related senses.
- New representations are added, strengthening the overlap in our hyper-dimensional conceptual space with previous experiences of the same word.
- Word meanings are constrained by competition from other words.
- Speakers avoid overgeneralizations by learning and gaining fluency with more appropriate labels for the intended meanings.

In the following chapters, we will see that each of these points applies to more abstract pairings of form and function, including ARGUMENT STRUCTURE CONSTRUCTIONS, introduced in the following chapter. We address the *explain-me-this* puzzle head on in chapters 4–6.

Chapter 3

Constructions as Invitations to Form Categories

In the previous chapter, we discussed the sorts of rich interpretations that individual words can evoke. In this chapter, we focus on the fact that when we interpret an utterance, we also importantly rely on the way the words are put together. For example, we understand the difference between *The boys ate an alligator* and *An alligator ate the boys*, despite the fact that the same words appear in both sentences, because the English TRANSITIVE CONSTRUCTION determines which noun phrase refers to the actor and which, the acted upon.[1] The transitive construction is an example of an ARGUMENT STRUCTURE CONSTRUCTION (ASC) (Goldberg, 1992a). ASCs determine the basic clause types in each language and constrain the interpretation of "who did what to whom."

We will see that ACSs contribute to the semantic interpretation of utterances in important and sometimes subtle ways, although their meanings are more abstract than the meanings of individual words. Relatedly, ASCs also often determine important aspects of INFORMATION STRUCTURE; for example, they often constrain which parts of a sentence convey new information and which aspects have already been evoked in the discourse. Occasionally, ASCs constrain which social contexts they can be used appropriately in, or even how verbs that appear in them should sound. Examples of each of these CONDITIONING FACTORS are described below in order to offer a flavor of the various dimensions that serve to characterize individual ASCs. The formal grammatical patterns of individual constructions encourage learners to seek out regularities in how they are to be used. That is, just as a famous paper was entitled, "Words are invitations to

1 See Dowty (1991), Fillmore (1968), and Foley and Van Valin (1984) for discussion of this generalization. The interpretation of an actor acting on an undergoer is clearly less pertinent to certain transitive sentences, including *This line meets that line* or *She underwent an operation*. This raises the question of whether there exist a number of different transitive constructions with different semantic properties (e.g., Ambridge and Lieven, 2015), or whether such sentences are nonprototypical extensions of a single construction, parallel to how words have prototypical and extended senses, as discussed in chapter 2. In chapter 4, we will see that since "constructions" constitute clusters of partially abstract exemplars, the answer depends on whether constructions are defined narrowly (individual clusters) or broadly (clusters of clusters).

form categories" (Waxman and Markow, 1995; see also Zettersten and Lupyan, 2018), grammatical constructions, like words, invite learners to form categories.

There is variation in how ASCs are used across dialects of the same language, but once the appropriate factors are identified for a given dialect, it is possible to predict with high precision which ASC is most appropriate to express a given message in a given context. This chapter ends with a brief survey of a few of the striking differences in ASCs that exist across languages, to give readers some appreciation of how much we need to learn on the basis of the language we witness (section 3.7).

3.1 Meaning (Semantics)

The DOUBLE-OBJECT CONSTRUCTION is to the study of language what the fruit fly is to biology: it has been the focus of more research than perhaps any other construction. I apologize in advance, dear reader, that the discussion here will not do much to rebalance the scales, since the double-object construction usefully illustrates each of the key aspects of ASCs mentioned at the outset: semantic, information structure, phonological, and dialect constraints. But several other ASCs are introduced as well in this chapter, and still others are discussed in chapter 8.

The double-object construction (3.1) involves, as you might guess from its name, two noun-phrase objects (Objs), as well as a verb (V). It evokes a quite specific meaning, that of transfer between two sentient beings.

Double-object construction

3.1 (Subject$_x$), V, Obj$_y$, Obj2$_z$[2] "X causes Y to receive Z"

To see that the construction contributes to interpretation, try to guess what the novel verb, *moop*, is likely to mean in (3.2):

3.2 She mooped him something. (double-object construction)

Your interpretation of (3.2) cannot rely on prior knowledge of *moop*, since *moop* is a made-up word. It turns out that the majority of people guess that *moop* means "give," and the rest similarly assume *moop* has a meaning that preserves an interpretation of literal or metaphorical TRANSFER (e.g., "tell") (Ahrens,

2 The representation refers to grammatical relations (Subject, Obj[ects]) instead of grammatical categories (NP), because some NPs are predicates, not objects. For example, in *He called them fools*, the word *fools* is an NP but it is not an object, in relational terms; notice that adjectives can equally well appear in this position with little change in meaning: *He called them foolish.*

1995; Ellis and Ferreira-Junior, 2009; Goldberg, 1995; Goldwater and Markman, 2009; Kako, 2006; Rowland and Noble, 2010).

When the double-object construction occurs with familiar verbs, it also requires an interpretation involving transfer (whether the transfer is intended, actual, or denied). For example, notice that the example in (3.3), another instance of the double-object construction, requires that Sam intends to give Chris the cake. This is not necessarily true when *bake* is used in a different construction as in (3.4), which is a transitive construction with an optional BENEFACTIVE *for* phrase:

3.3	Sam baked Chris a cake.	(double-object construction)
3.4	Sam baked a cake for Chris.	(transitive construction + benefactive)

The double-object construction in (3.3) can only mean that Sam bakes the cake with the intention of *giving* it to Chris. Notice that while (3.4) can also be used in this context, it can alternatively be used if Sam is baking the cake for someone else instead of Chris baking it, because Chris is not well. Alternatively, (3.4) can be used if Sam bakes the cake with the intention of throwing it at Chris (in this case, the technical term "benefactive" is a misnomer). The wider range of interpretations for (3.4) suggests that the fact that the sentence in (3.3) implies intended transfer is not a result of the verb *bake* nor any other particular word in (3.3).

Instead, linguists have long recognized that the semantics of "giving," or transfer from one animate being to another, is associated with the English double-object grammatical construction (Green, 1974; Oehrle, 1976; Partee, 1970; see also Goldberg, 1992b; Pinker, 1989; Rowland and Noble, 2010). This has since been further supported by experimental judgment studies (e.g., Ambridge et al., 2012b; Ambridge, Pine, et al., 2014) and corpus analyses (Goldberg et al., 2005). The semantic restriction of transfer to a recipient explains why (3.5) sounds odd:

3.5 ?She sent that place a package. (double-object construction)

An acceptable paraphrase of (3.5) involves a different construction: the *to-dative* or, to use a more general term, the CAUSED-MOTION CONSTRUCTION, as in (3.6):

3.6 She sent a package to that place. (caused-motion construction)

The caused-motion construction is used with a wide range of verbs, including those in (3.7) and (3.8), to express, as the name suggests, caused motion.[3]

3 The constructions in 3.7–3.8 and 3.9–3.10 are commonly referred to as part of a "locative alternation," but the caused-motion and the transitive + quasi-instrumental are treated as independent constructions here, because each can be described in a quite general way. Only

3.7 Pat loaded some hay onto the wagon. (caused-motion construction)
3.8 Pat sprayed the paint onto the wall. (caused-motion construction)

Some verbs can alternatively appear in a transitive construction that allows an optional quasi-instrument phrase (*with* <something>; for discussion of the *with* phrase as a quasi-instrument, see Goldberg [2002]). In this case, a causative interpretation is implied. That is, the combination of transitive + quasi-instrument phrase implies that the direct object complement is affected in some way: note that in (3.9) and (3.10) the wagon or wall are most naturally interpreted as full of hay or covered with paint, respectively, while this is not the case in (3.7) or (3.8) (Anderson, 1971).

3.9 Pat loaded the wagon with hay. (transitive + quasi-instrument)
3.10 Pat sprayed the wall with paint. (transitive + quasi-instrument)

In fact, acceptability in the transitive construction + quasi-instrument correlates with the degree to which the verb is understood to cause a change of state (Ambridge et al., 2012a; Gropen et al., 1989).

3.1.1 Evidence

Evidence for constructional meaning comes from a variety of experimental work. One type of study demonstrates that the meanings of verbs can be *coerced* by the constructions they occur in, and speakers are sensitive to this. For example, Ambridge, Noble, et al. (2014) asked three-year-olds and adults to choose between two possible interpretations of ungrammatical sentences such as ?*Bob laughed Wendy*, by means of an animated picture-choice task. One interpretation ("Bob made Wendy laugh") preserved the meaning of the transitive construction in which the verb appeared (X acts on Y). Another interpretation ("Bob laughed at Wendy") preserved the meaning of the verb—in this case *laugh*—at the expense of the meaning of the construction. Both adults and children selected the construction-congruent interpretation in the vast majority of trials (see also Michaelis, 2003; Naigles et al., 1992, 1993; Sethuraman et al., 1997).

Other studies have found that people make use of ASCs just as much, if not more than, the main verb, when explicitly asked to *sort* sentences into sets that have similar overall sentence meanings (Bencini and Goldberg, 2000; Gries and Wulff, 2005; Liang, 2002). In particular, participants were provided with a set of 16 sentences in which four verbs were used in each of four different constructions. When asked to sort the 16 sentences into four groups according to

a relatively small number of verbs appear in both constructions, and focusing only on those verbs obscures the larger generalizations (see Goldberg [2002] for discussion).

overall sentence meaning, they were as likely to sort all instances of the same construction together as they were to sort those that shared the same verb, even though the verb was an obvious shared feature.

Kaschak and Glenberg (2000) provided additional evidence of constructional meaning through a series of comprehension studies involving novel verbs created from familiar nouns ("denominal" verbs) (Clark and Clark, 1979). In particular, participants were asked to paraphrase sentences such as (3.11a, b), while a separate group was asked to define the denominal verbs involved.

3.11a Allyn crutched Tom her apple to prove her point. (double-object)
3.11b Allyn crutched her apple to prove her point to Tom. (transitive)

Participants were more likely to decide that transfer had occurred in (3.11a) than in (3.11b), and were more likely to decide that the novel verb (*crutch*) was a verb of transfer in (3.11a) than in (3.11b). As Kaschak and Glenberg conclude, "the syntax specifies a general scene," with the novel verb filling in more specific information (p. 508).

Goldwater and Markman (2009) likewise found that instances of the MIDDLE CONSTRUCTION, which does not evoke an agent or causer, tended to be judged nonsensical when followed by purpose clauses, which presuppose an agent or causer. This was not the case for instances of the passive construction, which does evoke an agent or causer.

3.12 ?The ripe tomatoes had sauced well to complement the pasta at the gala dinner. (middle)
3.13 The ripe tomatoes were sauced well to complement the pasta at the gala dinner. (passive)

Taking this line of work one step further, Kako (2006) has found that speakers glean the construction's meaning even when nonsense words appear instead of nouns *and* verbs—for example, *The rom gorped the blickit to the dax* (see also Lebani and Lenci, 2016). These studies collectively argue that ASCs play a role in speakers' ultimate interpretations of sentences.

A skeptic might argue that all of these tasks lend themselves to strategic responding. That is, each of the tasks involved has been explicit, asking participants: "does this make sense?" (Goldwater and Markman, 2009), "what does this mean?" (Kako, 2006), "paraphrase this sentence" (Kaschak and Glenberg, 2000), or "sort according to overall sentence meaning" (Bencini and Goldberg, 2000). Work demonstrating a role for constructional meaning in the acquisition of verbs ("syntactic bootstrapping") in younger children is compelling, but the aspects of meaning that have been demonstrated to date in that work have focused primarily on the number of arguments involved.

In fact, however, newer work has found that constructional meaning appears to be available without conscious reflection, on a time scale that is compatible with online sentence comprehension. In particular, the type of "Jabberwocky" sentences used by Kako (e.g., *She jorped him the miggy*) have been found to implicitly PRIME real verbs that are semantically related to the ASC of the sentence (Johnson and Goldberg, 2013). In particular, participants were asked to perform a LEXICAL-DECISION TASK, which required them to decide as quickly as possible whether individual words (and nonwords) were words. Participants' reaction times to decisions for real verbs were found to be faster when they were preceded by a semantically congruent construction. For instance, the words *give*, *handed*, and *transferred* were judged to be words faster when they appeared after a CONGRUENT prime sentence (e.g., *He daxed her the norp*) than when they appeared after an INCONGRUENT prime sentence (e.g., *She jorped it on the lorp*). The words *put*, *placed*, or *positioned* were responded to more quickly after a prime sentence that was congruent for them (e.g., *She jorped it on the lorp*). Notice that the target words included high-frequency associates of each construction (e.g., *give* for the double-object construction; *put* for the caused-motion construction), low-frequency associates (e.g., *handed*, *placed*, respectively), and semantically related nonassociates that actually rarely occur in the respective constructions (e.g., *transferred*, *positioned*), and faster reaction time (priming) was found for all three types of verbs. Especially notable was that people responded to the word *transfer* faster after a double-object Jabberwocky sentence, even though this particular verb rarely occurs in the double-object construction (see section 3.3). Thus, Jabberwocky-type instances of argument structure patterns prime verbs that are related to their abstract meanings.

Neurolinguistic work on individual constructions is still in its infancy. But in one experiment, participants passively read blocks of instances of the double-object construction and the caused-motion construction, as in (3.14) and (3.15) respectively.

Examples of the double-object construction

3.14a James e-mailed Matt a document.
3.14b Josh took David a notebook.
3.14c Jessica sold Mike a hot dog.

Examples of the caused-motion construction

3.15a James e-mailed a document to Matt.
3.15b Josh took a notebook to David.
3.15c Jessica sold a hot dog to Mike.

As in these examples, the same nouns and verbs were used across the two types of blocks and the corresponding sentences had very similar meanings. A newer type of analysis of fMRI data, multivoxel pattern analysis (MVPA), allows researchers to determine whether patterns of activation in particular areas are capable of distinguishing similar stimuli. Allen et al. (2012) were in fact able to distinguish the double-object and caused-motion constructions, even though propositional content, nouns and verbs, complexity, and frequency were controlled for. Critically involved in the discrimination was an anterior temporal region in the left hemisphere (left anterior Brodmann's area 22) that is known to be evoked by semantic composition, suggesting that the two constructions differ subtly in how their meanings are composed. A related study comparing actives and passives similarly found discrimination was possible in nearby temporal areas of the cortex also known to be involved in semantic combination (Frankland and Greene, 2015).

Collectively, these experimental studies demonstrate that ASCs are associated with meaning even when the main verb involved is not represented as a verb but as a noun (Goldwater and Markman, 2009; Kaschak and Glenberg, 2000), when the main verb is a nonsense word and thus not previously represented at all (Johnson and Goldberg, 2013; Kako, 2006), or when the verb *is* represented, but with a meaning that is at odds with that of the construction (Ambridge, Noble, et al., 2014; Naigles et al., 1993; Sethuraman et al., 1997). Evidence of constructional meaning exists even when the morphological form of the main verbs involved is controlled for, as in the sorting and fMRI studies (Allen et al., 2012; Bencini and Goldberg, 2000). Moreover, the priming study using Jabberwocky sentences and the fMRI studies using MVPA did not rely on explicit judgments or interpretation, but instead involved automatic processing (Allen et al., 2012; Frankland and Greene, 2015; Johnson and Goldberg, 2013). On the basis of this wide range of evidence, we can safely conclude that ASCs are associated with interpretations.

3.1.2 The Construct-i-con

A partial list of ASCs is provided in table 3.1. It is clear from the second column that the forms of ASCs are partly dependent on their interpretation. That is, certain formal complements have subscripts that indicate that they must be interpreted in a specific way (e.g., as a path). The more fully specified semantics is provided in the third column of the table.

We do not expect to find ASCs that make reference to extremely specific events, such as "actions related to a spoon" (although the word *spoon* used as a verb can convey that meaning). This is because ASCs must be relevant over a range of contexts, since they are, by definition, used with multiple verbs. In fact, ASCs capture "humanly relevant scenes of experience" (Goldberg, 1995, 1999), including scenes

TABLE 3.1. Incomplete list of English ASCs

Label	**Form** ***Examples***	**Meaning**
Double-object construction	Subj, V, Obj, Obj2* *She gave him something.* *She mooped him something.*	X causes Y to receive Z
Way construction	Subj$_i$, V, <poss$_i$> *way*, Oblique$_{path}$ *She made her way into the room.* "*Heather handstands her way out of the bathroom.*"	X creates a path and moves through it (to) Z
Intransitive motion construction	Subj, V, Oblique$_{path}$ *She went down the street.* "*Skiers whooshed down the slopes.*"	X moves (to/from) Y
Caused-motion construction	Subj, V, Obj, Oblique$_{path}$ *She put the ball in the box.* "*He sneezed the bullet right out of his right nostril.*"	X causes Y to move (to/from) Z
Resultative construction	Subj, V, Obj, Predicate$_{AP}$ *He made her crazy.* *She kissed him unconscious.*	X causes Y to become Z
Transitive causative construction	Subj, V, Obj, (Oblique$_{instrument}$) *He broke the plate (with a hammer).*	X directly causes Y to undergo a change of state (using Z)
Rely-on construction	Subj, V, Oblique$_{on}$ *She relied on him.* *She lived on vegetables.*	X relies on Y (for physical or psychological nourishment)
Communication conative construction	Subj, V, Oblique$_{at}$ *She yelled at the duck.*	X directs forceful action at Y

The abbreviations in the table above should be interpreted as follows: AP, adjective phrase; Obj, noun phrase object; poss, possessive; Oblique, prepositional phrase argument; Subj, subject; V, verb.

* The commas between complements in ASCs' formal representations are intended to imply that the word order is not fixed in advance. We return to this point in section 3.2.

Each ASC is defined by a set of grammatical relations, semantics, and associated information structure properties (not shown). For present purposes, these constructions may be interpreted as lexical templates (Hovav and Levin, 1998; Müller, 2007; Müller and Wechsler, 2014), or as multiword ASCs (Goldberg, 1995, 2013, 2014).

of moving or changing state, people or forces acting on one another, ongoing states, and so on, with metaphorical extensions allowing these basic scenes to apply to quite abstract topics as well as literal ones (Fillmore, 1968; Goldberg, 1995, 1997; Lakoff, 1987; Pinker, 1987; Talmy, 2003). At the same time, ASCs often have somewhat subtle constraints, as proposed and refined by many researchers (e.g., Bergen and Chang, 2005; Booij, 2002; Croft, 2003; Culicover and Jackendoff, 2005; Goldberg, 1992a, b, 1995, 1999, 2002, 2006; Hwang, 2014; Jackendoff, 1992, 1997, 2002 ; Michaelis, 1996; Stefanowitsch and Gries, 2009).

Constructions form a network of interrelated knowledge within our hyperdimensional conceptual space, just as our knowledge of words does (Booij, 2010; Fellbaum, 2010; Goldberg, 1995, 2016; Goldberg and van der Auwera, 2012; Jackendoff, 1997; Kim and Sells, 2013; Lakoff, 1987; Goldberg and Michaelis, 2017; Sung and Yang, 2016). In this way, our knowledge of language forms a CONSTRUCT-I-CON, which includes words, partially filled words (aka morphemes), *and* representations that are larger than single words, all represented in a complex dynamic network, much as we have long known to be true of the lexicon. For example, the constructions listed in table 3.1 are better captured by a representation that includes the relationships among them, as schematically illustrated in figure 3.1. Bidirectional arrows between constructions are intended to capture the idea that the constructions mutually motivate each other and share aspects of their representation. The resultative construction (e.g., *She drove him crazy*) is a metaphorical extension of the caused-motion construction (e.g., *She drove him to Chicago*), and the *way* construction (e.g., *She drove her way across the country*) is simultaneously related to both the intransitive motion construction and the caused-motion construction (Goldberg, 1995). Relating constructions to other constructions allows us to explain many constraints that would otherwise need to be stipulated, as discussed elsewhere (e.g., Booij, 2010; Goldberg, 2016; Goldberg and Michaelis, 2017; Goldberg and van der Auwera, 2012; Janda, 1990).

Because the construct-i-con is a network, some of the constructions in table 3.1 subsume more specific constructions that have been proposed in the linguistic literature. For example, the double-object construction subsumes expressions that can be paraphrased using *to* and those that can be paraphrased using *for*. See Goldberg (2002) for evidence supporting this higher-level analysis.

Double-object construction

3.16a Zach sent Sue a cake. (cf., Zach sent a cake to Sue.)
3.16b Zach baked Sue a cake. (cf., Zach baked a cake for Sue.)

Similarly, the caused-motion construction subsumes *to*-datives (3.17a, b), and so-called figure-locatives, which tend to use the prepositions *into* or *onto* (3.17c, d):

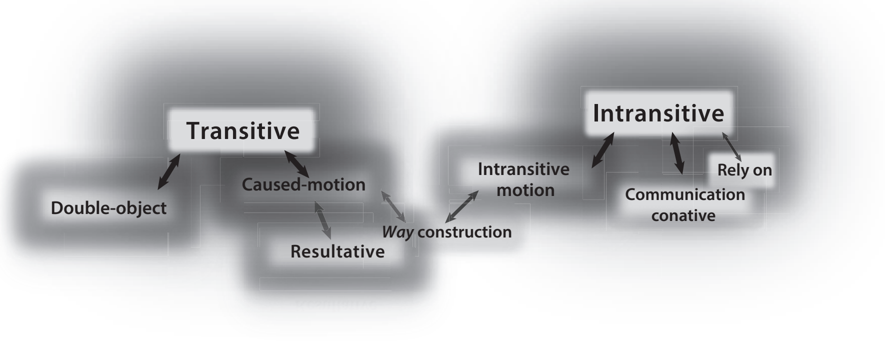


FIGURE 3.1. Argument structure constructions as a network of clusters of abstract, structured, distributed, partially overlapping representations.

Caused-motion construction
3.17a Zach sent a cake to Sue.
3.17b Zach sent a cake to the truck.
3.17c Zach loaded the cake into the truck.
3.17d Zach put the cake onto the truck.

3.1.3 Compatibility

Verbs can sometimes be used in constructions in which they do not typically occur. In these cases, the construction *coerces* an aspect of interpretation that is not evoked by the verb in other contexts (Goldberg, 1995; Michaelis, 2003, 2005). For example, the verbs *sneeze* and *think* invite a construal in which they imply caused motion when they are used in the caused-motion construction as in (3.18) and (3.19):

3.18 "He sneezed the bullet out of his right nostril."
3.19 "His dad basically thought him into his mother's womb."

Similarly, although the verb *kiss* is not normally considered a change-of-state verb—one can kiss a wall without affecting the wall in any way—it is possible to *construe* the act of kissing as having an effect. We therefore find that *kiss* is compatible with the resultative construction, as in (3.20):

3.20 She *kissed* him unconscious.

Another construction, the WAY CONSTRUCTION (e.g., *she made her way into the room*) is associated with the creation of a path through obstacles or despite difficulty (Goldberg, 1995; see also Israel, 1996; Jackendoff, 1990; Perek, unpublished MS). While walking is not particularly difficult and therefore is less than fully acceptable in the *way* construction (3.21a), in the right context, walking can be construed as a challenge, in which case its acceptability improves, as in the attested example in (3.21b):

3.21a ?She *walked her way* across the room.
3.21b "[The disabled bride] *walked, yes walked, her way* down the aisle."

At the same time, general semantic compatibility is not enough. Notice that the sentences in (3.22a–c)—just the sort of odd-sounding examples we set out to address—are interpretable, but are unconventional, and they are judged to sound odd (e.g., Ambridge et al., 2012a; Robenalt and Goldberg, 2015).

3.22a ?She *disappeared* the rabbit.
3.22b ?She *filled* the water into the cup.
3.22c ?He *drank* himself ill.

There is no a priori reason why the verbs and constructions in (3.22) are incompatible. This becomes clear when we see that analogous examples *are* acceptable in certain other languages. For example, while verbs of disappearance or other "internally caused" changes (e.g., *giggle*, *laugh*) are not used causatively in English (Levin, 1993), their translations in Austronesian and Mayan languages are (Dixon, 2000). While *fill* is restricted to the causative construction in English, it can appear in either the caused-motion or the causative construction in German (*Sie füllte Wasser in das Glas*; *Sie füllte das Glas mit Wasser*) (Ambridge and Brandt, 2013).

Other languages, including French, Hindi, and Dyirbal, are much more restrictive than English, in that they require the verb to itself lexically encode the same type of event that is conveyed by the ASC (Dixon, 2000). For example, one would not translate *She kissed him unconscious* directly into a transitive sentence in Dyirbal, regardless of how passionate the kiss might be. Verbs meaning "give" can be used in the general caused-motion construction in English, but not in Russian (Levin, 2008). Unless we wish to claim that the verb meanings are critically different in these different languages, it is not feasible to argue that a verb's meaning directly determines its distribution.

Moreover, as outlined in section 3.7, there are marked cross-linguistic differences in the formal expression of ASCs, indicating that the constructions are emergent generalizations (Croft, 2001; Dixon, 2000; Haspelmath, 2015; Mithun, 1986), just as we saw was the case for words. To summarize, it is true that people generally only say things that are interpretable, but this does not explain why some interpretable expressions are judged to be less acceptable than others.

3.2 Form (Syntax)

The formal properties of ASCs correspond to what is traditionally captured by *valency* (Herbst and Schüller, 2008; Tesnière, 1953): the number and type of complements involved in a clause. ASCs do not necessarily specify word order, as word order can instead be captured by independent constructions—such as the subject-predicate construction and the verb phrase construction—that combine with ASCs (Goldberg, 2014, 2016; Stallings et al., 1998; Wasow, 2002). That is, I assume that the same double-object construction is involved when it is combined with a question or cleft construction, as in (3.23b, c):

3.23a	They gave him *a satchel*.	
3.23b	"*What* did they give him?"$_{\text{COCA}}$	(double object + question construction)
3.23c	"It was *a satchel* they gave him."$_{\text{COCA}}$	(double object + cleft construction)

We have so far been taking for granted grammatical relations such as subject, direct object, secondary object, as well as syntactic categories such as verb and noun phrase. These categories must themselves be learned from the input because there are no syntactic cross-linguistically valid tests that identify subjects, objects, verbs, adjectives, nor any other syntactic category, beyond their prototypical semantic functions (e.g., Boas, 1911; Croft, 2001; Haspelmath, 2010). The way fieldworkers decide which units of a new language should be called "nouns" is by identifying which words refer to concrete objects and then determining which other words share overlapping distributions. The analogous procedure is true for verbs: fieldworkers identify which words to call "verbs" by identifying which words refer to actions; then all words that share similar distributions are considered verbs.

Properties of "nouns" and "verbs" that we might expect to be universal turn out not to be. For instance, Croft (2001) has observed tense-aspect-mood inflection cannot be taken as critical for determining the category of verbs cross-linguistically, because words that translate into English as nouns, adjectives, and adverbs, as well as verbs, are inflected for person, aspect, and mood in Makah, a Native American language, while no words are inflected for these categories in Vietnamese. Therefore, unless of course one is willing to say that all words are verbs in Makah, and no words are verbs in Vietnamese, we cannot use these inflections as criteria for a syntactic category, "verb."

It is even difficult to come up with criteria that hold of all members of a grammatical category *within* a language (Culicover, 1999; Herbst and Schüller, 2008, ch. 3; Langacker, 2008). For example, in English, adjectives (e.g., *ripe*) typically can appear attributively (as in *a ripe banana*) or predicatively (as in *the banana seems ripe*), but certain words are only acceptable when used in one way (*the child seems asleep*; ?*the asleep boy*) or the other (*the ultimate test*; ?*the test seems ultimate*). In short, speakers need to learn how the grammatical categories behave in their language by generalizing how they witness their language being used. At the same time, languages necessarily do contain generalizations because if they did not, it would be impossible to use language creatively (e.g., Steels, 2005).

3.3 Sound Patterns (Phonology)

ASCs can sometimes even be picky about how verbs *sound*. It turns out that the English double-object construction prefers verbs that sound as if they are of Germanic and not Latinate origin (Green, 1974).

3.24a She told me a story.
3.24b ?She explained me a story.

3.25a She gave/bought/showed him a book.
3.25b ?She transferred/purchased/displayed him a book.

While few English speakers are aware of the historical origins of the words they use, Germanic (sounding) verbs are generally short (e.g., *give*, *buy*, *ask*, *tell*), while Latinate (sounding) verbs tend to be longer and often contain recognizable parts (*trans*-, *re*-) or contain stress on the second syllable (*transfer*, *purchase*, *reQUEST*, *exPLAIN*). A sensitivity to the Germanic vs. Latinate distinction has been confirmed in several experimental studies that used nonsense verbs (Ambridge et al., 2012b; Gropen et al., 1989). We can also see the dispreference of Latinate verbs in the double-object construction by considering how English speakers use semantically related pairs such as *tell/explain*, *give/transfer*, *buy/purchase*, and *ask/request*. In the 520-million-word COCA corpus (Davies, 2009), the Latinate verbs strongly prefer prepositional paraphrases, while the Germanic verbs prefer the double-object construction, as illustrated in figure 3.2.[4] But as is often the case with language, there are exceptions to the exceptions. In particular, *guarantee* strongly prefers the double-object construction despite being (and sounding) Latinate.

4 In order to facilitate searches with the unparsed corpus, search strings included pronominal direct objects: e.g., give.[v*] [pp*] [d*]; or give.[v*] [pp*] to. That is, the search returned the verb (*give*) followed by a pronoun (*him/her/them*), followed either by a determiner (*a/the/some*; for the double-object construction) or by a preposition (*to*, *for*, or *of*, depending on the prepositional paraphrase).

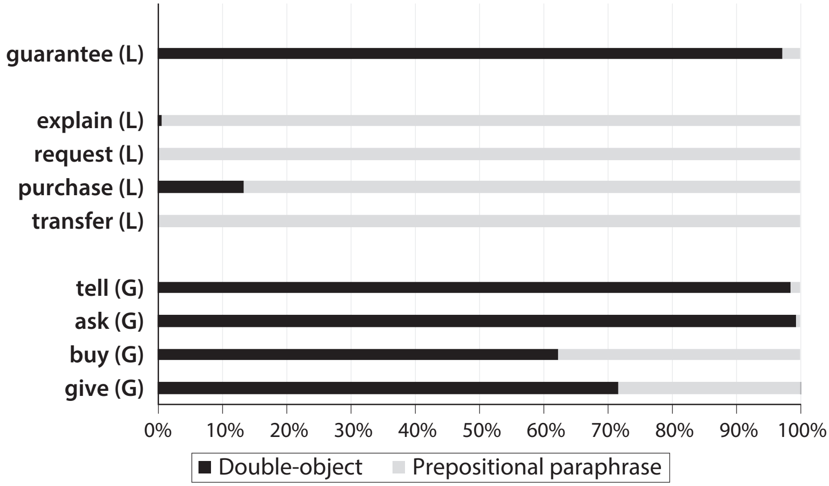


FIGURE 3.2. Percentages of the double-object and prepositional paraphrases for a subset of Latinate (*L*) and Germanic (*G*) (sounding) verbs in the COCA corpus.

This observation is relevant to the *explain-me-this* puzzle, since *explain* is Latinate, sounds Latinate (with its first syllable, *ex-*, and stress on the second syllable), and resists appearing in the English double-object construction, along with most other Latinate-sounding verbs. But, obviously, children are not born knowing this odd, English-specific fact. It has been suggested that Germanic verbs contain an invisible feature ("G") (Pesetsky, 1996), but this begs the question of how children know that this feature exists, which verbs it is assigned to, and what it implies about those verbs' distribution (see also Pinker, 1989, 5.2). Latinate-sounding verbs are not difficult to interpret in the construction, and they are not even all dispreferred (recall *guarantee*). Thus, we still owe an account of how it is that speakers *learn* to avoid using most Latinate-sounding verbs in the double-object construction, while allowing *guarantee* to freely occur in the construction. Therefore, the phonological generalization only helps to clarify the *explain-me-this* puzzle. We have not satisfactorily resolved it yet.

The probabilistic constraint against using Latinate-sounding verbs in the double-object construction seems to have emerged historically for reasons that are unclear. In the eighteenth century, the construction *was* readily used with *allow*, *deliver*, *inform*, *procure*, *repeat*, *provide*, and *produce*, which are all Latinate (sounding) verbs (Bybee, 2015; Colleman and De Clerk, 2011). At that time, the construction was also used with a wider range of interpretations, as illustrated in (3.26–3.28) (attested examples are cited by Colleman and De Clerk [2011], with original citation dates in parentheses):

3.26 "[no one] shall rob me this rich purchase" (1613)
3.27 "the young Benedictine holding him the torch as he wrote" (1767)
3.28 "an English woman clean me my house" (1881)

The fact that the distribution of the double-object construction has narrowed over the past two centuries is evidence that the constraints on the construction today could have been different; they *were* different in the not so distant past. Therefore, again, we still need to explain how native speakers come to know the ways in which the double-object construction is used today.

3.4 Discourse Context (Information Structure)

Constructions typically have constraints on how information is "packaged" in a sentence by taking into account what is understood to be the listener's and speaker's shared knowledge or common ground. That is, constructions determine INFORMATION STRUCTURE: which parts of an utterance are intended as the focus, and which are backgrounded or taken for granted; which arguments must be already given or accessible in the discourse, and which must be new.[5]

Our friend the double-object construction is associated with certain information-structure constraints, in addition to its semantic and phonological constraints. In particular, a number of corpus studies have found that the RECIPIENT argument overwhelmingly has already been introduced in the discourse (typically expressed with a pronoun), while the transferred entity (the THEME) tends to be newer information. For example, a typical instance of the double-object construction is given in (3.29):

3.29 Sally told <u>her</u> <u>a story</u>. (double-object construction)

The double-object construction rarely occurs when the recipient argument is newer or longer than the theme, as in (3.30):

3.30 Sally told <u>a little girl</u> <u>the story</u>. (double-object with less felicitous information structure)

Another way to say this is that the double-object construction strongly prefers its recipient argument to already be TOPICAL in the discourse, while its theme

5 For general work on information structure, see Ariel (1991), Birner and Ward (1998), Chafe (1976), Clark (1996), Clark and Haviland (1977), Du Bois (1987), Erteschik-Shir (2007), Goldberg (2004), Gundel (1985), Halliday (1967), Halliday et al. (2014), Lambrecht (1994), Michaelis and Lambrecht (1996), and Birner and Ward (1998).

argument is more likely to be part of what is in FOCUS (for variations on this formulation, see Arnold, Wasow, et al., 2000; Bresnan et al., 2007; Collins, 1995; Dryer, 1986; Erteschik-Shir, 1979; Givón, 1979, 1984; Goldberg, 1995, 2006; Green, 1974; Hovav and Levin, 2008; Oehrle, 1976; Thompson, 1990, 1995; Wasow, 2002).

3.5 Social Context

Oftentimes, ASCs are applicable across social contexts. Occasionally, as languages change, historically older constructions come to be associated with a more formal register. An example of this is the *IS-TO* CONSTRUCTION analyzed by Goldberg and van der Auwera (2012), as illustrated by (3.31):

3.31 "Arguments are to be avoided; they are always vulgar and often convincing." (Oscar Wilde)

This construction has a range of interpretations, but what is important for present purposes is that it is somewhat restricted when used as an indirect command. That is, it would sound rather rude for a child to use the construction when speaking to a parent, or for a student to use it to address a teacher. The infelicity of the context is indicated by "#" in (3.32) and (3.33):

Adult son or daughter to parent (indirect command):
3.32 #You are to arrive home by 11 p.m.

Adult son or daughter to parent:
3.33 #Cars are to drive on the right side of the street.

The same is not true of other indirect commands, such as the following:

3.34 You are supposed to arrive home by 11 p.m.
3.35 Cars are supposed to drive on the right side of the street.

This restriction on the is-to construction in American English does not appear to hold in British English, where the construction remains more commonly used (Leech, 2003, 229). In fact, constructions are often used in subtly different ways across dialects.

3.6 Variation across Dialects

As already reviewed, in standard American English, the double-object construction strongly prefers the recipient argument to be shorter and more topical than the theme argument. This predicts that the double-object construction should be dispreferred when the theme argument is the third person pronoun, *it*, since *it* is always topical, and this prediction holds (Green, 1974):

3.36 ?She gave him it. (infelicitous in Standard American English)

However, in certain dialects of British English, the same constraint is significantly weaker. There is still some tendency to disprefer *give me it—give it to me* is three times as frequent in the BNC corpus—but the relative difference in bias against using *it* as theme in the double-object construction is greatly reduced. This is clear in table 3.2 on the following page, which compares tokens of *give me it* and *give it to me* in large corpora of American and British English.

Hughes and Trudgill (1996) have observed that the *order* of recipient and theme roles can be reversed in certain dialects of northern British English, as in (3.37a, b) (see also Siewierska and Hollmann, 2007):[6]

3.37a She gave the book him. (Hughes and Trudgill, 1996, 16)
3.37b "Give it me"$_{\text{BNC}}$

Certain nonstandard dialects of American English have taken other sorts of liberties with the double-object construction. Webelhuth and Dannenberg (2006) note that many speakers from the southern part of the United States use the double-object construction as follows:

3.38a "I love me some cheesecake."$_{\text{NOW corpus}}$
3.38b "I'm gonna get me a burger."$_{\text{NOW corpus}}$

This extended use of the double-object construction has been dubbed the "Southern double-object construction," and it highlights the subject argument's involvement in the event or state denoted by the verb, without necessarily implying real or metaphorical transfer (Webelhuth and Dannenberg, 2006). Still other differences in the double-construction in English spoken in India are discussed by Hoffmann and Mukherjee (2007).

The double-object construction is of course not the only construction subject to dialect variation and historical change (see, e.g., Barðdal, et al., 2015; Delbecque et al., 2005; Mondorf, 2014; Östman and Trousdale, 2013; Rickford, 1997; Wulff et al., 2007). Western Pennsylvania English speakers use certain verbs (*need*, *wash*) with a verbal past participle in a way that is not acceptable

6 See also Bresnan and Ford (2010) for an in-depth analysis of the double-object construction in Australian English, which differs from standard American English in still other ways.

TABLE 3.2. Token frequencies of *give me it* and *give it to me* in American and British varieties of English

Corpus	Double-object *Give me it*	To-dative *Give it to me*	% of double-object expressions
COCA American English (520 million)	8	412	**1.9**
BNC (100 million)	26	80	**24.5**

Token frequencies were collected in the COCA corpus (Davies, 2009) and BNC (Davies, 2004). The distribution in these dialects is distinct (Fisher's exact test $p < .0001$).

in Standard American English, as in (3.39) (Tenny, 1998, 592; see also Kaschak, 2006; Murray et al., 1996):

3.39 "it's something that needs addressed."$_{\text{NOW}}$

Appalachian English speakers allow a "split-subject" construction in which the subject position is filled by a nonreferential subject (*there* or *they*) and the logical subject is expressed as a quantified phrase appearing after the verb:

3.40a They didn't nobody live up there. (Zanuttini and Bernstein, 2014)
3.40b There can't nobody ride him. (Montgomery and Hall, 2004)

The fact that these differences emerge in systematic ways within a given dialect and differ across dialects implies that learners' knowledge of constructions must be shaped in dynamic ways by the language that is witnessed. When we look beyond dialects to unrelated languages, we see that ACSs can differ in even more dramatic ways.

3.7 Variation across Languages

All languages have ways to convey scenes in which "someone acts," "something causes something else to move somewhere," "something undergoes a change of state," and "someone experiences something." Languages *must* have means to express these basic scenes of human experience because this is the basic stuff people wish to communicate. Many have assumed that ASCs and their associated semantics are the same in all languages, and that therefore there may not be much to actually learn on the basis of the language that is witnessed (e.g., Baker, 1997; Landau and Gleitman, 1985; Pinker, 1989). While it is true that relationships between form and meaning are not arbitrary (e.g., Bybee et al., 1997; Givón, 1979; Goldberg, 2006; Haiman, 1983; Lakoff, 1987), there nonetheless exist striking cross-linguistic differences in basic clause types and their interpretations across languages. A few of these are surveyed below.

3.7.1 One-Participant Events

If we consider even the most simple type of events, those that involve only one participant, we find that various languages use a vexing array of ASCs. In English, intransitive clauses might be described simply as involving a subject and a verb: *She left*; *It broke*. English does not formally distinguish single intransitive arguments that are agentive (e.g., *The man ran*) from those that are not (*Her stocking ran*). However, many other languages do, and how they do differs. For instance, in Ewe, a Niger–Congo language spoken in Ghana, intentional actions obligatorily appear with a second argument. For example, "run" is expressed *fú du* (literally, "move.limbs course"), "swim" is *fú tsi* (literally, "move.limbs water"), "blow" is *gbɔ ya* (literally, "breathe air") (Ameka, 2006; Essegbey, 1999). Other languages mark intentionality on a single argument, but they differ in whether individual *verbs* determine whether the action is treated as intentional or whether agentive marking depends on whether the action is construed by the *speaker* to be intentional in particular contexts (Aĭkhenvald, 2000; Dixon, 1994, 70–83).

Moreover, what counts as a "subject" in languages is tricky to determine. Subjects cannot be identified by word order: while the majority of the world's languages express an agentive argument before other arguments ("S" before "O"), some languages routinely position the agentive argument clause finally (Tagalog, Malagasy). In fact, there are no formal criteria that can select, across all languages, either "subjects" or "objects" (Croft, 2001; Fried, 1995; LaPolla, 2009).

The grammatical "tests" that linguists use to identify subjects across languages tend to identify either an intentional or active argument (the actor) or an argument that is already given and central to the discourse (the topic), but both the actor and the topic plainly correspond to functional, not syntactic, categories. While some languages require the agentive argument to be topical, other languages systematically distinguish topic and agent arguments (Foley, 2008; Givón, 1979; LaPolla, 1994).

English treats single arguments of intransitive events in the same way it treats the more active argument in two- or three-argument events: both are considered subjects. In English, we also treat single arguments that are neither topical nor active as subjects, but this type of subject is marked in English by a special sentence accent (Lambrecht, 1994). That is, in answer to the question

What happened? an English speaker could say *My* CAR *broke down*, with *my car* being the subject argument even though the car has not already been evoked in the discourse and is not an actor. Lambrecht (1994) points out that French speakers, on the other hand, are more likely to use a special construction that avoids treating the new information ("my car") as subject: *J'ai ma* VOITURE *qui est en* PANNE ("I have my car that is in breakdown") (Lambrecht, 1994).

3.7.2 Two-Participant Events

When there are *two* arguments in a clause, it is useful to be able to distinguish which is the agentive participant (recall, *the alligator ate the boys* vs. *the boys ate the alligator*), and in fact languages overwhelmingly distinguish the two arguments in one way or another. But languages differ in whether the agentive argument in a transitive construction patterns like the single argument in an intransitive construction. In English, it does, and it is sensible to refer to both as subjects (*They* ate soup; *The sticks* broke). In other languages (e.g., Dyirbal), the single intransitive argument patterns with the *undergoer* argument in transitive expressions. The single intransitive argument patterns more like "soup" in "They ate soup" (Dixon, 1977; Du Bois, 1987). In Iroquoian languages such as Oneida, animate agents and animate undergoers are typically indicated by prefixes on the verb, while inanimate arguments are optionally expressed as independent phrases (Koenig and Michelson, 2017). Many other languages routinely express arguments that correspond to direct objects in English as part of the main verb, as English does only occasionally (e.g., *duck-hunting*); these languages vary in the degree to which arguments are "incorporated" into the verb, and within each language individual verbs can vary in their degree of incorporation (Mithun, 1984). In some cases, the bare noun and verb semantically combine to express meanings that are expressed by simple verbs in English. The following examples are from Farsi (spoken in Iran):

3.41a *gush dodan* (ear + give: "to listen")
3.41b *xamiyaze keshidan* (yawn + pull: "to yawn")

Complex predicates vary widely in their specifics but exist in many languages (e.g., Alsina et al., 1997; Evans, 1997; Family, 2006).

3.7.3 Three-Participant Events

English and many other languages allow verbs to occur with three arguments, as in the double-object and caused-motion constructions. While Chinese and English are often both said to involve the "same" double-object construction, the

Chinese construction that has two object noun phrases allows a broader range of meanings than the English construction. In particular, the construction can be used to mean that something is taken from the recipient rather than given to the recipient, as in (3.42):

3.42 *yue1han4 to1 le ma3li4 yi4 ben3 shu1.*
John steal Mary one classifier-book.
"John stole a book from Mary." (Chung and Gordon, 1998)

Double-object ("applicative") constructions in Bantu languages allow an even broader range of meanings: one or the other of the two object arguments can be an instrument, a location, or a benefactive in the event (e.g., Alsina and Mchombo, 1993). And not all languages allow a single verb to have more than two arguments. For example, in Lao, various specialized constructions are available to convey three semantic participants—only a handful of verbs allow three arguments without an additional verb (Enfield, 2002). Many languages allow all semantically recoverable or irrelevant arguments to be omitted, and within this set of languages, there is systematic variation in the extent to which speakers take advantage of this possibility (e.g., Bickel, 2003).

3.7.4 Serial Verb Languages

Languages also vary in how many predicates are used to express an action. While English generally allows only one verb per clause (but see Goldberg, 2006, 50–54), many other languages routinely combine verbs to express a single clause, as in the following example from Akan, a language spoken in Ghana:

3.43 *akwadaa no bɛ-weá a-kɔ dan no mu*
CHILD definite *future*-CRAWL *potential*-GO ROOM definite CONTAINING.
REGION
"The child will crawl into the room." (Ameka and Essegbey, 2013, 24)

In (3.43), the Akan verbs "crawl" and "go" are combined to create the meaning expressed in English by a verb and preposition combination ("crawl into"). This type of serial verb construction is also common in Japanese, Chinese, and Dyirbal (Aĭkhenvald and Dixon, 2006; Seuren, 1990; Slobin, 2004).

While this section has emphasized how constructions that serve the same general functions vary, it is also important to bear in mind that the purpose of each construction and its relationship to other constructions in the language typically *motivates* its form (see, e.g., Givón, 1979; Haspelmath, 2001). To understand this, compare the constructions in a language to useful artifacts such as roofs. Roofs serve to protect their inhabitants and their belongings from the elements. Their specific functions constrain the range of reasonable forms they may take, but they vary in their specific forms (they may be flat, pyramidal, lean-to, gazebo-like).

3.8 Constructions Are Combined (Recursively)

Actual utterances are typically composed of multiple constructions: the OPEN SLOTS of a construction are filled by instances of other constructions in a potentially recursive way. For example, one ASC involves a verb and a clausal complement, as illustrated in (3.44):

3.44 "I thought that we had made an impact." (verb + clausal complement construction)

The clausal complement of *think* can be filled by any clause that describes a thought, and since thoughts can be about other thoughts, the open slot may be filled by a clause which itself contains an open slot for a clausal complement, as in (3.45):

3.45 "She'd thought that he believed that's what he'd done."

The fact that constructions have open slots that may be filled by constructions that themselves contain open slots allows our finite network of constructions to license an open-ended set of potential sentences (von Humboldt, [1832] 1999; see also Everett [2005] and Futrell et al. [2016] for evidence that this property does not necessarily hold in all languages).

As we have seen, slots in constructions are generally restricted in terms of their forms and their functions, and any combination of constructions that fills a slot in another construction must satisfy those constraints (e.g., Ambridge and Goldberg, 2008; Chaves, 2012; Goldberg, 1995, 2014; Goldberg and Perek, forthcoming). That is, the forms and the functions of constructions that are combined must be compatible. When they are not, the resulting utterances are judged to be unacceptable to varying degrees, depending on the degree of incompatibility. Once we appreciate that each construction has its own constraints, the constraints on their combinations should follow without stipulation.

3.9 Summary

We have seen that argument structure constructions (ASCs) impose semantic constraints on which verbs they occur with, and they are sensitive to discourse contexts, social register, and dialect differences. They may also be constrained in fairly arbitrary ways, as exemplified by the relative dispreference of ?*Explain me this* or ?*drive him ill*. The range of distinct ASCs that exists across languages makes clear that children need to learn a great deal about how to express "who did what to whom" in their languages. The following three chapters address how they manage this task, successfully learning to use the ASCs in their languages in creative but constrained ways.

Chapter 4

Creativity

Coverage Is Key

In the previous chapter, we saw that ARGUMENT STRUCTURE CONSTRUCTIONS (ASCs) provide the basic means of expressing who did what to whom. We have seen that the choice of one ASC over another can be conditioned by a variety of factors; constructions can constrain which arguments are already topics in the discourse and which provide new information. They may be constrained for use in certain social contexts. Sometimes, an ASC requires a particular adjective, preposition, or particle, or a range of any of these. We saw that ASCs occasionally even care about the sounds of verbs. We need to learn these specific conditioning factors on the basis of the language we are exposed to, since, as we have seen, they are not universal (recall especially sections 3.6 and 3.7). Therefore, we cannot appeal to the conditioning factors or constraints themselves as an explanation for which verbs can occur with which constructions because that would beg the question of how the constraints are learned. Moreover, the existence of constraints on form, meaning, information structure, and social context does not account for the fact that frequencies, variability, and competition also play important roles, as we will see in this and the following chapters.

This chapter argues that the meaning and distribution of words, combinations of words, and constructions rely on the nature of our memory. In particular, partially abstracted (lossy) structured exemplars dynamically cluster within our hyper-dimensional conceptual space. The clustering gives rise to the formal, semantic, and contextual constraints we reviewed in the previous chapter: new expressions are licensed to the extent that they comfortably fit within an existing cluster. What it means to "comfortably fit" will be defined in terms of a notion of COVERAGE introduced in this chapter. It is argued that the factors we know to be relevant to induction generally—variability, frequency, and similarity—explain the special case of linguistic productivity. Let's unpack how this works.

4.1 Knowledge and Memory

If we consider what we know about any familiar domain, it quickly becomes obvious that we have amassed a vast interrelated network of partially abstracted, structured knowledge. Consider what we know about dwellings, for example. We know that houses have doors, while tents and igloos do not. We also recognize the types of doors that we are likely to find on houses versus skyscrapers versus train cars versus churches versus garages. We appreciate the differences between mansions and dorm rooms, and we also have particular memories of individual homes we have lived in or visited. We know what type of furniture is likely to be found in which types of rooms, and which types of rooms we are most likely to encounter in which types of circumstances (a dining room for a dinner party; a bedroom for a sleepover). We take most of this knowledge for granted and have not been taught it explicitly. Much of it remains implicit: we have a sense of when an automatic door should open as we approach it, we expect the steps in a staircase to be the same height, we are not surprised by the sounds we hear as we move between train cars. We not only know each of these facts but we understand how they relate to each other; we understand how floors, walls, ceilings, and doors are positioned within a dwelling, and we know where light switches, doormats, and rolls of toilet paper are likely to be found.

Decades of research in cognitive psychology have emphasized that **memory is importantly ASSOCIATIVE**, in that individual memories are not modularly encapsulated from one another but instead form interrelated networks (for reviews, see Anderson and Bower, 2014; Hinton and Anderson, 2014). New knowledge is integrated with prior knowledge (Vygotsky, 1978), and each memory trace is therefore inextricably linked to other, related memory traces. Moreover, we do not lay down new memory traces of experiences "blindly." We instead devote more attention to aspects that are informative or unexpected, as we aim to predict what we will experience next in order to make inferences and respond appropriately. We infer causal relationships and prior states, and we recognize subparts and their relationships to larger entities. We learn an integrated network of knowledge, and each new experience serves to update and extend our previous knowledge by strengthening or adding to existing connections and representations, or by weakening prior connections.

Our memory is also CONTENT ADDRESSABLE. This means that when we want to recall something, we hone in on memories whose representations share attributes with whatever we are searching for. If we think about fruit, the most accessible types of fruit immediately come to mind (e.g., apples, oranges, bananas). If we consider pets, we are very likely to access dogs and cats. We do not search through concepts randomly. Neither do we search serially in order of how often we have thought of them before, although more highly entrenched concepts are more easily accessible, other things being equal. **Knowledge of language is knowledge**, and what we know about memory and knowledge applies to language as well.

4.2 Memory for Language

When we speak, our utterances typically include phrases that we have heard and used many times before. Even speakers known for their creativity rely heavily on recognizable LEXICALLY SPECIFIED CONSTRUCTIONS. For example, the comedic genius Jon Stewart was asked by an interviewer which of two political TV shows was funnier, *Crossfire* or *Hardball*. Jon Stewart's response is transcribed below. The compounds and phrases that are recognizably familiar—the lexically specified constructions—are italicized.

> CROSSFIRE or HARDBALL? Which is funnier? Which is more *soul-crushing, do you mean*? Both are equally dispiriting in their . . . *you know*, *the whole idea* that *political discourse* has degenerated into shows that have to be entitled CROSSFIRE and HARDBALL. And *you know*, "I'm *Gonna Beat Your Ass*" or *whatever they're calling them these days* is *mind-boggling*. CROSSFIRE, especially, is completely an apropos name. It's what *innocent bystanders* are caught in when gangs are fighting. And it just *boggles my mind* that that's given *a half hour*, *an hour a day* to . . . *I don't understand how* issues can be dissected from the left and from the right *as though* . . . even *cartoon characters* have *more than left and right*. They have *up and down*.[1]

The use of lexically specified constructions is a hallmark of native-like speech. As Pawley and Syder put it, "The fact is that only a small proportion of the total set of grammatical sentences is nativelike in form—in the sense of being readily acceptable to native informants as ordinary, natural forms of expression, in contrast to expressions that are grammatical but are judged to be 'unidiomatic,' 'odd' or 'foreignisms'" (1983, 193). Notice that the following alternative choices would have made Jon Stewart's response far less native-like (and far less entertaining):

> **Response without lexically specified constructions**:
> CROSSFIRE or HARDBALL? Which is funnier? Which *causes less enthusiasm, do you intend*? Both are equally disheartening in their . . . *you are aware*, *the complete* idea that *talk of politics* has degenerated into shows that have to be entitled CROSSFIRE and HARDBALL. And *you are aware*, "*You will be Defeated*" or *whichever names they are labeling them currently* is *upsetting*. CROSSFIRE, especially, is completely an apropos name. It's what *uninvolved people* are caught in when gangs are fighting. And it just *jiggles my brain* that

1 Jon Stewart, interview by Bill Moyers, *Now*, PBS, 11 July 2003, http://www.pbs.org/now/transcript/transcript_stewart.html. Incidentally, Stewart's critique of *Crossfire* on the show itself was credited with its cancellation.

that's given *0.5–1/24 hours* to . . . *I do not comprehend in what way* issues can be dissected from the left and from the right *in the manner* . . . even *characters in cartoons* have *things in addition to right and left*. They have *down and up*.

Native speakers rely more on lexically specified constructions than non-native speakers do (Dąbrowska, 2014; Ellis, 2008; Lewis, 2008), a point we return to in section 6.4.

Many researchers have assumed that we retain expressions only if they have been witnessed with high frequency. In previous work, I have made this assumption myself, stating "patterns are stored as constructions even if they are fully predictable as long as they occur with *sufficient frequency*" (Goldberg, 2006, 5; emphasis added). But this assumption is nonsensical, since "sufficient frequency" cannot involve some number, *n*, unless speakers retain some memory trace of an instance that has occurred with a frequency of $n - 1$ so that the frequency can be increased upon subsequent exposure; by the same reasoning, a phrase with frequency $n - 1$ cannot be recorded unless a frequency of $n - 2$ was recorded. Applying this reasoning recursively, it becomes clear that speakers need to be able to implicitly record a single exposure (frequency of 1), at least long enough that a subsequent instance might be encountered so that the memory can be strengthened (see also Bybee, 2010). Memory traces that are reinforced become stronger, or more ENTRENCHED (Savage et al., 2003), and sufficient entrenchment may be essential for fluent recall. But if no memory traces were retained upon a single exposure, then each exposure would be the same as the very first exposure, and entrenchment could never increase. No constructions would ever have "sufficient frequency."

The idea that we retain any verbatim language is not entirely uncontroversial. Certain early studies are often misconstrued as demonstrating that we retain *no* memory for the forms of utterances (e.g., Sachs, 1967). But the work had actually emphasized the fact that memory for meaning, or "gist," is *stronger* than memory for form. This shouldn't be terribly surprising, since meaning can essentially be "read off" any sentence that is remembered verbatim. A study by Potter and Lombardi (1990) is regularly cited as demonstrating that we have no memory for linguistic forms. The study had participants read a series of sentences while they also witnessed near synonyms of some of the words contained in the sentences either before or after reading each sentence. The near synonyms interfered with participants' memory, and they commonly intruded in participants' recall and recognition of the sentences. But it is worth noting that even in the face of this interference, participants' accurate verbatim recognition of fairly long sentences (11–15 words) was nontrivial, at roughly 75% correct.

In fact, **memory is cheap**. There is a good deal of evidence that we retain an enormous amount of information about the language(s) we witness, including the ways in which particular words have been previously combined. In a series of experiments, Gurevich et al. (2010) found evidence for above-chance recognition memory for entire clauses after hearing a long story (300 words), even when all nouns, verbs, and adjectives were controlled for across two different versions of the story, and even though participants were not warned that their memory would be tested. In one experiment, participants listened to one of two versions of a story, and then were asked to return for a different experiment a week later. When they were brought back to the lab, participants were simply asked to "retell" the story to the experimenter. Strikingly, they tended to reuse the formulations of whichever version of the story they had witnessed the previous week.

Evidence that we record combinations of particular words comes from studies with adults and children as young as three (Arnon and Clark, 2011; Arnon and Snider, 2010; Bannard and Matthews, 2008; Ellis et al., 2008; Matthews and Bannard, 2010; Reali and Christiansen, 2007; Tremblay et al., 2011; see also Arnon et al., 2017). When asked to repeat frequent versus infrequent strings (e.g., *back in the box* vs. *back in the car*), with the final words matched for frequency (*box* vs. *car*) and the final bigram also matched for frequency (*the box* vs. *the car*), people were faster, and made fewer errors, when repeating the higher frequency strings, even though all of the strings were fully grammatical and semantically sensible.

Dąbrowska (2014) has found that native (British) English speakers are able to correctly select various two-word collocations with above-chance accuracy as being more "natural or familiar" than four other phrases that were chosen to be semantically interpretable. For example, 80% of native British English speakers recognized *absolute silence* as more familiar than *pure silence*, *sheer silence*, *stark silence*, *supreme silence*; more than 90% recognized *arouse suspicions* as more familiar than *incite suspicions*, *kindle suspicions*, *revive suspicions*, *stimulate suspicions.* Again, the more entrenched phrases were recognized, even though all of the combinations were fully grammatical and semantically interpretable.

We do not need to assume that a memory trace is retained for every utterance ever witnessed, forever. And we surely do not retain witnessed word combinations as atomic units, as this would quickly result in a combinatorial explosion (Baayen et al., 2013). But memory traces have internal structure—they are not atomic. Thus, representations of related memories overlap neurally, which mitigates the concern about a combinatorial explosion.

The case for verbatim memory of language should not be overstated. Our brains are prediction machines, so certain information is more likely to be retained than other information. It stands to reason that memory for *meaningful* phrases is privileged, since language users' twin goals are to predict meanings from forms (comprehension) and to select forms to express meaningful messages (production). Retaining connections between words that are high in mutual information is valuable in making predictions. We also know that surprising information or information that was not predicted has a stronger influence on memory, because surprising information creates a stronger error signal, which results in greater weight changes and more learning. Since our goal is to predict forms given messages-in-context and messages-in-context given forms, our network of knowledge of language optimizes these predictions.

There is evidence that initial memory traces are consolidated during sleep, where they are integrated with prior knowledge in the cortex (Rothschild et al., 2017; Stickgold, 2005). In this way, memory traces that overlap with existing memories are strengthened within distributed representations of language. Memory traces that are not consolidated or well integrated with prior knowledge may rely more heavily on the hippocampus (Kim et al., 2014; Marshall and Born, 2007; Schlichting and Preston, 2016).

Speakers do not necessarily retain each recognizable lexically specified construction in a fully specified form. If we consider how the construction *arouse suspicion* is used by native speakers by examining corpus data, we find that *suspicion* behaves as a typical noun: it can appear in the plural (*arouse suspicions*); with a possessive or definite determiner (*arouse their suspicion*; *arouse the suspicions*), modifier (*arouse grave suspicions*), or complement phrase; or some combination thereof (*aroused new suspicions about the smugglers*). The word *arouse* likewise behaves like other verbs: it can be used in any tense, and it can appear in the passive construction (*suspicions were aroused by*) or as a deverbal adjective (*his newly aroused suspicions*). To capture this flexibility, a lossy representation is useful. In particular, we can represent the lexically specified construction schematically as follows: $\text{AROUSE}_{\text{v}}\{\text{X}_{\text{arouser}}, \text{SUSPICION}_{\text{arousee}}\}$. This representation is simply intended to capture the fact that the specific noun, *suspicion*, is the "arousee" of *arouse*. It does not specify the number of the noun or tense of the verb (as indicated by the use of capital letters). The representation allows the construction to be related to other uses of *arouse* that occur with nouns other than *suspicion*, as they, too, involve an explicit or implicit "arouser" argument and an "arousee" argument. It also allows *suspicion* to be related to other lexically specified constructions (*divert suspicion*; *suspicious coincidences*) within our network of knowledge of language. Since our memory is content addressable, when we intend to convey a message that includes the idea of "creating suspicions," this lexically specified construction will become activated, making it accessible and available for use. When it is then witnessed by the listener, it will become more entrenched and even more accessible in the future. We return to this point below (section 4.5), but first we review evidence that we retain the sorts of memory traces that are needed for ASCs: memory traces of which verbs have been used in which constructions.

4.3 Verbs in ASCs

The rich frame-semantic meanings of particular words are part of our knowledge of language (chapter 2), and we know much more about words than the range of meanings they can convey. Verbs in particular can be quite finicky about how they are used, and there is good evidence that we know which predicates have been witnessed in which ASCs. This is clear in part from the fact that nearly synonymous verbs often have distinct distributions, as illustrated by (4.1–4.4):

4.1a "He had helped her get hired"$_{\text{COCA}}$
4.1b ?He had aided her get hired.

4.2a "She provided them with a hidden, safe space"$_{\text{COCA}}$
4.2b ?She gave them with a hidden, safe space.

4.3a "We managed to educate our children"$_{\text{COCA}}$
4.3b ?He succeeded to educate our children.

4.4a "They succeeded in educating children."
4.4b ?They managed in educating children.

Other predicates—including adjectives, prepositions, and particles—can also be particular in their distributions, especially in the ways they combine with other predicates (Boas, 2003; Goldberg and Jackendoff, 2004; Gries, 2003; Herbst, 2014). For example, although *to sleep* and *asleep* can be used in overlapping contexts, when paired with *cry* in the resultative construction, *to sleep* is preferred (Goldberg, 1995, 192):

4.5a "I'll cry myself to sleep"
4.5b ?I'll cry myself asleep.

The dispreference of (4.5b) as compared to (4.5a) is evidenced by a comparison of the number of instances of each in the COCA corpus (Davies, 2009): there are over 100 tokens of "cry <oneself> to sleep" but only one example of "cry <oneself> asleep," even though *to sleep* and *asleep* occur with roughly equivalent frequency in the corpus generally.

Another way to see that predicates are particular about how they can be used is that the meanings of their combinations are often not predictable from the ways in which each word is used in other contexts. This is abundantly clear in the case of complex predicates such as verb + particle combinations. For example, while *take it off* can mean to disrobe, *take it on* does not mean the opposite (cf.

put it on). The intransitive *take up with someone* means roughly to begin to date, while the transitive counterpart *take something up with someone* means roughly to initiate a confrontation. *Take something over to* means roughly to bring, while *to take someone down* means to conquer, and *to take someone out* can imply either a date or a murder (e.g., Goldberg, 2016).

These types of examples require that we retain specific information about how verbs, and predicates generally, are used. There is also considerable evidence that we retain not only the fact that a particular combination of verb and construction has occurred, but also the relative frequency with which it has occurred. For example, *confirm*, like several verbs, is acceptable with either a direct object complement or a clausal complement, as illustrated in (4.6) and (4.7):

4.6	She confirmed [the date of our visit]$_{\text{direct-object}}$	(direct-object complement)
4.7	She confirmed [that the date was soon]$_{\text{clause}}$	(clausal complement)

Searches of large corpora reveal that *confirm* is statistically more likely to occur with a direct object than with a clause (e.g., Garnsey et al., 1997). Several other verbs that share the same direct-object LEXICAL BIAS include *emphasize*, *hear*, *understand*, and *write*, while verbs with the opposite lexical bias (that are clause biased) include *argue*, *believe*, *claim*, *suggest*, and *suspect*.

Other things being equal, when a verb is used in a way that is consistent with its lexical bias, the sentence is processed faster and with greater accuracy (Clifton et al., 1984; Ferreira and Henderson, 1990; Garnsey et al., 1997; MacDonald et al., 1994; Stallings et al., 1998; Trueswell et al., 1993). In fact, speakers even tend to utter the verb itself more quickly when it is used in a context that matches its bias (Gahl and Garnsey, 2004; see also related work by Hay and Bresnan, 2006; Tily et al., 2009).

Why am I emphasizing verbs and other predicates, instead of nouns, in this chapter? It is because, within the domain of argument structure, verbs and other predicates play an outsized role (Boas, 2003, 2014; Hare et al., 2009; Herbst et al., 2004; Müller, 2002; Pinker, 1989). Noun phrases tend to be much less particular about their distributions in ASCs. That is, while verbs commonly display the sort of distributional idiosyncrasies that give rise to the *explain-me-this* puzzle, noun phrases only very rarely do. Why this is the case is discussed in the following section.

It is worth keeping in mind that while predicates are central to the characterization of ASCs in a way that noun phrases are not, which slots of a construction are finicky depends on the function of the construction at issue. For example, particular question words (*what*, *who*, *why*, *how come*) play a central role in the case of question constructions (Rowland and Pine, 2000); in order to learn subject-verb agreement (e.g., *am* vs. *are* vs. *is*; *walk* vs. *walks*), the (number of the) subject argument is crucial. Learners must determine which elements are relevant for each construction on the basis of the functions of each construction, together with distributional evidence. We return to this issue in chapter 6.

4.4 Why Noun Phrases Are Open Slots in ASCs

When children hear that *Sally fepped*, then unless there is something very special about Sally, they know that they, their little brothers, or their friend Harry could all in principle *fep*, regardless of what *fep* means (i.e., it is possible to say *They fepped*. *My little brother fepped*. *Harry fepped*). That is, empirically, noun phrases tend to be substitutable in ASCs, as long as their meanings are semantically appropriate (Imai et al., 2005; Kam, 2009; Naigles et al., 2009; Theakston et al., 2015; Tomasello et al., 1997). This creative freedom follows from two related facts (Theakston et al., 2015).

First, children attend closely to the semantics being conveyed. For example, they are aware from early on that if one human entity can perform some action, others most likely can as well, or at least can pretend to. Similarly, if a child learns the meaning of *nibble* and hears *The bunny nibbled the carrot*, she will know that the bunny can likewise nibble an apple or a pear. If the semantics of the verb requires a very specific type of argument, then that argument will be correspondingly constrained. For example, only trees can be *felled*, matrices *diagonalized*, or shrimp *deveined* (McCawley, 1978). These examples demonstrate that learners are capable of learning quite specific constraints on noun phrase slots; noun phrase complement slots are as variable as the verb's semantics allows.

A second source of support for children learning that noun phrases are generally substitutable for one another comes from distributional evidence. Children witness hundreds of simple phrases in which the noun phrases (NPs) vary but the rest of the phrase is constant (e.g., *This is <NP>*; *<NP> want <NP>*; (Braine et al., 1990; Lieven et al., 1997; Theakston et al., 2015; Tomasello, 2003). This fact ultimately stems from semantics as well. That is, many predicates can be applied very broadly to any animate entity, or to any entity that can be acted upon, or to anything at all that can be construed as an entity (*This is <NP>*; *<NP>'s here*).

Occasionally, ASCs do impose particular lexical choices in a noun phrase slot. For example, consider the *WAY* CONSTRUCTION illustrated in (4.8–4.10):

4.8 "He fought his way out of the house."
4.9 "Sonja was always bullying her way into the center of things"
4.10 "They bike their way down Pennsylvania Avenue to the White House."

This construction necessarily includes a possessive determiner that is coreferential with the subject, followed by the specific noun, *way*. The requirement of extremely specific noun phrases is also a hallmark of *idioms*, which have quite specific meanings and which can require specific determiners or modifiers as well as specific noun phrases (e.g., *kick the bucket* does not mean the same thing as *kick a bucket* or *kick the blue bucket*). But because almost any action can be performed by various actors, and the vast majority of actions can be performed on various types of entities, noun phrase arguments are much less particular than predicates when it comes to ASCs.

4.5 Simple Entrenchment

It is easier to perform all types of motor actions, including speech, when we reuse a plan that we have witnessed or used before for a certain purpose, instead of creating a wholly new plan (MacDonald, 2013; Montag et al., 2017; Schmid, 2017). If the listener intends to formulate a message-in-context, a previously witnessed formulation that conveys the same message-in-context will be more accessible and therefore easier to produce than any novel formulation would be. For example, recall that sentences are read more quickly, and repeated faster and more accurately, when the main verb is used in a way that is consistent with its lexical bias. Mechanistically, when we witness an expression used to convey a particular message, a lossy representation of the expression becomes more strongly associated with its intended message-in-context in the mind of the listener. When we have a familiar message we want to communicate, we make use of linguistic knowledge that is accessible for that purpose in the given context. Of course, we do not simply utter *whichever* words and constructions happen to be the most accessible. Since our memory is associative and content addressable, our knowledge of language relates formal properties, functional properties, and aspects of context, and we access the knowledge that serves our intended purpose. When we speak, we begin by formulating at least part of the message we intend to convey and we access those words and constructions that are appropriate and sufficiently accessible for that message.

Accessibility motivates the fact that speakers tend to *use* familiar formulations over novel formulations that might be intended to express equivalent messages. Additionally, speakers *prefer* familiar formulations, and this has been dubbed the Idiom Principle (Erman and Warren, 2000; Sinclair, 1991; Wray, 2002). Our judgments are not directly explained by the fact that familiar formulations are easier to access, though, since the novel formulations are typically provided when we are asked for judgments, and so demands on memory are minimal. We tell ourselves that we value creativity and nonconformity, and so why should we *prefer* familiar formulations over creative ones?

The answer is that we consider creative uses "wrong" when there exists a conventional alternative way to express the same message, because we view language normatively: we consider there to be "right" ways to use our language. In the type of cases this book aims to address, the "right" ways to use language are familiar ways, whenever familiar ways are suitable to express our intended message-in-context. That is, we prefer the types of formulations we've witnessed in our speech communities (or in those communities we wish to belong to). This idea is expressed by the first CENCE ME principle: speakers balance the need to be expressive and efficient while obeying the normative conventions of their speech communities.

It is somewhat ironic that certain judgments of "syntactic ill-formedness," then, are at root sociolinguistic, given the outsized role such judgments have played in theories of syntax ever since Chomsky (1957). But this idea predicts that subpopulations that want to distinguish themselves from the larger community will be the ones most likely to innovate and adopt new language conventions. In fact, sociolinguists have long recognized that it is adolescents—those who are at precisely the age when they aim to distinguish themselves from the previous generation—who are the primary drivers of language change (Bybee and Slobin, 1982; Wagner and Tagliamonte, 2016; Labov, 2011; Tagliamonte and D'Arcy, 2009; Trudgill, 2011).

We have summarized evidence that speakers retain a great deal of specific information about how individual verbs and other words are used. At least part of the solution to the *explain-me-this* puzzle stems from the fact that familiar formulations tend to be reused due to accessibility, and tend to be preferred over novel formulations because language is a shared cultural system. But when we wish to convey messages that are novel to varying degrees, we must generalize beyond the resources in memory: we need to use language creatively.

4.6 Creativity and Productivity

English speakers can creatively use familiar verbs in ASCs that are novel for those verbs, as in examples (4.11a–i):

4.11a	"She smiled him in the door"	(caused-motion)
4.11b	"she can love the anger away"	(caused-motion)
4.11c	"I can't drink away their visits."	(caused-motion)
4.11d	"I sloshed down the street"	(intransitive motion)
4.11e	"the B-17 screamed into the room"	(intransitive motion)
4.11f	"somebody leaked me a focus group tape"	(double-object)
4.11g	"would prompt the editors to land me a date fit for a princess."	(double-object)

4.11h "They would swat them dead immediately after being bitten." (resultative)
4.11i "We translated our way through a series of bizarre sentences" (*way* construction)

The fact that many constructions are used PRODUCTIVELY is also clear from the fact that we have intuitions about how novel verbs can be used. Arnold Zwicky long ago provided the example of a made-up verb of communication, *greem*, that was assigned the interpretation "to speak with a loud, hoarse quality." As he observed:

> You know that . . . it will be possible to greem for someone to get you a glass of water, to greem to your sister about the price of doughnuts, to greem "Ech" at your enemies, to have your greem frighten the baby, to greem to me that my examples are absurd, and to give a greem when you see the explanation . . . [these constructions] are all systematically associated with the semantic representation of manner-of-speaking verbs. (Zwicky, 1971, 232)

Other cases of novel verbs are created on the fly from familiar nouns as new verbs (Clark and Clark, 1979; Goldberg, 1997; Kaschak and Glenberg, 2000). Attested examples of this phenomenon are provided in (4.12a–e):

4.12a "I'm gonna go benadryl myself to sleep."
4.12b "Blog him out of jail!!" (cited by Hwang, 2014)
4.12c "I cannot believe she utubed the lyrics!"
4.12d "[he had a lot of work to do if] he was to podium."
4.12e "She necklaced me with her arms"

These productive uses of constructions contrast with other cases that indicate that constructions are *not* fully productive (?*Explain me this*). That is, ASCs are typically *partially* productive, and this partial productivity highlights the *explain-me-this* problem we set out to address. Before focusing on how ASCs are constrained, the following sections explore in more detail when, how, and why ASCs come to be used productively.

4.7 Coverage: Clustering of Partially Abstract Exemplars

A simple way to think about productivity is via the notion of COVERAGE, an idea borrowed from the nonlinguistic literature on induction (Goldberg, 2006, 98; Osherson et al., 1990; Suttle and Goldberg, 2011). Specifically, a potential productive use of an existing construction (a COINAGE) is acceptable to the

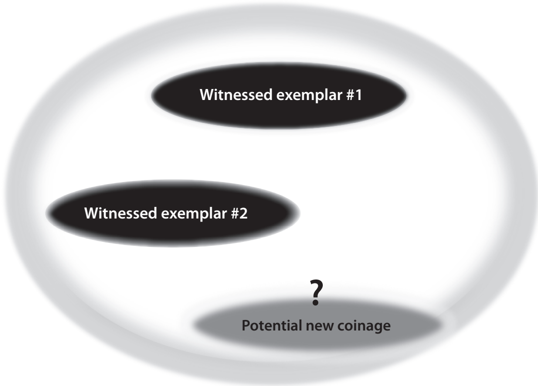


FIGURE 4.1. The smallest ad hoc category in a structured, high dimensional conceptual space that includes two previously attested instances (in dark gray) and a potential new coinage (in light gray). The extent to which the dark gray instances cover the category correlates with how acceptable the coinage is judged to be.

degree that the category which would be required to include the previously attested examples *and* the coinage is well attested within the hyper-dimensional conceptual space in which exemplars cluster. In figure 4.1, the potential coinage is represented by the lighter grey oval, and two previously witnessed exemplars that are clustered close to the potential coinage are represented by black ovals. Coverage predicts that the potential coinage will be judged acceptable to the extent that the ad hoc category required to contain all three exemplars (the large transparent circle) is well attested or well "covered"; i.e., acceptability will vary inversely with the amount of empty space required by the ad hoc category, as described below. The exemplars and the category itself are represented in a hyper-dimensional space that includes semantic, syntactic, information-structure, and phonological dimensions and/or social context. Figure 4.1 is intended to be a projection of this structure onto two dimensions, such that closeness in space represents similarity.

The degree of coverage corresponds to the degree to which the attested instances fill or "cover" the category. Coverage relates the number of different verbs witnessed in a given construction (a construction's TYPE FREQUENCY), the semantic and phonological VARIABILITY of witnessed types, and the SIMILARITY of the coinage to attested types. Each of these factors—type frequency, variability, and similarity—has been independently found to be relevant, as described below. Any new coinage will be judged to be acceptable to the extent that it falls within a well-attested cluster (as long as there is no competing more accessible alternative; see chapter 5).

The formal aspect of constructions plays a special role in how clusters are formed, in that recurrent formal patterns can serve as "invitations" to form categories before semantic and discourse properties of constructions are well understood. That is, early, if imperfect, formal regularities can be gleaned via distributional analysis. The formal surface regularities of constructions invite learners to seek other types of regularities across exemplars, through a process of STRUCTURAL ALIGNMENT (Gentner, 1983; Kotovsky and Gentner, 1996; Markman and Gentner, 1993; Norman and Rumelhart, 1975). Structural alignment involves relating two or more distinct relational structures, and the process highlights both similarities and differences between the representations. Humans are especially talented at structural alignment when compared with other primates such as baboons (Fagot and Thompson, 2011; but see Flemming et al., 2013). Dedre Gentner has argued extensively that aligning relational structures allows for scientific insights, metaphorical extensions, and, most relevantly here, higher-level linguistic generalizations (Gentner, 1983, 2003; Markman and Gentner, 2001; see also Goldwater, 2017; Tomasello, 2003).

Formal parallels between utterances highlight additional similarities and dissimilarities. For example, by aligning the abstract relational structure of two utterances such as *I love you* and *You want a cookie*, the shared relational structure, "animate entity experiences attitude toward something," becomes more salient, as do the differences (e.g., a pronoun object vs. a lexical noun phrase object). The process of aligning exemplars relies on both formal properties and the meaning of the exemplars. For example, Markman and Gentner (1993) showed people a picture of a truck towing a car and one of a car towing a boat. The cars were clearly highly similar, but they played different roles in the overall event depicted in the two pictures: The car was acted on in the first towing picture, while the car was the actor in the second picture. When people were simply asked to indicate which entity in the second picture "matched" the car in the first, they tended to choose the other car. But when a separate group was asked to first compare the pictures and rate their overall similarity and *then* perform the same task, people tended to choose the *boat* in the second picture, which played the same role in the overall scene of towing as the car did in the first picture (see also Gentner and Markman, 1997; Goldwater, 2017; Tomasello, 2003). The USAGE-BASED CONSTRUCTIONIST approach capitalizes on the fact that learners attend to and retain aspects of both the form and interpretation of utterances, as this leads to exemplars clustering in the hyper-dimensional space we use to represent language so that more general constructions emerge (Falkenhainer et al., 1989; Gentner and Forbus, 2011). The semantic, formal, sound, and social dimensions associated with each construction are formed by generalizations across the partially abstracted exemplars that have been witnessed.

Similarities among attested instances shape the domain of a generalization. That is, constraints EMERGE as exemplars that share the same surface form cluster together. For example, exemplars of the English double-object construction (e.g., *give him some money*) will almost all share an implication of potential transfer, and the construction rarely occurs with Latinate-sounding verbs (recall section 3.3). This leads new coinages of the construction, if they are to be judged acceptable, to likewise imply potential transfer and involve shorter, more Germanic-sounding verbs (Ambridge, Pine, et al., 2014; Gropen et al., 1989).

Differences among attested instances of the same formal pattern increase a pattern's variability and encourage productivity. That is, variability begets more variability as attested exemplars cover a larger swath of the representational space. A construction that has been witnessed occurring with 1000 different verbs (type frequency = 1000) is almost certainly going to contain more variability among the verbs than a construction that has only been witnessed with 10 verbs (type frequency = 10). For this reason, variability among attested exemplars is sometimes conflated with type frequency. But, to the extent that type frequency can be teased apart from variability, variability appears to be more predictive than type frequency in judgments of acceptability (Adelman et al., 2006; Gries, 2003; Johns et al., 2016; Kachergis et al., 2017; McDonald and Shillcock, 2001; Suttle and Goldberg, 2011).

At this point, I ask the reader to mentally bookmark the idea of COMPETITION, which is the focus of chapter 5, because competition among constructions is at least as important to productivity as coverage. But, before we turn to chapter 5, the rest of this chapter is devoted to evidence and formalization of the notion of coverage.

4.7.1 Evidence

In a series of experiments, Suttle and Goldberg (2011) found that variability of attested instances, type frequency, and similarity of a coinage to attested instances interact in ways that are predicted by the notion of coverage. Specifically, participants were provided with attested utterances of a fictitious language, "Zargotian," and then asked to judge how likely it was that a final utterance would also be acceptable in Zargotian. An example is given in (4.13):

4.13 Assume you can say these sentences:
Scrape-*nu* the *vip* the *hap.*
Load-*nu* the *yib* the *vork.*
Flip-*nu* the *loof* the *rolm.*

How likely is it, on a scale of 1–100, that you can also say:

Rumple-*nu* the *pheb* the *jirm.*

The stimuli systematically varied in type frequency (participants were given one, three, or six distinct exemplars), variability (exemplars were chosen from one or three different semantic classes), and degree of semantic similarity between the

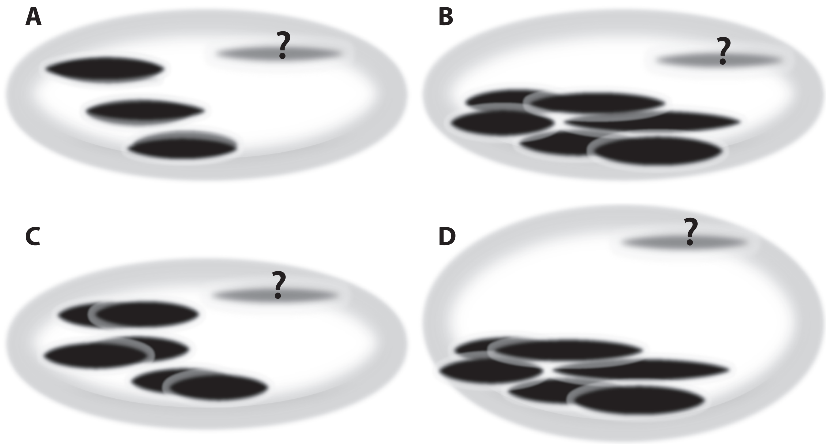


FIGURE 4.2. Sample stimuli involving varying degrees of coverage due to differences in type frequency (*A* vs. *C*), variability (*B* vs. *C*), and semantic similarity (*B* vs. *D*). Based on data from Suttle and Goldberg (2011), represented pictorially.

target utterance and its closest attested neighbor (determined by latent semantic analysis [Landauer, 2006]). Ten verb classes were varied across participants and items: verbs of breaking, loading, bending, cooking, cutting, acquiring, throwing, hitting, holding, and cognition.

The findings confirmed that when coverage was relatively high, a coinage was judged to be more acceptable. For example, the role of type frequency was investigated by comparing judgments on the situations depicted in figure 4.2A and C, which differ in terms of the number of attested instances. As predicted by coverage, participants judged the potential coinage to be more acceptable when the type frequency was increased, as depicted in figure 4.2C.[2]

A comparison of figure 4.2.C and D illustrates the role of similarity among witnessed exemplars when speakers judge the acceptability of a new coinage. Speakers judge new coinages like that in D as less acceptable than the one in C when the only difference is the degree of similarity between the coinage and previously witnessed exemplars (Suttle and Goldberg, 2011, experiment 3; see also Barðdal, 2008; Bybee and Eddington, 2006; Croft and Cruse, 2004; Desagulier, 2015; Kalyan, 2012; Langacker, 1987; Peng, 2016; Zeschel, 2012).

2 Multiple verbs from the same semantic classes used in the condition represented by 4.2A were used in condition 4.2C in an effort to increase type frequency without increasing variability. Yet, as noted above, type frequency and variability are hard to keep entirely independent, so increasing type frequency also increases variability to some extent.

We can see that type frequency interacts with variability, just as coverage predicts. In particular, an increase in type frequency of attested tokens *inhibits* generalization if variability is low and the potential coinage is not part of the same cluster of related tokens, as depicted in figure 4.2B compared with 4.2A (see also Bowerman and Choi, 2001; Bybee, 1995; Janda, 1990). In particular, witnessing more exemplars that form a tight cluster in conceptual space leads to reduced generalization beyond the dimensions of that cluster. This type of relationship between type frequency and variability has been recognized in work on nonlinguistic categorization (e.g., Rhodes et al., 2010). It has also been widely reported or hypothesized as an important factor in linguistic generalization (e.g., Barðdal, 2008; Bybee, 1985, 1995; Clausner and Croft, 1997; Emberson et al., forthcoming; Goldberg, 1995; McDonough and Kim, 2009; Plunkett and Marchman, 1991, 1993; Tomasello, 2003; Xu and Tenenbaum, 2007). High frequency with low variability leads to the formation of a narrow category that is less easily extended to new cases that would fall outside the range of previously witnessed variability.

Further support for the general notion of coverage comes from Perek (2016), who investigated the nature of productivity over historical time by examining the "<verb> *the hell out of* <noun phrase>" construction exemplified in (4.14), that serves to intensify the meaning of the verb:

4.14 "Santas that would *scare the hell out of Jesus*."

The study found the first instance of the construction in COHA to be from 1928 (Davies, 2010):

4.15 "Swap generals with us and we'll *lick the hell out of you*"
(*lick* as used here meant "beat," or "defeat" [Perek, 2016, n. 9]).

Perek examined the semantic distribution of verbs used in the construction in each of four 20-year time periods between 1930 and 2009, using distributional semantics and multidimensional scaling of attested verbs. A hierarchical clustering algorithm identified clusters of varying densities, including verbs of understanding (*understand*, *explain*, *analyze*), verbs of abstract annoyance (*annoy*, *irritate*, *embarrass*, *intimidate*), and the most dense class: verbs of physical injury (*slap*, *pound*, *kick*, *push*, *smash*, *slam*, *bang*, *knock*, etc.). Results demonstrated that the degree of density of a given subclass of verbs during one period strongly correlated with how many new verbs were added to that verb cluster in the following two-decade time period. In other words, clusters with higher density tended to attract new members, just as the notion of coverage predicts.

Another study, by Schuler et al. (2016), found that seven-year-olds and adults tended to generalize a plural form when five nouns were witnessed with that

FIGURE 4.3. Two hypothetical generalizations, with examples that follow the generalization indicated by dark ovals, and exceptions indicated by white cloud-like gaps. The cluster above has greater coverage and should therefore be more available for generalization than the one below.

plural form and four nouns were witnessed with other “irregular” plural forms, but participants did not generalize the plural if they witnessed only three nouns using it and six nouns with a variety of other exceptional plurals. The relative difference in generalizabilty is predicted by the notion of coverage as well—as illustrated in figure 4.3, since the regular plural generalization has greater coverage on the top than on the bottom.

4.7.2 Token Frequencies

While type frequency refers to the number of distinct types that have been witnessed, TOKEN FREQUENCY refers to the number of instances, or “tokens,” that have been witnessed. In the specific case of ASCs, type frequency refers to the number of distinct verbs that have been witnessed in a given construction, and token frequency refers to the number of times a given verb occurs in a construction, with possibly different noun phrases. Since we do not encode every aspect of every exemplar we encounter, it is possible that multiple instances of the same verb in a given construction are represented in an undifferentiated way. According to this perspective, a verb with higher token frequency may serve to simply strengthen the representation of that verb without increasing coverage (see figure 4.4, in which darker shading is used to indicate higher token frequency). This view would not predict that increased token frequency should lead to increased productivity, because additional tokens do not increase variability or coverage. Supporting this view is the fact that several researchers have observed that high token frequency varies *inversely* with productivity (Baayen, 2007; Bybee, 1985, 1995; O’Donnell, 2015). This is expected if increasing the token frequency of a single verb + ASC does not increase coverage.

At the same time, is possible that distinct tokens of a verb in an ASC may create a narrowly circumscribed cluster or cloud of closely related, partially overlapping exemplars for that verb + ASC, as has been proposed in the case of individual words and phonological patterns (Elman, 2004; Johnson, 2006; Kleinschmidt and Jaeger, 2011; Pierrehumbert, 2001). On this view, each token would cover an overlapping but slightly distinct area of conceptual space (figure 4.4C). Increasing the token frequency of a construction should not increase coverage of the larger generalization across verbs but it would allow more variability for new coinages that involve that same verb + ASC. That is, a usage event that

FIGURE 4.4. Three categories with equal type frequency (e.g., number of distinct verbs). An increase in token frequency (of each verb) can be represented either as in *B*, wherein distinct tokens of the same type are represented identically (higher token frequency only indicated by stronger [darker] representations), or as in *C*, wherein individual, albeit lossy, representations are recorded, creating a cloud of closely related tokens for each type.

includes a particular verb, v^i, would overlap more with another usage event of the same verb, v^i_2, than it does with usage events with a different verb, v^w, but v^i and v^i_2 would not be identically represented as long as something about their linguistic or nonlinguistic contexts distinguishes them. There is circumstantial evidence in favor of this idea that individual tokens increase the breadth of a narrow cluster in conceptual space (as represented in figure 4.4C). In particular, verbs that occur with particularly high token frequency in a given ASC often allow more variability in what is acceptable for those verbs. For example, we saw that the double-object construction typically involves transfer between one sentient being and another. However, *give*, the most frequent verb in the double-object construction, can be used with *in*animate "recipient" arguments (4.16) in a way that other verbs cannot (4.17):

4.16a "he **gave** *that place* a town hall and public library"
4.16b "Williams **gave** *the area* its nickname in 1876"
4.16c "He then **gave** *the land* a deep ploughing"

4.17a ?He **granted** *that place* a town hall and public library.
4.17b ?Williams **told** *the area* its nickname in 1876.
4.17c ?He then **offered** *the land* a deep ploughing.

In a parallel way, *make*, the most frequent verb in the resultative construction, is more productive in that construction than other verbs. That is, *make* allows a wide-open cornucopia of adjectival complements, as in (4.18):

4.18 The soup made him sick/ill/crazy/happy/funny/tall/healthy/greenish.

Other verbs, on the other hand, can be quite finicky in the resultative construction: one can eat oneself *sick* but not *ill*; we can drive someone *crazy* but not *sick* (e.g., Boas, 2003; Goldberg, 1995). Thus, verbs that occur with high token frequency in a given construction are often more permissive than other verbs in the same construction. Why should this be? It seems that when a single verb occurs with high frequency, that verb's variability tends to increase as well (since, again, frequency and variability are highly correlated). Therefore the coverage of that verb + ASC tends to be broader than that of other verbs, encouraging more additional variation for new cases of the same verb. That is, the facts related to *give* in the double-object construction and *make* in the resultative construction suggest that individual tokens serve to create their own subclusters; verbs with higher token frequency cover more area in conceptual space than do lower-frequency verbs in the same construction.[3]

Coverage provides a unified account of the nature of generalization, because it explains how constraints that differ across different languages can emerge from the language speakers' witness. Coverage also predicts that productivity should be a matter of degree and should depend on variability, similarity, and frequency. If the variability of attested cases sufficiently covers the conceptual space that is required to include a possible productive use, that productive use is judged to be relatively acceptable. We will see that what *constrains* productivity is the existence of an alternative formulation that conveys the intended meaning and is more accessible at the moment of speaking. We return to this point in the following chapter where we focus on the role of competition via statistical preemption.

4.8 Modeling Coverage

Coverage is captured by computational models that cluster new instances to previously witnessed instances within our hyper-dimensional conceptual space. Clustering algorithms in Bayesian models do this by assigning each new utterance (each USAGE EVENT) to an existing cluster that maximizes the fit between the new usage and the cluster, while taking into account the prior probability of each cluster.

Barak et al. (2014) have proposed such a model in which each usage event is represented by a vector of feature values (F_i), that includes a representation of the verb's semantics, the utterance's semantics, and the argument structure's syntactic properties (for related work see also Alishahi and Stevenson, 2008; Arbib, 2017; Beekhuizen et al., 2014; Bryant, 2008; Parisien and Stevenson, 2010; Perfors et al., 2010). The model learns incrementally just as human learners

3 There is more to say about the role of token frequency and its interaction with type frequency, but as this topic has been covered in depth elsewhere, we leave it aside (see, e.g., Boyd and Goldberg, 2009; Casenhiser and Goldberg, 2005; Ellis and Ferreira-Junior, 2009; Goldberg, 2006, ch. 4–6; Goldberg et al., 2004; Madlener, 2016; McDonough and Nekrasova-Becker, 2014; Wonnacott et al., 2012, 2017).

do. It assigns the very first usage event its own cluster; the next usage event is then assigned either to the existing cluster, if it is sufficiently similar to the previous usage event, or to a new cluster (the degree of dissimilarity that is tolerated is set by a parameter).

Following Bayes's Rule (4.19), the probability that a cluster, k, is the appropriate cluster for a usage event, F_i, is proportional to the relative frequency or probability of k ($P(k)$) multiplied by the likelihood that F_i fits cluster k, which is determined by the degree of match between F_i and the combined weighted features of the previous usage events represented by k.

4.19 **Bayes's Rule** $P(k|F_i) = \dfrac{P(k) * P(F_i|k)}{P(F_i)}.$

The model considers all existing clusters and a new one, and it assigns F_i to the best match—that is, the cluster with the highest probability, given the features of F_i:

4.20 **BestCluster(F_i)** = argmax $P(k \mid F_i)$, $k \in$ clusters.

Argmax takes the probabilities of F_{i} matching with each existing cluster and a new cluster; it returns the cluster with the highest probability. Early on, the model is more likely to posit new clusters than it is later on, as the frequency of existing clusters ($P(k)$) is relatively low initially, so the match between the usage event and the existing clusters, $P(F_i \mid k)$, has less influence.

The model was designed to capture the learning process. But as each cluster contains a combination of formal and functional features, the model can be used to make relevant predictions about production and comprehension as well. In particular, to capture the production of language, vectors with semantic but not syntactic information can be used as input to predict which syntactic construction is best suited to express the semantics provided. To capture comprehension, vectors that contain only information about the form of the verb and the syntactic properties of the utterance can be used.

The model has several advantages. First, it learns incrementally, just as human language learners do. That is, each new usage event is assigned to a cluster as it is witnessed (see Christiansen and Chater [2016] on the importance of online incremental learning). Another advantage is that the total number of clusters that are ultimately posited is not predetermined—the number depends on how similar usage events are to one another. If a new usage event is an insufficient match to any existing cluster, a new cluster is created. Finally, language learning, production, and comprehension all make use of the same representations.

At the same time, the model has a number of limitations that are shared with most other current modeling work. The linear vectors that serve as input do not capture the sort of nuanced structured meanings that words convey (recall section 2.2), let alone the sorts of complicated propositional content that utterances convey. Moreover, the model requires that the input vectors specify some range of features, whereas human learners have to figure out which semantic and formal dimensions are relevant, since these differ in different constructions, and in analogous constructions in different languages (see sections 3.6–7 and 6.1–2). Additionally, the current model has not yet attempted to capture multiclausal utterances, speech acts, or information-structure properties of utterances. In short, this and other current incremental-learning models require much more work before they approach humanlike knowledge of language.

My goal in referencing Barak's model is a modest one. The formalism simply provides a way of talking about what is involved in the type of clustering described here. It also provides a means to investigate the roles of similarity, type and token frequency, and variability (e.g., Barak and Goldberg, 2017). Finally, explicit models of argument structure and verb partial productivity can be usefully compared (e.g., Ambridge and Blything, 2016; Barak et al., 2016).[4]

Constructions correspond to clusters in the model, and they emerge from overlapping representations in a hyper-dimensional conceptual space (see also Daelemans and van den Bosch, 2005; Wible and Tsao, 2017). Increased type frequency (the number of distinct verbs) witnessed in a given cluster tends to lead to greater variability (recall figure 4.2C vs. A), and increased variability in a cluster increases the likelihood that a new usage event will match that cluster (figure 4.2C vs. B). When witnessed verbs in a given cluster share relevant overlapping features, those overlapping features carry more weight. This results in the cluster increasing its attraction to other verbs with those features, while the cluster's appeal to other verbs that have distinct feature values will decrease (figure 4.3). By searching for the cluster with the best match for the particular message that is to be conveyed and the particular verb, the model also captures the role of competition between constructions, which is the focus of the following chapter.

4.9 Summary

Constructions are commonly productive within a circumscribed semantic, pragmatic, syntactic, morphological, and/or phonological space. The notion

4 This work systematically compares Barak's Bayesian model to an earlier feed-forward connectionist model that also aimed to learn the distribution of verbs in the double-object and caused-motion ("*to*-dative") constructions (Ambridge and Blything, 2016). Both models included representations of verb meanings and were trained on the same type and amount of data. Barak et al. (2016) demonstrates that the clustering algorithm of the Bayesian model correlates better with human judgments on several measures. The advantage is likely due to the ability of Barak's model to allow for subclusters, which the minimal connectionist model did not do.

of coverage detailed in this chapter captures the idea that new uses of verbs must fit, or be able to accommodate, the constraints of the constructions they appear in (Ambridge, 2013; Ambridge et al., 2009, 2012a, b; Blything et al., 2014; Coppock, 2009; Goldberg, 1995, 1999, 2010, 2013, 2016; Gropen et al., 1989, 1991; Pinker, 1989). Since instances of each construction cluster together, generalizations about semantic, information structure, syntactic, morphological, and phonological constraints emerge, and new expressions that are witnessed are associated with existing clusters in hyper-dimensional conceptual space. At the same time, coverage accounts for the fact that the variety of previously attested exemplars correlates positively with the acceptability of new coinages. That is, speakers take previous usage into account when deciding whether or how far to extend an existing construction. If they have already witnessed a construction being extended with a wide variety of exemplars, they are more willing to use it productively themselves. Productivity begets productivity.

We have emphasized the following points regarding how constructions emerge and are used creatively:

1. Learners create lossy memory traces of formal patterns and their associated messages-in-context within their hyper-dimensional conceptual space.
2. New memory traces are related to previously existing memory traces, creating emergent clusters that relate formal properties with relevant aspects of the message-in-context.
3. The emergent clusters are constructions: learned pairings of form and function.
4. A construction is strengthened, becoming more accessible, when new representations overlap with existing representations. The range of a construction is broadened when witnessed instances are more variable.
5. Constructions are accessed in a content-addressable way during both comprehension and production, which encourages entrenched formulations to be reused.
6. Novel expressions are licensed by existing constructions to the extent that the existing combination of constructions **covers** the hyper-dimensional space required to include the novel expression.

We have seen that speakers extend constructions productively when an extension provides an appropriate and accessible way to express an intended message. In the following chapter, we will see that productivity is inhibited when there exists a more readily accessible alternative formulation to express the intended message-in-context.

Chapter 5

Competition

Statistical Preemption

The previous chapter emphasized the role that clusters (or neighborhoods) play in supporting generalizations. But clusters (or constructions) don't generalize, people do. That is, which constructions a speaker uses depends on what the speaker wants to say. This simple idea predicts that productivity is curtailed by the existence of an alternative formulation that conveys the intended message-in-context and is more accessible at the moment of speaking. This chapter focuses on how this STATISTICAL PREEMPTION, or competition in context, works (see also Boyd and Goldberg, 2011; Clark, 1987; Foraker et al., 2009; Goldberg, 1995, 2006, 2011a, b; Kim and Yang, 2017; Marcotte, 2005; Ramscar, 2002; Robenalt and Goldberg, 2015, 2016).

5.1 Constraining Morphology and Meaning

Researchers have long recognized that competition in context plays an important role in constraining generalizations when new complex words are created. For example, the suffix -*er* can be productively added to verbs to create agentive nouns. Familiar examples include *teacher*, *skier*, *listener*, *speaker*, and newer coinages include *blogger* and *texter*. But because we already use *cook* and *spy* as agentive nouns, the novel formulations, ?*cooker* and ?*spier* are judged to be ill formed. Similarly, although -*ness* is quite productive in combining with adjectives to form nouns (*fastidiousness*, *boorishness*), other adjectives sound odd if they occur with -*ness*. For example, ?*warmness* and ?*youngness* are odd, because they are preempted by *warmth* and *youth*; likewise, ?*honestness*, ?*jealousness*, and ?*caringness* are preempted by *honesty*, *jealousy*, and *caring*, respectively. Each of these novel formulations is "blocked" by an already existing and familiar word (Aronoff, 1976; Kiparsky, 1982). That is, speakers do not freely create new words when a readily available alternative that conveys the intended meaning-in-context already exists; moreover, they judge such overgeneralizations to sound awkward and infelicitous. In fact, we only call a generalization an *over*-generalization when there already exists a different, conventional formulation that native speakers use to express the intended meaning-in-context.

This idea should also be familiar from the discussion of word *meanings* in section 2.5, where it was argued that competition from other words helps learners avoid overgeneralizations. For example, children who initially overgeneralize the word *ball*, using it to refer to other round things such as the moon, will learn to narrow their use of the word *ball* not by witnessing additional instances of balls or additional instances of the word *ball* but by learning the word *moon*. Since semantically related words compete with one another in order to express a particular message-in-context, consistently witnessing the conventional word (*moon*) used to express "moon" in those contexts, *moon* will statistically preempt *ball* from being used to mean "moon" in the same type of contexts.

Notice that the appeal to the meaning or message-in-context is critical. Multiple words can coexist as long as they are used in distinguishable contexts (recall e.g., *thrifty* vs. *stingy*; *fast* vs. *quick*; *dog* vs. *chien* [French]). In the present chapter, I argue that the same type of competition that is relevant to learning word forms and word meanings also explains why native speakers judge certain syntactic overgeneralizations to be ill formed (e.g., ?*explain me this*).

5.2 Statistical preemption

How exactly is statistical preemption relevant to the fact that native speakers find the examples on the left in (5.1) decidedly odd?

5.1a	?She explained her the news.	(cf. She told her the news.)
5.1b	?She disappeared the rabbit.	(cf. She hid/banished the rabbit.)
5.1c	?She considered to go to the farm.	(cf. She wanted to go to the farm.)
5.1d	?the afraid boy	(cf. the scared boy)

There is nothing syntactically illicit about any of the examples, or if there is, it is not anything that is obvious to naïve learners. This is clear from the examples on the right within parentheses, which share identical syntax and yet are fully acceptable. The left-hand examples in (5.1) are also easy enough to understand. The key idea behind statistical preemption is simply that native speakers have learned a more conventional way to express the intended messages-in-context. We favor formulations that have become entrenched through previous exposure for expressing our intended messages, because we implicitly understand those formulations to represent the "right" way to use our language (recall section 4.5).

It may be unclear how or why being familiar with a verb in one construction could possibly lead that verb to be judged odd in a different construction, since we have already seen—emphasized!—that each ASC has its own unique function (Ambridge, et al. 2012a, b; Bowerman, 1988; Embick and Marantz, 2008;

Goldberg, 1995; Pinker, 1989). Indeed, verbs typically do occur in multiple constructions, including some quite novel formulations, as in (5.2):

5.2a " I . . . coughed him [a moth] out of my mouth."
5.2b "Sarah . . . winked her way through the debates."
5.2c "She'd smiled herself an upgrade."

Each of the verbs, *cough*, *wink*, and *smile*, has been witnessed frequently in the *intransitive* construction, and only rarely, if ever, in the various transitive constructions in (5.2); yet these novel examples are judged more acceptable than examples like those in (5.1) (Robenalt and Goldberg, 2015). Thus, witnessing a verb in one construction, even extremely frequently, does not automatically prevent that verb from being used felicitously in other constructions (see section 7.3 for further discussion).

Importantly, however, the sorts of ill-formed novel examples that we aim to account for, repeated below in (5.3) (left side), have conventional close paraphrases (right side) that have been consistently witnessed *instead of* the unacceptable types of examples. That is, statistical preemption predicts that we judge the examples on the left to be unacceptable because we have learned that native English speakers use the constructions on the right to express the intended messages-in-context.

5.3a	?Explain me something.	≈ Explain something to me.
5.3b	?He disappeared the rabbit.	≈ He made the rabbit disappear.
5.3c	?She considered to go.	≈ She considered going.
5.3d	?the afraid boy	≈ the boy who was afraid

The fact that an existing formulation must *compete* with a potential novel formulation in order for the latter to be preempted explains why the novel expressions repeated below in (5.4), are *not* preempted by the (highly frequent) intransitive uses of *cough*, *wink*, and *smile*: The intransitive expressions do not convey the same messages. Therefore, they would not be activated in the same contexts, and are thus not in competition.

5.4a	I . . . coughed him out of my mouth.	≠ I coughed.
5.4b	Sarah . . . winked her way through the debates.	≠ Sarah winked.
5.4c	She'd smiled herself an upgrade.	≠ She'd smiled.

The current emphasis on competition distinguishes statistical preemption from a distinct, albeit related, proposal that I refer to as CONSERVATISM VIA ENTRENCHMENT. Conservatism via entrenchment is the idea that unacceptability can be directly inferred from absence, without regard to meaning or context. For example, if a speaker is very familiar with a particular verb being used in one ASC, then conservatism via entrenchment predicts that the verb will resist being extended for use in *any other* ASC, regardless of whether the previously witnessed and novel messages compete or not. According to conservatism via entrenchment, the more frequently a verb has been witnessed in a language in any other construction, the more resistant it should be to being used in any new way. This idea has been labeled "negative entrenchment," or simply "entrenchment" (Ambridge et al., 2012b; Stefanowitsch, 2008), but elsewhere (Langacker, 1987), *entrenchment* is used to refer simply to frequency that results in increasing familiarity. For clarity, then, I refer to the use of frequency as indirect negative evidence of how words can*not* be used as *conservatism via entrenchment*, and refer to frequency as simply a proxy for familiarity as *simple entrenchment* (section 4.5).

All usage-based researchers agree that simple entrenchment is important: more frequent formulations are more accessible and are preferred. The question is whether witnessing positive evidence of a verb used in one way serves as indirect negative evidence that that verb cannot be productively used in *any other way*. Much of the time, the predictions made by conservatism via entrenchment and statistical preemption are highly correlated (Ambridge et al., forthcoming). However, note that the verbs *cough*, *wink*, *smile*, and *sneeze* are highly frequent in the intransitive construction and are rarely if ever witnessed in any of the transitive constructions in (5.2)/(5.4). Conservatism via entrenchment therefore predicts that speakers should judge the examples in (5.2)/(5.4) as unacceptable, with the degree of unacceptability dependent on each verb's frequency, and *not* whether there exists a readily available competing alternative way to express the intended messages. On the other hand, since statistical preemption relies critically on the idea of competition in context, it predicts that verbs may be used creatively when speakers want to express particularly novel messages (such as those in [5.2]/[5.4]). Immediately below, we summarize experimental evidence in support of statistical preemption. For further discussion of conservatism via entrenchment, see section 7.3.

5.3 Evidence

Statistical preemption has been investigated in relation to argument structure patterns in several experimental studies. An early study by Brooks and Tomasello (1999) found that if six-year-olds witnessed *Ernie's making the cow cham* (periphrastic causative), they were less likely to respond to "What did Ernie do to the cow?" with *Ernie chammed the cow* (transitive causative), than they were if they had witnessed the intransitive construction (*The cow is chamming*) (see also Brooks and Zizak, 2002). That is, children were willing to use *cham* in a new way to express their intended message (*Ernie chammed the cow*)—unless they had witnessed a competing alternative way to express the same message (*Ernie made the cow cham*).

Several other studies have exposed adult participants to *novel* constructions: that is, novel word-order patterns that were assigned novel abstract functions. In one study, the two novel word order constructions differed in terms of their information structure (Perek and Goldberg, 2015). In other studies, two novel word order constructions differed in terms of their semantics (Perek and Goldberg, 2017; Thothathiri and Rattinger, 2016). For instance, in Perek and Goldberg (2017), one transitive novel word order construction was associated with a strong effect on the undergoer, while another construction was associated with a weaker effect on the undergoer. Static images of the sample short videos are provided in figure 5.1. Participants learned six novel verbs, each with a unique meaning. In a *preemption* condition, a group of participants witnessed three verbs occurring only in the construction with a strong effect on the undergoer; two verbs occurring only in the other construction with a weak effect on the undergoer, and one verb stubbornly occurring in the "weak effect" construction regardless of whether the effect on the undergoer was actually strong or weak. After exposure, participants were asked to produce and judge the acceptability of various novel descriptions of new scenes, and they demonstrated an awareness of the distinct functions of the two constructions. Yet the same participants tended to use the stubborn verb only in the construction in which it had been witnessed; that is, it was preempted from being used in the alternative construction.[1]

This was not an effect of conservatism via entrenchment. We know this because, in a separate condition, a different group of participants witnessed three novel verbs exclusively in one construction and the other three novel verbs exclusively in the other construction. At test, participants witnessed actions that either had a strong or weak effect on the undergoer, and were asked to use the novel verbs to describe the scenes. Results in this condition were very different: people displayed a strong tendency to use the verbs in whichever construction was better suited to express the intended message, regardless of whether the verb had been witnessed in that construction during exposure. They also judged novel sentences to be acceptable as long as the scene and construction were

1 Likewise, if two novel constructions are assigned identical functions, people display a strong tendency to restrict verbs for use only in the construction in which the verb has been witnessed (Perek and Goldberg, 2015; Thothathiri and Rattinger, 2016; Wonnacott et al., 2008). This fact is also consistent with statistical preemption: if two constructions share the same function, witnessing a verb in only one of them statistically preempts its use in the other.

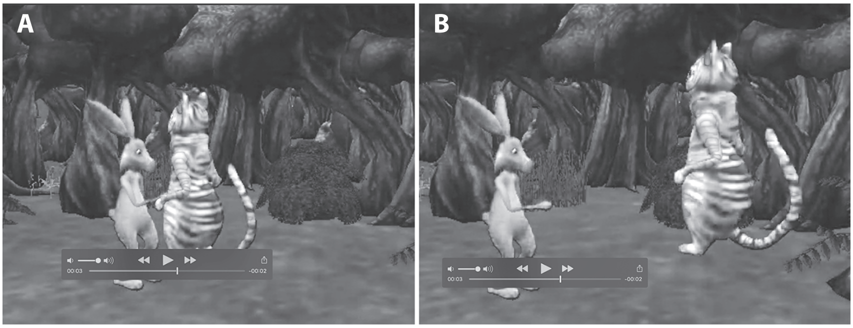


FIGURE 5.1. *A*, a screen shot from a sample video depicting a weak effect on the undergoer—a rabbit punches a cat, causing the cat to jump only slightly; *B*, a screen shot from a corresponding video showing a strong effect on the undergoer—the same action makes the cat jump out of the picture frame. Reprinted from Perek and Goldberg (2017) with permission from Elsevier.

semantically congruent. That is, participants tended to generalize the constructions for use in appropriate discourse or semantic contexts, largely ignoring evidence of verb-specific behavior during exposure (Perek and Goldberg, 2017).

Learners' markedly different behavior in the two conditions provides evidence that competition in context is key to restricting verbs' distribution. Participants were likely to generalize in the second condition presumably because the constructions had not been in competition during exposure; instead, each verb + ASC pairing witnessed was appropriate for expressing the semantics of the witnessed scene. Only in the preemption condition, in which one verb was witnessed stubbornly occurring in one construction, even when the other construction was better suited to describe the scene, did learners restrict that verb's distribution during later production and judgment tasks. Thus, speakers are willing to be creative in order to effectively express their intended messages, but creativity is curtailed if a familiar formulation has already been witnessed expressing that same type of message.

In fact, in the preemption condition of Perek and Goldberg (2017), speakers tended to restrict other verbs to some extent as well. That is, despite the fact that speakers recognized the distinct semantics of the constructions (in the case of new novel verbs), witnessing one verb being statistically preempted from use in the other construction led to more conservative behavior of other verbs witnessed during the initial exposure as well. We return to this important point again below.

While we have so far focused largely on the idiosyncratic behavior of verbs, another relevant study investigated certain *adjectives* that are oddly finicky in their distribution. Specifically, there exists a class of adjectives that begin with an unstressed schwa sound ("a-") and resist appearing prenominally, a position otherwise prototypical for adjectives. Examples of these A-ADJECTIVES are provided in (5.5) (Huddleston and Pullum, 2002).

***A*-adjectives**

5.5a ?an asleep child	5.5e ?an alive monster
5.5b ?an afraid man	5.5f ?an ablaze building
5.5c ?an alone boy	5.5g ?an afloat ship
5.5d ?an aware woman	5.5h ?the abloom flowers

The resistance of *a*-adjectives to being used prenominally is not due to an across-the-board semantic or phonological restriction, since near synonyms (5.6) and other adjectives with similar phonology (5.7) happily appear prenominally:

Semantic near synonyms

5.6a a sleeping child	5.6e a living monster
5.6b a scared man	5.6f a burning building
5.6c an isolated boy	5.6g a floating ship
5.6d a mindful woman	5.6h blooming flowers

Phonologically related non-*a*-adjectives[2]

5.7a an adult male	5.7c an acute sickness
5.7b an astute comment	5.7d an aloof woman

We can conclude that while *a*-adjectives are partly defined by the way they sound, the fact that they resist prenominal use does not strictly follow from general phonological or semantic factors.

The restriction on *a*-adjectives is motivated by the history of the adjectives as prepositional phrases in Old English. For example, *asleep* evolved from *on sleep* in Old English, *ablaze* came from *on blaze*, and so on (Long, 1961). As prepositional phrases, it made sense in Old English that they did not occur before nouns, since prepositional phrases do not occur in this position (e.g., ?*the on drugs man*). Of course, speakers today are generally unaware of the historical origin of these adjectives, so the historical motivation is not an explanation of how or why we still avoid saying ?*the asleep boy*.

2 The key distinction between the *a*-adjectives in (5.5) and the non-*a*-adjectives in (5.7) appears to be that each of the words in (5.5) involves *a*- together with a recognizable stem (Boyd and Goldberg, 2011, Coppock, 2009). That is, *alive* is "a + live"; *asleep* is "a + sleep"; *alone* is "a + lone." While *fraid* in "afraid" is not an independent word, it is arguably recognizable in *frighten* (and *fraidy-cat*). The adjectives in (5.7) are not segmentable in the same way. That is, /dult/ in *adult* is not meaningful, and neither are the syllables /stut/, /cyut/, and /luf/ in *astute*, *acute*, and *aloof* ("cute" is a word, but it has a completely different meaning).

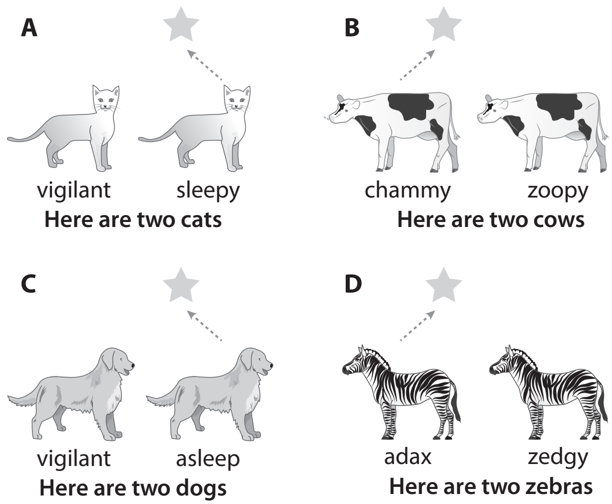


FIGURE 5.2. Example stimuli used in the production study of Boyd and Goldberg (2011): *A*, semantic near synonyms (e.g., *sleepy*); *B*, novel non-*a*-adjectives (e.g., *chammy*); *C*, familiar *a*-adjectives (e.g., *asleep*); *D*, novel *a*-adjectives (e.g., *adax*). Participants were asked to describe what happened when one of each pair of animals moved toward the star.

Boyd and Goldberg (2011) investigated how speakers could learn the restriction on *a*-adjectives for novel cases (e.g., *afek*). In each of three experiments, four types of adjectives were used, as illustrated in figure 5.2. These were familiar *a*-adjectives (e.g., *asleep*, *afraid*), familiar non-*a*-adjectives (*sleepy*, *scared*), novel *a*-adjectives (*afek*, *adax*) and novel non-*a*-adjectives (*chammy*, *tooky*). Undergraduates watched scenes that involved two identical animals, which were only differentiated by adjective labels. One of the animals moved to a star on the computer screen, then the screen went blank, and participants were asked to describe what had just happened. Of interest was whether speakers would describe the scenes by using the prenominal construction (e.g., ***the sleepy cow*** *moved to the star*) or the somewhat more cumbersome relative-clause construction: *The cow* ***that was asleep*** *moved to the star*. As expected, speakers reliably used *sleepy* and other familiar non-*a*-adjectives in the prenominal construction, and they avoided using *asleep* and other familiar *a*-adjectives this way, instead favoring the relative-clause construction. In the first experiment, participants showed a slight tendency to avoid using novel *a*-adjectives (e.g., *afek*) prenominally, instead producing them in relative clauses more often than they did adjectives like *sleepy* or *chammy*. This tendency apparently resulted from participants tentatively including *afek* and *adax* in the cluster of *a*-adjectives. The tendency to generalize from familiar exemplars to other exemplars that are relevantly similar exemplifies the idea of *coverage* discussed in the previous chapter.

In a second experiment with a new group of undergraduates, a small amount of preemptive exposure was provided. In particular, under the guise of explaining the experiment, two of four novel *a*-adjectives were witnessed occurring in the relative-clause construction (*The fox that's afek moved to the star*). Other adjectives were witnessed prenominally in order to avoid simply priming the relative-clause construction. The preemptive exposure had a striking effect: it resulted in a strong reduction in prenominal attributive responses for *all* novel *a*-adjectives, and no reduction for novel non-*a*-adjectives (e.g., *chammy*, *tooky*). The fact that the preemptive exposure led to an avoidance of prenominal use, not only for the two novel *a*-adjectives that had been witnessed in the preemptive relative-clause context but also for the two novel *a*-adjectives that had not been witnessed, indicates that speakers are capable of forming clusters that behave differently than the general case—i.e., "exceptions" need not necessarily be learned purely on an item-by-item basis. This is the same type of generalization we saw in the novel-construction-learning study (Perek and Goldberg, 2017): speakers generalize distribution to other examples that are judged to be relatively similar. In this way, it is possible to formulate the "restriction" in positive terms—speakers learned to use all of the novel *a*-adjectives in a relative clause by witnessing relevantly similar *a*-adjectives being used that way.

Finally, a third group of participants provided evidence that speakers can be quite savvy about what counts as a preemptive context. In this final study, two novel *a*-adjectives were again witnessed in relative clauses, but this time there was an independent reason for use of the relative clause, in that the novel *a*-adjectives were combined with a complex adjective phrase (e.g., *the hamster that's* ***ablim and proud of himself***). Since complex adjective phrases are unacceptable in the prenominal position for independent reasons (e.g., ?*the proud of himself hamster*), learners should not assume that *ablim* is responsible for the relative-clause use. Indeed, participants behaved shrewdly in the face of this "pseudo"-preemptive exposure: they essentially ignored it, and prenominal *a*-adjective descriptions were provided much as they were in the first experiment. Thus, speakers are willing to use language creatively unless they have witnessed a different formulation being used to express the intended message-in-context. When the use of an alternative formulation is attributable to extraneous factor, we do not appear to treat it as evidence of statistical preemption.[3]

3 Note that this last finding is not consistent with conservatism via entrenchment, which would predict that speakers should continue to use the novel *a*-adjectives in relative clauses simply because they had been witnessed in relative clauses. Instead, however, participants seemed to recognize that the witnessed relative-clause uses were not in competition with the prenominal use.

For learners to make use of statistical preemption, they must be able to recognize—at least implicitly—which constructions are in competition to express a particular message in a given type of context. We expect this type of learning to be incremental, with the speed of learning dependent on the particular constructions involved, how readily speakers appreciate their contextual constraints, and how often relevant instances are witnessed. In fact, not until age six or seven is the productive transitive causative preempted in experimental contexts (e.g., ?*She laughed him*) (Brooks and Tomasello, 1999), and not until age 10 do children display systematic avoidance of *a*-adjectives like *asleep* in experimental settings (see section 7.8).[4] There is also some evidence that children with language delays have more trouble than typical learners in taking competition into account (Manela-Arnold et al., 2010).

The idea that there is a connection between our judgments of ill-formedness on novel sentences and the existence of a readily available conventional alternative finds general support in certain neuroscience work. It is possible to detect electrical activity on the scalp evoked by brain responses to different types of stimuli, using event-related potentials (ERPs). One component wave of the complex ERP waveform is the P600, which is a positive wave form that peaks roughly 600 milliseconds after the onset of particular types of stimuli. When it was first discovered, the P600 was thought to index only syntactic processes, because it was reliably evoked when participants read sentences that contained syntactic errors (Friederici et al., 1996; Hagoort et al., 1993; Osterhout and Holcomb, 1992), such as those in (5.8):

5.8a ?The cat won't *eating*.
(cf. The cat won't *eat*.)
5.8b ?Every Monday he *mow* the lawn. (Osterhout and Holcomb, 1992)
(cf. Every Monday he *mows* the lawn.)

However, later work revealed that the P600 is also elicited by spelling errors (5.9a) (Münte et al., 1998), simple arithmetic mistakes (5.9b) (Núñez-Peña and Honrubia-Serrano, 2004), and incorrect word choices when the intended word is identifiable (5.9c) (Stearns, 2012):

4 Yang (2016) challenges the idea that statistical preemption can explain the restriction on *a*-adjectives by arguing that children do not receive enough of the required type of evidence before they are three years old. However, this is not surprising given children's delayed recognition of the restriction. See sections 7.4 and 7.8 for discussion (and Goldberg [2015] for a more detailed response).

5.9a The toddler was not yet too years old.
5.9b 2+ 2 = 5
5.9c The storyteller could turn any story into an amusing antidote.

Notice that in each example, including the syntactic violations, a preferable alternative formulation "comes to mind." That is, we recognize and are able to mentally correct each error. Thus, a compelling interpretation of the P600 is that it is evoked when listeners monitor and correct their input (van de Meerendonk et al., 2009). When we judge expressions such as ?*Explain me this* to be odd, the suggestion is that we are implicitly comparing it to the "correct" way to express the intended message-in-context, *Explain this to me*.

5.4 Recasts

Caregivers don't overtly correct children's errors very often because they are more likely to respond to the content of children's utterances than their form. But there is evidence that caregivers do sometimes *recast* children's utterances, implicitly correcting utterances that are ill formed. There is good evidence that children are exposed to recasts of words that have been mispronounced or used incorrectly (Bohannon and Stanowicz, 1988; Demetras et al., 1986; Hirsh-Pasek et al., 1984; Saxton, 1997). Moreover, at least within the domain of words, children appear to be quite responsive to recasts (Saxton, 1997; Strapp et al., 2008), although this remains somewhat controversial (Morgan et al., 1995). Some studies have reported rates of self-correction after a recast of 10%–50%, depending on the study and the child (Chouinard and Clark, 2003; Farrar, 1992).

Interpreting recasts as possible corrections for misused ASCs is tricky, since discourse factors typically constrain the use of ASCs, and the discourse context typically shifts subtly from one moment to the next. For example, the following would be perfectly normal mini-conversations. They both involve a parent's use of a different construction than that used by the child, but they should *not* be construed by the child as corrective, since the child's formulation is perfectly acceptable.

5.10 Child: Only *boys who were tall* made the team. (*tall* used predicatively)
Parent: One day you'll be *a tall boy*. (*tall* used attributively)
5.11 Child: I *gave the dog my sandwich*. (*give* in double-object construction)
Parent: You *gave your sandwich to THE DOG*!? (*give* in caused-motion construction)

What is needed for a recast to serve as preemptive exposure is for the child to recognize that the adult used a different formulation, not because of changing discourse demands, but because the adult's formulation was more appropriate in the original context. In this way, only a subset of recasts serve as evidence that one construction preempts another.

In other ways, preemption applies much more broadly than recasts do. Since listeners routinely predict what speakers are going to say (e.g., Pickering and Garrod, 2013), a witnessed formulation will serve as preemptive exposure whenever the witnessed formulation expresses the same message that the listener had anticipated. Moreover, since the witnessed formulation will be strengthened when it is witnessed in a given type of context, the link between that formulation and its intended message-in-context will be incrementally strengthened even when no other coinage was anticipated.

5.5 Explain Me This

We saw in sections 4.7–8 that sufficiently high coverage of an area in our conceptual space licenses generalizations within that space. Coverage relies on the idea that overlapping, lossy structured representations of witnessed exemplars cluster together within in our hyper-dimensional associative memory. Accessibility is increased by the activation of a cluster of closely related instances due to their overlapping representations. Relevant to an explanation of why native speakers find it odd to say ?*Explain me this* is the fact that *most* verbs that sound Latinate resist being used in the English double-object construction (recall section 3.3). Instead, these verbs (including *explain*, *transfer*, *return*, *detail*, *transport*) prefer an alternative way to express the intended message (e.g., *Explain this to me*). The existence of the cluster of verbs serves to reinforce the behavior of each member.

The notion of statistical preemption adds the key idea that constructions *compete* with one another to express a particular aspect of our message-in-context. The existence of a better—more entrenched—means of expressing a particular message-in-context explains our tendency to *use* the more entrenched construction: it is more accessible within our content-addressable associative memory. The reason speakers *judge* the novel formulations to sound odd is that they view language as a normative enterprise. Just as we consider *spier* to be "wrong" because we already have *spy*, we judge *explain* in the double-object construction (e.g., ?*Explain me this*) to be odd because we have consistently witnessed the alternative (e.g., *Explain this to me*) in contexts in which the double-object construction would otherwise have been appropriate (recall section 4.5). In order to determine the *strength* of preemption, the gradient probability in (5.12) is the key relevant measure (Goldberg, 2011b):

Probability that construction Z statistically preempts construction W
5.12 *P*(Z | contexts in which W is at least as appropriate)

For example, recall that in general the double-object construction is favored in discourse contexts in which the recipient argument is expressed by a pronoun and the theme is expressed by a lexical argument (e.g., *tell him a story*; section 3.4). By one estimate, the probability of *tell* occurring in the double-object construction with a pronoun recipient and lexical theme is .99 (Goldberg, 2011b). This is not the case for *explain*, however, which occurs instead in an *alternative* construction 99% of the time (as in [5.14]) (Goldberg, 2011b).

5.13 ?She explained him a story. (double-object construction)
5.14 She explained a story to him. (caused-motion [*to*-dative] construction)

That is, in contexts in which the double-object construction would be preferred for verbs that allow either construction (e.g., *tell*), the caused-motion (*to*-dative) construction is systematically used instead in the case of *explain*. Statistical preemption is *statistical*, because very occasional uses can be due to errors or the intentional flouting of norms for the sake of playfulness or memorability (as in the title of this book).

Thus, we can all agree that distinct phrasal constructions are virtually never semantically and pragmatically identical. Since each distinct construction is used in its own semantic and discourse niche, the possibility of verbs occurring in both members of a pair of constructions with closely related meanings (i.e., appearing in ALTERNATIONS) provides speakers with more expressive power, and many verbs do regularly occur in various alternations (Levin, 1993). When a verb is regularly witnessed occurring in two constructions that have closely related meanings, preemption does not extinguish one formulation in favor of the other. Instead, both versions survive, and each is used in contexts that conform to the distinguishable functions of the two constructions, as in (5.15) and (5.16):

5.15 She told the boy a story. (double-object construction)
5.16 She told a story to the boy. (caused-motion [*to*-dative] construction)

Similarly, *break* is regularly witnessed both intransitively as in (5.17a)—including in the periphrastic causative as in (5.17c)—and transitively as in (5.17b) (with a causative meaning). Because both (5.17b) and (5.17c) are regularly witnessed, neither preempts the other; instead, speakers learn to differentiate the two constructions. In this case, the transitive causative construction (e.g., [5.17b]) tends to be used for direct or intentional causation while the periphrastic causative (e.g., [5.17c]) tends to be used to express indirect or unintentional causation (Goldberg, 1995; Hopper and Thompson, 1980; Shibatani, 1976).

5.17a	The vase broke.	(intransitive)
5.17b	She broke the vase.	(transitive causative)
5.17c	She made the vase break.	(periphrastic causative + intransitive)

The fact that *some* verbs occur in two constructions with slightly different construals does not undermine the effectiveness of statistical preemption for other verbs. When learners repeatedly witness a verb used in one construction and not another that otherwise would have been appropriate, they learn to use the favored construction in lieu of the other option. That is, alternations provide *opportunities* for learners to witness one construction being used in contexts that would normally favor the other construction (if another verb were used). The rest of this chapter details how this works.

5.6 Calculating the Probabilities

The probability of one construction statistically preempting another one depends on the same calculations as VERB BIAS. For instance, the probability of the caused-motion (CM) construction statistically preempting the double-object (DO) construction for a given verb, $verb_i$, is determined by the formula in (5.18):

5.18 $P(\text{CM and } verb_i \mid \text{contexts appropriate for DO})$
$= \#(\text{CM}_{\text{DO context}} \text{ and } verb_i)/\#([\text{DO} + \text{CM}_{\text{DO context}}] \text{ and } verb_i)$

Recall that there is ample evidence that we retain the relative probabilities of verbs' appearance in competing constructions; a verb's bias toward one construction over another leads to, for example, faster reading times when the verb occurs in its preferred construction, when other factors are controlled for (see section 4.3; Clifton et al., 1984; Ferreira and Henderson, 1990; Garnsey et al., 1997; Linzen and Jaeger, 2014; MacDonald et al., 1994; Trueswell et al., 1993). The judgment that a verb + ASC combination is "wrong" arises when speakers have consistently witnessed a different way to express the intended message-in-context.

5.7 A Secondary Factor: Confidence

Suppose that the first time a learner hears *explain*, she hears it used in a caused-motion construction with a pronominal recipient and lexical theme (e.g., *Explain the story to me*). At that moment, the probability of witnessing *explain* in the caused-motion construction rather than the double-object construction is 1, but, clearly, the learner should not be confident from a single example that the double-object construction is preempted for *explain*. On the other hand, if a learner hears *explain* used in the caused-motion construction 100 times without ever hearing it used in the double-object construction, the probability hasn't changed—it is still 1—but the *confidence* of preemption should increase. We should note that confidence of one construction preempting another is not a simple linear function of frequency; it is not likely that confidence doubles when a person hears a second example, or that confidence increases twofold when exposed to 2000 as opposed to 1000 examples. We can capture the fact that confidence grows more slowly by appealing to the natural log function, which is commonly used in psycholinguistic work involving frequencies.

***Confidence* of statistical preemption of construction Z over construction W for *verb*$_i$**

5.19 ln*F*(Z when W would be at least as appropriate)

As represented in (5.19), confidence is estimated by the natural log of the frequency (*F*) of Z appearing when W would be appropriate. It is important to distinguish the key probability measure (in [5.18]) from this secondary measure of confidence when determining whether evidence for statistical preemption exists in corpus and judgment data. We return to this point in a discussion of certain work that has argued against a role for statistical preemption, as that work had used only the calculation in (5.19), not that in (5.18) (see section 7.3).

The idea that the frequency of a competing form plays a key role in judgments of acceptability of a novel form is supported by a large-scale judgment study of light verb constructions by Bonial (2014). Speakers were found to disprefer novel expressions like *give sanction* more when there existed a nearly synonymous alternative (*give permission*) that was high in frequency than when the only alternatives were lower frequency. Other experiments that have explicitly investigated a confidence measure on acceptability judgments on novel verb + ASC combinations have been reported by Robenalt and Goldberg (2015, 2016). These studies investigated 20 pairs of semantically related verbs—one with markedly higher frequency than the other (e.g., *disappear/vanish*). Novel sentences were then created for each pair of verbs. Importantly, not all of the novel sentences had a readily available competing alternative paraphrase. This was determined by asking a group of native speakers to paraphrase the novel sentences. Those sentences that were given the *same* paraphrase by more than half of the native speakers were considered to have a competing alternative. Those sentences that were paraphrased in a

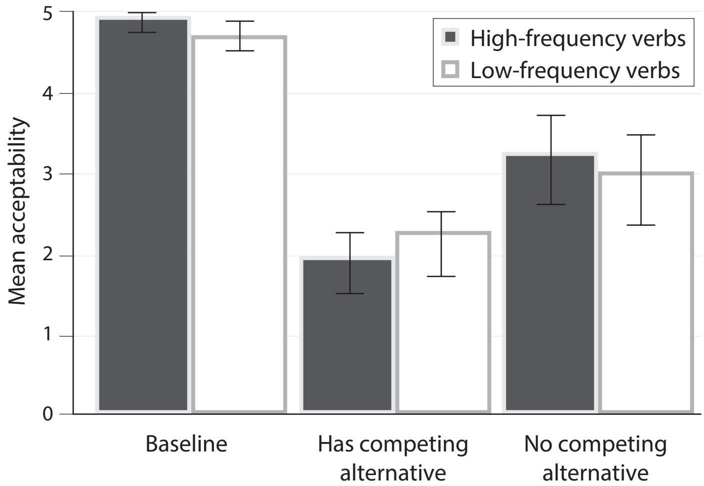


FIGURE 5.3. Acceptability ratings on pairs of sentences containing higher- and lower-frequency nearly synonymous verbs: in their familiar ASCs (baseline), in novel sentences for which there was a readily available alternative paraphrase (has competing alternative), or in novel sentence types for which there was no readily available alternative paraphrase (no competing alternative). 5, perfectly acceptable; 1, not at all acceptable. Error bars indicate standard error. Sentence length and judgments of plausibility were included as covariates. From Robenalt and Goldberg (2015).

broader variety of ways were considered not to have a competing alternative (see figure 5.3). For example, more than half of participants paraphrased the novel sentence in (5.20a) the same way (5.20b), so (5.20a) was determined to have a competing alternative. On the other hand, participants did not agree on a paraphrase for the novel sentence in (5.21a), but instead suggested a variety of different paraphrases (e.g., 5.21b–d), so (5.21a) was classified as not having a readily available competing alternative.

Novel sentence with a clear competing alternative
5.20a The dictator flooded propaganda into the city.
Agreed-upon paraphrase:
5.20b The dictator flooded the city with propaganda.

Novel sentence without a clear competing alternative
5.21a The woman screamed the children out of the ice-cream store.
Sample paraphrases:
5.21b A woman screamed and the children ran out of the ice-cream store.
5.21c The woman screamed at the children, and they left the ice-cream store.

5.21d The woman caused the children to run out of the store by screaming at them.

A separate group of people then rated the original novel sentences (e.g., [5.20a] and [5.21a]) on how natural or acceptable they sounded. Judgments in which each verb was used in its most typical ASC were also collected as baseline judgments.

The baseline, familiar types of sentences were all judged highly acceptable, with those containing the higher-frequency verbs judged to be even more acceptable than their lower-frequency counterparts. This aspect of the results illustrates the fact that more familiar formulations are judged to be more acceptable than less familiar formulations, in accord with the notion of simple entrenchment (recall section 4.5). Also consistent with simple entrenchment is the fact that both types of novel sentences were judged to be less acceptable than the baseline sentences, as is to be expected since novel sentence types are, by definition, unfamiliar.

At the same time, when the two types of novel sentences were compared—those with and those without a clear competing alternative—an interesting pattern emerged. Novel sentences that had a competing alternative (e.g., [5.20a]) were judged less acceptable than novel sentences for which there was no clear competing alternative (e.g., [5.21a]). Moreover, a previous finding that novel sentences containing lower-frequency verbs are judged more acceptable than novel sentences containing higher-frequency verbs as indicated in (5.22) was replicated (Ambridge et al., 2008; Brooks et al., 1999; Theakston, 2004; see section 7.3 for discussion). That is, novel sentences with a competing alternative as in (5.22) were judged higher in acceptability when they included the lower-frequency verb (*inundate* occurs less frequently than *flood*):

5.22 The dictator inundated propaganda into the city.
> The dictator flooded propaganda into the city.

Importantly, the preference for lower-frequency verbs in novel constructions only held for the sentences that had a competing alternative. In the case of sentence pairs that did not have an agreed-upon competing alternative, there was no significant effect of verb frequency (5.23) (*shriek* occurs less frequently than *scream*):

5.23 The woman shrieked the children out of the ice cream store.
= The woman screamed the children out of the ice cream store.

These facts make sense once we recognize that it is the frequency of the competing alternative that matters: the more often a competing alternative has been witnessed, the more confident speakers will be that the novel combination of verb + ASC is unacceptable. If there is no competing alternative, however, then the frequency of the verb does not predict acceptability.

To summarize, whether a novel sentence has a competing alternative plays a key role in judgments of acceptability as follows. Speakers prefer familiar verb + ASC pairings to less familiar ones (simple entrenchment; section 4.5). Greater familiarity with one verb + ASC pairing leads to less acceptability with *novel* alternatives that compete (statistical preemption). The more frequent the competing alternative is, the more confident speakers will be that the novel sentence is unacceptable. If there is no conventional way to express the intended message-in-context, then the frequency of the verb in other constructions does not matter. Instead, acceptability is determined by coverage, as discussed in chapter 4.

5.8 Mechanism: Error-Driven Learning

Before young children are exposed to vast amounts of language in a variety of different contexts, they sometimes produce unconventional formulations that are considered errors by native speakers. Such errors become suppressed over time in the following way. As we comprehend utterances, we attempt to anticipate what speakers will say next. There is a great deal of evidence for this (e.g., Eberhard et al., 1995; Johnson et al., 2013; Kutas and Hillyard, 1984; McRae et al., 1998; Pickering and Garrod, 2013; Stephens et al., 2010). And when a listener anticipates hearing a particular construction, mismatches between the input from the speaker and what had been anticipated by the listener provide an error signal that is used to improve future predictions through a process of ERROR-DRIVEN LEARNING (Chang et al., 2006; Elman, 1990; Pickering and Garrod, 2013; Rescorla and Wagner, 1972; Schultz et al., 1997; Sutton and Barto, 1998).

Even if a different formulation is only partially active, a subtly different type of learning, referred to in the memory literature as RETRIEVAL-INDUCED FORGETTING (RIF) can occur (Anderson, Bjork, et al., 2000; Anderson, Green, et al., 2000; Newman and Norman, 2010; Norman et al., 2007; Storm and Levy, 2012). RIF is used to describe the situation in which a representation is only partially activated in a particular context and a competing form is fully accessed instead; in this case the partially activated form is subsequently harder to retrieve in that context. Evidence for RIF has been found at the behavioral level and at the level of individual neurons: strong excitatory input leads to long-term synaptic strengthening, but *moderate* excitatory input leads to long-term synaptic weakening (Artola et al., 1990).

The effect of RIF has been demonstrated behaviorally in the following type of paradigm. Anderson and Spellman (1995) had a group of subjects learn paired associates—e.g., *fruit-apple*, *fruit-pear*, *fruit-kiwi*, *furniture-table*, *sport-tennis*, *furniture-chair*, and so on. Participants were then provided incomplete cues and asked to recall a subset of the pairs. For instance, one incomplete cue had the form:

5.24 fruit-pe___

Note that since *pear* is only partially cued in (5.24), subjects can be expected to partially activate other prototypical associates of *fruit*—e.g., *apple*. But since *apple* is not compatible with "pe___", it will be suppressed in favor of *pear*. Subsequently, subjects' memory for *fruit-apple* is weakened when compared to other unwitnessed pairs that had not been partially activated (e.g., *sport-tennis*). RIF can be thought of as a type of error-driven learning, with the key difference being whether the error was fully activated (a typical assumption in error-driven learning) or only partially or moderately activated (RIF).

Either RIF or more standard error-driven learning can explain statistical preemption. We can assume that if a listener expects to witness *disappear* used causatively (*The magician disappeared the rabbit*), but instead hears the periphrastic causative (*The magician made the rabbit disappear*), the periphrastic form will be strengthened and the causative version will be incrementally weakened. This predicts that a construction that is in competition will be weakened in a particular type of context, whenever another form "wins" (is used) in that type of context. If a competing construction is *not* even partially activated, there will be no suppression. This effect allows us to fine-tune our expectations as we gain experience. Because error-driven learning and RIF are domain-general mechanisms, no special process is required to account for their effects in language.

5.9 What Coverage Adds to Statistical Preemption

Importantly, different languages allow creative novel uses of verbs to differing extents, for different constructions (Goldberg, 1997). While English (and Chinese) allow novel uses of verbs, such as *sneeze the foam off the cappuccino*, Spanish and Hindi, for example, do not (Slobin, 1996b; Talmy, 1985). English and Chinese speakers, but not Spanish or Hindi speakers, witness a wide variety of verbs being used in each construction, and coverage determines whether an additional verb is a *candidate* for being extended. Statistical preemption then restricts candidates for a generalization when there already exists a familiar way to express the intended message-in-context.

A combination of statistical preemption and coverage is relevant to learning unusual distributions of very low frequency or novel verbs. Recall that the preemptive evidence for ?*the afek fox* was generalized to other novel *a*-adjectives (e.g., ?*the ablim cow*; Boyd and Goldberg, 2011); likewise, the restriction learned via statistical preemption on one verb was generalized to other verbs in the experimental context of Perek and Goldberg (2017). The clustering mechanism outlined in sections 4.7–8 predicts a degree of generalization across similar items in just this way (Barak et al., 2014). That is, clustering of "exceptions" leads speakers to generalize the behavior of exceptional cases.

Statistical preemption is predicted to only apply when there is a construction that is related closely enough to the target construction to compete with it (Bermel and Knittl, 2012; Goldberg, 1995, 127). This implies that constructions with particularly unique functions should not have the type of exceptions that tend to exist in the case of alternations. Instead, such NICHE CONSTRUCTIONS should be broadly generalizable without idiosyncratic exceptions, as long as they have been witnessed with a variety of verbs (sufficiently high coverage). I know of no systematic study of this, but anecdotally it seems to be the case. For example, the *way* construction conveys a very specific meaning involving creating a real or metaphorical path and motion through it (Goldberg, 1995), as in the examples in (5.25).

5.25a "The bandit . . . *robbed and murdered* his way through central California"
5.25b "he *thwacked* his way through more than 30 small watersheds"
5.25c "He *navigated* his way past wind-lashed trees and floating logs"$_{\text{COCA}}$

The productivity of this construction is well accounted for by coverage alone (Israel, 1996). At least, I know of no verbs with appropriate semantics that are restricted from appearing in this construction. The same appears to be true of the TIME-AWAY construction (Jackendoff, 1997).

One question that has not been fully resolved is whether a construction serves to preempt another construction *only* if the unwitnessed construction could reasonably have been expected to occur in the given discourse context. For example, if a learner witnesses a verb in the *to*-dative in contexts that are unsuitable for the double-object construction, does that serve as evidence that the double-object construction is not possible? To be concrete, would witnessing (5.26), which cannot be paraphrased with (5.27) for independent reasons, serve as evidence that (5.28) is not acceptable?

5.26 She blicked it to the man.
5.27 ?She blicked the man it.
5.28 She blicked me the book.

Arguably, witnessing (5.26) is irrelevant to determining whether (5.28) is possible because in American English, the double-object construction does not allow the theme argument to be *it* (?*She gave the man it*) (section 3.4). Recall that speakers appear to be capable of recognizing when an alternative formulation is required for independent reasons; participants seem to discount this type of input, treating it as essentially irrelevant to statistical preemption (section 5.3). Still, it is possible that such input serves as weaker preemptive evidence that could accrue over experience. Future work is needed to better quantify any effects it may have, since, to date, most work targeting preemption has investigated the nonoccurrence of constructions that would be at least as appropriate in a given context as the construction that is witnessed instead.

5.10 Summary

The previous chapter described how and why constructions or combinations of constructions tend to become conventionally used to describe certain types of messages-in-context. The hyper-dimensional conceptual space used to represent language is part of our associative memory. Therefore, lossy representations of language experience cluster together on various dimensions, including those related to form (phonology, grammatical categories, order, morphology) and function (meaning, information structure, register, genre, dialect). Dynamic categories that correspond to words and abstract constructions emerge from overlapping aspects of representations that are strengthened. When we speak, we combine familiar constructions in order to best express our intended messages-in-context. More entrenched and better-covered constructions are easier to access, which results in more conventional language being used more often, which further strengthens the association between conventional forms and particular messages-in-context.

The current chapter adds two additional points to these ideas:

- To the extent that speakers recognize that a way to express an intended message-in-context already exists, any novel formulation will be judged to be relatively ill formed, because we aim to conform to our speech community.
- If speakers anticipate a *non*conventional formulation, it will become suppressed over time via statistical preemption, a type of error-driven learning.

In the following chapter, we turn our attention to age effects and the role of accessibility.

Chapter 6

Age and Accessibility Effects

Children have been known to put diapers on their heads and marbles in their mouths, but in other ways they are less than fully creative. Imagine a young child brought to an elaborate carnival that boasts a dizzying array of games, newly rearranged each day. Dauntingly, the rules for each game vary from booth to booth, and no one is able to instruct the child as to how to play. Instead, adults and older children are busy playing the games themselves, skipping from booth to booth; their reactions convey a deep involvement, as they often stare fixedly, weep openly, or laugh heartily.

Young children would no doubt be intrigued and would want to play these games themselves. Some understanding of the carnival would come quickly; an impression of the sights and sounds, tacit knowledge of the physical layout of the most colorful booths or knowledge of which ones had the longest lines. But only slowly would most children be able to venture out and toss their first ring. Most likely, young children would initially stand close to others, trying to comprehend their intents and imitate their strategies. Learning to anticipate the rules of each game would require watching and listening, and children may initially misunderstand certain games, especially when the rules are complex or subtle. Children may not recognize that they need to stand behind the yellow line, they may misunderstand how the prizes are awarded, and they may toss the rings without great coordination. But, over time, children would learn to play the games effectively and efficiently, the way others do.

Language is not a carnival, but it does involve a dizzying array of conventional language games or constructions, each with its own quirks of form and function (Austin, 1962; Wittgenstein, 1953). To develop adult-like competence in a language is to learn how to use each word and each grammatical construction, in order to produce and comprehend new combinations in contextually appropriate ways without conscious effort. By the age of three, children are typically able to converse about tickling, hugs, and cotton candy.

To push the carnival analogy a bit further, when we as adults visit a carnival in a new place, some things are familiar and others are quite new. Someone from New Jersey who pops balloons with darts like an expert may struggle if the darts are a different size and weight at a carnival in New Orleans. An American would recognize a Ferris wheel at a carnival in France but might not appreciate that there is mayonnaise atop the fries. A French person visiting the United States may fail to notice the age restrictions on drinking or the weight restrictions on some rides. That is, prior knowledge affects our observations and our ability to adapt to new ways of doing things.

A child's ultimate goal is to learn how formal patterns are paired with functions, as linking forms to intended messages is required for understanding, and linking intended messages to forms is required for speaking. Children also aim to conform to the conventions of their speech community, since they want to be accepted as members of the social group. Thus, children must learn the constructions of their language(s), as constructions are language's conventional means of linking forms and functions. In this chapter, we focus on aspects of how learners' age and differences in prior experience impact language learning, offering an account of certain apparent contradictions in the literature, by situating the findings within the usage-based constructionist approach to language developed here. The chapter consists of four sections.

The first two sections address the following apparent paradox. On the one hand, there exist a raft of experiments that indicate that children initially tend to stay close to the input they witness— they tend to be *conservative* and prefer to use words, particularly verbs and other predicates, in the same ways they have previously witnessed them being used (section 6.1). At the same time, other work has found that children are more likely to "regularize" complex input than adults, producing a system that appears to be more general than what they had been exposed to (section 6.2). If regularization indicates robust generalization, it is hard to see how children can be both more conservative *and* more likely to generalize than adults. In this chapter, however, it is argued that the empirical observations are not contradictory, but are actually flip sides of the same coin.

In order to use language appropriately, children must learn various factors that determine each construction's appropriate use. That is, they must learn the various factors that *condition* each construction. As we saw in chapter 3, relevant conditioning factors can include meaning, information structure, register, dialect, and morpho-phonology. That is, the possible range of relevant conditioning factors is quite large and variable. It therefore takes children time to learn which factors are relevant to each construction.

In addition, recall that speakers aim to conform to the conventions of their speech community. To do this, they need to access from their associative memory the (combination of) constructions that most appropriately express their intended messages-in-context. If speakers fail to access an optimal (combination of) constructions, they will choose constructions that are *good enough*. In this chapter, we will see that failure to access an optimal solution can lead children to be conservative, sticking closely to familiar formulations, as discussed in section 6.1. Alternatively, if a complex system has not yet been learned well enough for a child to access previously witnessed formulations, the children may regularize or *simplify* the system by making use of only a subset of possible options, as discussed in section 6.2. In section 6.3, we review work that finds that children generalize appropriately at a relatively young age; the experimental designs used in this work typically offer children SCAFFOLDING within the experimental context, which serves to facilitate children's access to appropriate constructions before those constructions are fully mastered.

Section 6.4 addresses the unique challenges that adult language learners face. That is, while typically developing *children* are ultimately capable of fully learning the appropriate generalizations, subregularities, and idiosyncrasies that exist in every natural language, *adult* language learners commonly persist in making certain errors, such as ?*Explain me this*.

Learning to Learn: Figuring Out Which Dimensions Are Relevant to Which Situations

A single blade of grass rarely captures our attention, and neither are we likely to tune in to radio recordings made in unfamiliar languages. Instead, we tend to pay attention to aspects of our environment that are neither too boring nor too unfathomable. Infants likewise show a preference for stimuli that are neither too simple nor too complex (Kidd et al., 2012). As infants' understanding of the world develops, what counts as too simple or too complex changes as well. We are all particularly interested in whatever is informative and relevant to our goals, and infants need to learn which aspects of the environment are informative and relevant to their goals by scaffolding from prior learning.

For example, in the earliest stages of postnatal development, infants are sensitive to a variety of cues, including contrast, change, familiarity, and rhythm (Bardi et al., 2011; Farroni et al., 2005; Moon et al., 1993). On the basis of these cues, attention is drawn to motion and other visible changes, to faces, and to the sound of familiar languages, and this in turn leads infants to learn much more about the world. Tummeltshammer and Amso (2017) have found that infants as young as six months are able to efficiently use predictive cues. In particular, the infants were exposed to arbitrary arrays of colored objects arranged in repeated spatial configurations; the arrays of objects predicted where a face would appear. Infants quickly learned to attend to a particular location in anticipation of seeing a face (see also Emberson et al., 2017). Other work has shown that children learn to focus on what another person's *hands* are doing, more than the person's face (Fausey et al., 2016; Yoshida and Smith, 2008), likely because children can predict more about what will happen next on the basis of intentionally moving hands than by looking at a person's face.

Part of learning involves learning what to ignore as well as learning what to attend to. In fact, we have long known that between the ages of six months and a year, infants learn to take certain phonetic distinctions for granted—essentially ignoring them—when those distinction are wholly predictable (Kuhl et al., 1992; Werker and Tees, 1984a). The challenge of figuring out exactly how we learn which dimensions to attend to and which to ignore is beginning to be addressed in computational work (Liu et al., 2017; Russeck et al., 2017), but we are a long way from fully understanding this process. In what follows, we take it for granted that children have already learned at least a few nouns, verbs, and pronouns in order to focus on how they learn to flexibly use argument structure constructions (ASCs).

6.1 Younger Children Are More Conservative

Imagine that I illustrate a new word that means "slam shut" by slamming my laptop computer closed and saying *This is fapping*. If I then encourage you to do the same to your computer, and ask *What did you do?*, you might respond *I fapped my computer (you idiot)*, productively using the word *fap* in the transitive construction. That is, as an adult, if you learn a new word (e.g., *fapping*), you are willing to productively use it in an ASC that is distinct from what you had witnessed, as long as the words' meaning and the ASC are compatible and appropriate in the context.

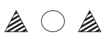
Young children are much less likely to do this. Study after study has shown that children initially tend to produce only minor variations of the formulations that they have witnessed in the input. While they readily substitute *I* for *you*, or *my dog* for *your cat* (recall section 4.4), they initially avoid more creative extensions and combinations of patterns (Akhtar and Tomasello, 1997; Baker, 1979; Bates and MacWhinney, 1987; Bowerman, 1982; Braine and Bowerman, 1976; Brooks and Tomasello, 1999; Dittmar et al., 2008; Lieven et al., 1997; MacWhinney, 1982; Savage et al., 2003; Theakston et al., 2001; Tomasello, 1992, 2000, 2003; Wonnacott et al., 2012).

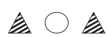
That is, children are more CONSERVATIVE than adults. Tomasello (2000, 2003), for example, has documented children's tendency to be more conservative at younger ages in dozens of experiments. In one experiment by Akhtar (1999), two-year-old children who witnessed a *familiar* verb (e.g., *push*) being used in a non-English word order (e.g., *Elmo the car pushed*) tended to "correct" what they had heard to produce *Elmo pushed the car*. But the same experiment showed that when two-year-olds witnessed a *novel* verb being used in a non-English word order (*Elmo the car gorped*) they were likely to reproduce the unusual word order. Older children were progressively more likely to correct both familiar *and* novel verbs (producing, e.g., *Elmo gorped the car*). These results imply that two-year-olds have learned the conventional way to use *push* and they follow this convention, but they do not assume that a new verb (*gorp*) follows the same pattern. That is, they do not immediately recognize that *gorp* is relevantly similar to other transitive verbs like *push*.

The relatively conservative behavior of younger children when compared with older children and adults is not entirely due to differences in the amount of exposure to language children have received. This is clear because a number of experiments have controlled for the amount of exposure by using novel constructions, and children still show more conservative behavior than adults. For example, Boyd and Goldberg (2009) exposed groups of five-year-olds, seven-year-olds, and adults to a novel construction that involved a non-English word order (e.g., *The doctor the player mooped*), paired with scenes of "approaching" in various manners labeled by novel verbs. After brief exposure to 16 pairs of scenes and utterances, adults readily generalized: they accurately associated completely new instances of the construction with completely new scenes of approach. The five-year-olds, on the other hand, appropriately recognized and interpreted the particular exemplars they had witnessed, but were unable to comprehend novel instances of the construction; the seven-year-olds displayed intermediate performance.

This is not to claim that five-year-old children are incapable of generalizing abstract constructions: they have already learned most of the ASCs of English by this age. But younger children are less successful than older children or adults at forming appropriate generalizations when provided with the same limited amount of exposure. This is an important point we return to below.

In order to align linguistic experiences in a way that they meaningfully cluster together, learners must recognize which dimensions are *relevant*. Children get better at this skill—which requires recognizing similarities and parallels among distinct tokens—as they acquire more knowledge both about language and about nonlinguistic contexts, but they are less adept at it than adults. For example, seven-month-old infants fail to recognize parallels among stimuli that follow an alternating (ABA) pattern when the stimuli involve three colored shapes (Marcus et al., 1999). That is, infants do not recognize the ABA pattern inherent in a series of stimuli such as (6.1a) and (6.1b) below, while older infants (and adults) do.

6.1a 
6.1b 

In order to recognize the shared pattern across (6.1a) and (6.1b) we need to implicitly appreciate that color and shape are relevant. Essentially, children must create an ad hoc category, "colored shapes," such that the shapes in 

and in ■ ▲ ■ are both instances of an ABA pattern. This is clear insofar as infants *do* recognize the same ABA pattern if the stimuli are created from a domain that they are more familiar with, such as pictures of dogs (e.g., beagle–collie–beagle, pug–poodle–pug), presumably because in this case infants are already familiar with a category "dog" and therefore are better attuned to relevant similarities and differences (Saffran et al., 2007).

Another example comes from Potter (2016, study 2) who exposed five- and six-year-old children and adults to a list of novel words that contained both systematic vowel (V) combinations and systematic consonant (C) combinations (all words were, e.g., CoCe or CaCi, and bVtV or gVdV). In a familiarity judgment task that followed five minutes of exposure, both children and adults showed above-chance recognition of items that had been witnessed. But while adults recognized new examples that obeyed either the vowel generalizations or the consonant generalizations as more familiar, children were at chance on novel words that had not been witnessed. Adults were in this way able to implicitly recognize the generalizations while children were not (see also Ferman and Kami, 2010; Raviv and Arnon, 2017).

The fact that younger children are less likely to generalize appropriately than older children or adults when exposed to the same input is also illustrated by research on "false" memory (Brainerd and Mojardin, 1998; Brainerd et al., 2002, 2008). In a standard paradigm, for instance, participants are exposed to a list of words that are related to a particular target word that does not itself appear. For example, if the target word is *doctor*, participants are exposed to words such as *nurse*, *sick*, *hospital*, *patient*, *cure*, and so on. After a short delay, adults are *more* likely than children to falsely believe that they had seen *doctor* in the original list. (Brainerd et al., 2002; Fisher and Sloutsky, 2005). That is, children tend to respond in a way that is more closely based on the words that were actually witnessed in the input instead of inductively "filling in" gaps in their experience via generalization.

While most children are quite good conversationalists by the time they are three, by that age they have witnessed between 8 million and 30 million words, a nontrivial amount of language (Hart and Risley, 1995). We have seen that children are demonstrably less adept at generalizing when compared with adults (Tomasello, 2000; 2003), even when the input is controlled for (Boyd and Goldberg, 2009; Potter, 2016; Schwab et al., 2018). Whether the multiyear process of language learning is impressively fast or more plodding is a matter of perspective. What is important for our USAGE-BASED CONSTRUCTIONIST approach is that children *learn* the grammatical patterns of their language via induction on the basis of their general cognitive skills.

To summarize, generalization requires that instances be represented in such a way that they cluster together along certain dimensions within our hyper-dimensional associative memory (chapter 4). Children's reduced ability to generalize stems from their reduced ability to appreciate the relevant dimensions that are required for generalization. That is, children's relatively conservative linguistic abilities stem from the fact that they are less likely than older children and adults to recognize and take advantage of relevant SIMILARITIES or parallels among exemplars. Other times, younger children are less likely to recognize or take advantage of relevant DISTINCTIONS among exemplars. This is the topic of the following section.

6.2 Younger Children Are More Likely to Simplify in Production

Both children and adults aim to respect the patterns in the input they witness. As we have seen, this follows from the first CENCE ME principle: speakers balance the need to be expressive and efficient *while obeying the normative conventions of their speech communities.* If a language contains irregular or exceptional cases, the normative response is to reproduce those irregular or exceptional cases. If the input systematically varies in a probabilistic way, the normative response is to produce it in the same probabilistic way.

But learners must be *able to learn the input well enough* to respect its distribution in their own productions. As reviewed in this section, languages are perceived to be particularly complex when they involve UNCONDITIONED (i.e., random, unpredictable) variation. When faced with such variation, children are more likely than adults to simplify it. A classic example comes from a deaf child, referred to as Simon, in work by Singleton and Newport (2004). Simon, like 90% of deaf children, was born to hearing parents. His parents learned enough American Sign Language (ASL) to communicate with their son, but as non-native signers, their ASL skills were imperfect. For example, in ASL, certain "classifiers" are used obligatorily; if you want to express that a car drove up a mountain, your hand should form a certain shape to indicate that *a vehicle* was in motion (as opposed to an animal or a person). Simon's parents used hand-shape classifiers more than half of the time, but they also regularly omitted them, in an unconditioned, haphazard way; that is, sometimes they remembered to use them and sometimes not. Simon, on the other hand, used the classifier hand shapes consistently in a way that approximated native signers (Singleton and Newport, 2004). That is, given that Simon could not discern any dimension that predicted when a classifier was used and when it was not (there was no conditioning factor to discern), he *simplified* his own productions.

Related work has demonstrated that when exposed to an artificial language in which a meaningless determiner appeared before nouns 60% of the time, adults tended to reproduce the determiner 60% of the time, essentially reproducing the input in a normative fashion. Children, on the other hand, tended to simplify the pattern, either by always producing the determiner or by always omitting it (Kam and Newport, 2005). Importantly, as the number of meaningless (and interchangeable) determiners was increased, adults also showed an increasing tendency to simplify, rather than successfully reproducing unconditioned variation (Kam and Newport, 2009; Ziegler et al., 2017). The fact that adults as well as children "give up" and simplify a language that they find unpredictable or overly complex suggests that the goal of learners of all ages is to respect the language they are exposed to; when they are unable to do that, they simplify, thereby making the language useable.[1] Because children have weaker working memory and metacognitive skills, they have a harder time keeping track of unconditioned variation and therefore are less likely to reproduce it accurately when compared with adults.

Simplification also occurs when the input *is* actually predictable, if learners fail to discern the relevant conditioning factors. Once again, this depends on both the complexity of the language and the sophistication of the learner. For example, Culbertson and Newport (2017) exposed children and adults to an artificial language in which adjectives only probabilistically preceded nouns (75% of the time; adjectives followed nouns the rest of the time), and numerals probabilistically followed nouns (or vice versa). Roughly half of the children (ages four to seven) failed to distinguish adjectives and numerals; they instead displayed a tendency to treat both categories alike, producing them both before, or both after, the noun. That is, without an appreciation of the fact that the semantic distinction between modification and quantification probabilistically predicted the difference in ordering, children simply treated adjectives and numerals alike, which led to a simpler system.

As was the case in the Culbertson and Newport (2017) experiment, it is more difficult for learners to discern which dimensions are relevant when the relevant factors are only probabilistically available or reliable (Samara et al., 2017). This is the case even when the relevant factor is quite salient, such as natural gender, as reported by (Schwab et al., 2018).

In particular, Schwab et al. (2018) exposed six-year-old children and adults to a mini-artificial language that included two meaningless "determiners" (*po* and *dax*). One determiner applied to three female characters and one inanimate object, while the other determiner was used with three male characters and a different inanimate object (e.g., *dax girl*; *po boy*). Thus, 75% of the entities were male or female, and natural gender perfectly predicted which determiner was witnessed for these entities.

1 For more discussion of the idea that languages change when new generations of speakers find them too difficult to learn, see Christiansen and Chater (2016) and Smith et al. (2017).

Remarkably, after repeated exposure, six-year-old children showed no evidence of detecting that gender was a conditioning factor, even though six-year-olds are well aware of the distinction between males and females (for better or worse), and even though they were demonstrably able to identify the intended gender of each character used in the experiment. The children instead simplified the system when asked to use the determiners with new gendered characters; in particular, they tended to produce only one determiner or the other. Adults, with their better metacognitive skills and stronger working memory, were highly accurate in their productions, appropriately extending the determiners with new gendered characters.

What are we to make of children's tendency to simplify complex input? Most conditioning factors in natural languages are only probabilistically available, and there are a great many different conditioning factors (recall sections 3.1–3.6). Without ample and/or targeted input, it is a challenging task to discern which factors are relevant for predicting complex systems, and natural languages are quite complex. Yet they are obviously learnable, since they have been learned. That is, given enough time and experience, children are *ultimately* successful at learning the conditioning factors that are relevant to the constructions in their language.

There is evidence that children are aware of complexity, even if they are unable to reproduce it. The evidence comes from a comparison of *production tasks*, in which learners need to accurately recall one of multiple options, and *judgment tasks*, in which learners are offered two options and need only to select between them. Several studies have found that learners simplify only in production tasks, regularizing their language insofar as they tend to overuse one option or another. When both options are *provided* in judgment tasks, demands on recall are minimized. In this case, learners display greater awareness of the language's complexity, even if they do not yet know how the complexity is conditioned. This has been highlighted on work with adults by Harmon and Kapatsinski (2017), and it is also evident in children's judgment tasks in Kam and Newport (2005).

To clarify, in the Schwab et al. (2018) experiment just described, a second group of six-year-olds took part in a judgment task where familiar and new characters were described using the determiners after the same exposure described above. In the judgment task, both determiners were offered to children, in order to reduce the accessibility demands. Specifically, children were asked, "Which do you think Mr. Chicken would say? *Dax boy* or *Po boy*?" In this task, children were more accurate on familiar determiner + noun combinations *and* they displayed an equal preference for both determiners. They still showed no evidence of discerning that natural gender was relevant when describing *new* gendered characters—they truly did not recognize natural gender as a conditioning factor.

Still, they were aware that the system was complex and that both determiners were equally appropriate, albeit with different words. Note that this is sufficient to allow memories of specific classifier + noun combinations to accrue over time, which would ultimately provide children access to the more complex system that involved conditioning factors of natural gender and specific words (in the case of the inanimate entities). A similar difference in behavior between production and judgment tasks was reported when variation in the input was unpredictable (Kam and Newport, 2005). The same pattern is also evident for *adults* when exposed to especially challenging systems: they simplify (or "regularize") in production, while revealing an awareness that the language is more complex and an intention to respect the complexities in judgment tasks that reduce demands on memory (Harmon and Kapatsinski, 2017; Kam and Newport, 2009).

Languages rarely contain truly unconditioned (random) variation, perhaps because it serves no communicative purpose and only adds to processing demands as it requires an arbitrary choice between options. Speakers of every age prefer conditioned over random variation (Kirby et al., 2008; Reali and Griffiths, 2009; Smith and Wonnacott, 2010). When learners *are* exposed to random variation, two possible long-term outcomes are possible. The first is that only one option continues to be used. This is the solution that the deaf child, Simon, converged on in his use of classifiers in ASL. Another option is that learners imbue what was unconditioned or random variation with a function, often sociolinguistic, such that a conditioned system emerges over time (Eckert, 2012; Janda, 1996; Smith and Wonnacott, 2010; Smith et al., 2017).

To summarize sections 6.1 and 6.2, younger children's reluctance to extend verbs for use in different constructions when compared with adults *and* their tendency to simplify complex input both stem from the same sources: their less robust ability to recognize which dimensions are relevantly similar or distinct across exemplars, and their reduced ability to access appropriate options from memory. That is, before children have mastered the conditioning factors that are needed to generalize each construction and distinguish it from other constructions, they are less able to access the most appropriate constructions (from an adult's perspective) and therefore will be more likely to rely on good-enough options.

Children display sensitivity to the complexities of the target language before they are capable of producing the complexity themselves. As the current proposal predicts, then, young children can be encouraged to generalize appropriately at relatively young ages when SCAFFOLDING is provided in the experimental context, which highlights the relevant dimensions and/or increases the accessibility of appropriate options. In section 6.4, we address why *adult* language learners typically persist in making certain errors that native speakers avoid.

6.3 Scaffolding Encourages "Early Abstraction"

We have reviewed evidence that, when provided the same amount of experience as adults, young children are less adept at recognizing the relevant similarities and differences that are needed to generalize appropriately. At the same time, it makes sense that children begin to form tentative generalizations immediately, as memory traces of bits of language are represented in their hyper-dimensional conceptual space. After all, by the time a child is roughly two years old, she has already witnessed anywhere from one million to four million utterances (Hart and Risley, 1995). Even tentative, simplified generalizations are sometimes sufficient for the task the child is faced with, particularly if scaffolding is provided that encourages appropriate generalization. Infants as young as 19 months have been argued to show evidence of "early structure mapping" or "early abstraction" (Yuan et al., 2012). This section argues that children at this tender age fall far short of adult-level mastery of the generalizations required for native-like language use, particularly when no scaffolding is provided in the experimental context. The fact that scaffolding is needed is further support for the idea that children find it challenging to access the most appropriate constructions, and that it takes time for them to learn to fluently use the relevant conditioning factors that determine how the constructions of their language are used.

To better understand the types of generalizations very young children are capable of and how they fall short of adult-like performance, it is worth considering a few experimental results in more detail. One interesting study demonstrated that 19-month-old children were able to distinguish English transitive (two-argument) sentences from intransitive (one-argument) sentences (Yuan et al., 2012). Novel verbs were used in an effort to show that children understood the distinction was a general one and not reliant on familiarity with particular verbs. Importantly, each child received one of three types of auditory pre-exposure before being tested on his or her comprehension of the novel verb *gorp*:

1. Transitive pre-exposure condition: children heard the novel verb in eight successive transitive sentences (e.g., *Mommy's gorping Daddy*!)
2. Intransitive pre-exposure condition: children heard the novel verb in eight successive intransitive sentences (e.g., *Mommy is gorping*!)
3. Mixed pre-exposure condition: children heard the novel verb in four transitive and four intransitive sentences.

After their pre-exposure, children witnessed two additional sentences (again, with *gorp*) that matched their pre-exposure, this time paired with two distinct scenes. One scene was semantically transitive (an agent acted on an undergoer) and the other scene was semantically intransitive (an agent performed an action on his own). Children who had witnessed the novel verb used transitively during both pre-exposure and at test spent more time looking at the transitive scene than the children who had witnessed the novel verb used intransitively during both pre-exposure and at test (67% vs. 53%). It may be tempting to view this finding as compelling evidence that 19-month-old children have a robust representation of the transitive construction, which involves an abstract actor and undergoer, as these quite young children appeared to interpret a novel verb as semantically transitive when it was witnessed in the English transitive construction. But notice that while no particular semantic context was provided during the pre-exposure, children did hear eight sentences with the novel verb, and each sentence contained two familiar arguments (e.g., *Mommy* and *Daddy*). The pre-exposure then provided the opportunity for children to *learn to expect two arguments* with the new verb, and only the semantically transitive scene matched this prediction.[2]

Moreover, children's performance in the Yuan et al. (2012) study was far less robust than adults'. To see this, consider children's performance in the mixed-exposure condition. The 19-month-olds who witnessed a novel verb used both transitively and intransitively during the pre-exposure did *not* show any preference for the transitive scene when they witnessed the verb used transitively at test. Yet a semantically transitive interpretation for *The duck gorped the bunny* was appropriate. Many verbs in English can be used transitively and intransitively with predictable differences in interpretation (e.g., *She broke it*; *It broke*). Instead, 19-month-old children only recognized the novel verb to be semantically transitive at test if they had *consistently* heard it with two recognizable arguments during pre-exposure. Indeed, an experiment with younger children found that 15-month-olds showed little evidence of learning that a novel verb was transitive, even in the transitive exposure condition, despite an increase in the number of sentences provided during pre-exposure and at test (Jin and Fisher, 2014).[3]

2 In an additional condition, Yuan et al. (2012) included a "bystander" in the intransitive scene at test—i.e., the scene included an animate entity who was not involved in the action. However, in this condition, the 19-month-olds' preference for the transitive event over the intransitive event was reduced from 67% to 56%, suggesting that the bystander foil was in fact a stronger lure than the intransitive scene that included only one entity. In any case, this result is hard to interpret since it was not specified whether 56% was greater than chance and it is also unclear that looking at the intransitive scene should be considered incorrect. If *gorping* were interpreted as a mental state verb ("scaring," "seeing"), the intransitive scene with bystander would provide an *appropriate* interpretation.

3 The study with 15-month-olds provided 12 sentences (instead of 8) with the novel verb during pre-exposure. Also, five (instead of two) repetitions were included during the testing phase. While the children in the transitive condition looked longer at the transitive scene

Much work, like this, has focused on whether children are capable of interpreting sentences containing one noun phrase as intransitive, and sentences containing two noun phrases as transitive (Fisher et al., 2010; Lidz et al., 2003; Yuan et al., 2012). But this "crowning jewel of structure-guided learning" (Pozzan et al., 2016, 4) sets the bar quite low. Insofar as language directs our attention, hearing two referring expressions in a single utterance *should* lead even young learners to expect two arguments in the corresponding scene. That is, we can expect that hearing nouns referring to entities directs listeners' attention to the referents of those nouns, so that if the choice is between a scene with two referents and a scene with only one, it's not surprising that children perform above chance (Goldberg, 2003). The fact that children don't perform as well as adults at distinguishing one-argument sentences from two-argument sentences by 19 months of age is evidence that nontrivial learning is required.

Other work has tested slightly older children on slightly more challenging tasks. For example, note that recognizing that two referring expressions should refer to two arguments is not the same thing as identifying which argument is the agent and which is the patient. In one study, two-year-old children successfully identified which scenes corresponded to *reversible* transitive utterances with novel verbs (Gertner et al., 2006). That is, two-year-olds correctly understood that if *The bunny is daxing the duck* then the bunny is acting on the duck and not the reverse. By the age of two, however, it is uncontroversial that children have reliably learned at least a handful of verbs and have learned how to interpret the arguments of those verbs. Relevantly, Gertner et al. (2006) included a practice session that made use of some of these familiar transitive verbs, and it seems that this provided essential scaffolding for above-chance performance (see also Arunachalam and Waxman, 2010; Arunachalam et al. 2013). In particular, a different study compared a group of 21-month-old children who received the same type of practice session involving familiar verbs and a separate group of children who did not receive the practice (Dittmar et al., 2008); this study found that only children who had the benefit of witnessing transitive verbs they were already familiar with subsequently performed above chance on the novel verbs, suggesting that the practice session provided children with a readily available analogical model that they made use of at test. The fact that the two-year-old children were able to benefit from witnessing familiar verbs

---

than the children in the intransitive condition did (56% vs. 40%), children's preference in a neutral condition was unexpectedly the same as in the transitive condition (56%). In a second attempted replication with the same design, children in the *in*transitive condition looked at the transitive scene 48% of the time, which is in striking distance of the 56% for the transitive condition in the first experiment (comparisons across experiments are not reported). These ambiguous results suggest that children's understanding of the transitive and intransitive constructions are far from robust at 15 months.

does indicate that they are at least transiently capable of recognizing the parallel between the familiar and novel verbs, but the fact that practice was required indicates that two-year-olds' facility with reversible transitives is not nearly as robust as adults'.

Contextual scaffolding has been found to be necessary for young learners to interpret constructions appropriately in other studies as well. For example, Conwell and Demuth (2007) found that three-year-olds were able to productively use the double-object or *to*-dative construction with a new verb, but only when they had witnessed several appropriate instances of each in the experimental context. That is, children's knowledge of various grammatical constructions begins in a tentative fashion and grows in strength over time and experience (Abbot-Smith et al., 2008; Boyd and Goldberg, 2012; Tomasello, 2000, 2003).

Tellingly, 21-month-olds mistakenly interpret intransitive sentences with conjoined subject (*The boy and the girl are gorping*!) as if they are transitive ("The boy is gorping the girl") (Gertner and Fisher, 2012; Noble et al., 2011; cf. 40-month-old children in Poznan et al. [2016]). This is expected if hearing two referents draws children's attention to two referents without a full appreciation of the English transitive construction (or the interpretation of *and*). The 21-month-old children tended to interpret only the first-named referent as the agent.

A general "agent-first bias" might be due to a processing preference to interpret (and produce) material that is more cognitively accessible before other material; this would explain why the animate-before-inanimate order generally dominates cross-linguistically. Alternatively, the tendency to interpret the first noun phrase as the agent may be due to previous experience with English or other languages that exhibit an agent-first bias (Abbot-Smith et al., 2017; Boyd et al., 2009; MacDonald, 2013; Wonnacott et al., 2008).

The CENCE ME proposal suggests that children begin early on to align bits of their language experience within their hyper-dimensional conceptual space. This allows for the fact that even very young children occasionally produce combinations of verbs and constructions that they have not witnessed in the input, as the following examples from my own children illustrate (ages in parentheses) (Goldberg, 2006, 60):

Aliza

6.2a ?*Up and down the neigh.* (20 months) ("Raise and lower the horse.")
6.2b ?*Come Abbi.* (20 months) ("Make Abbi come.")
6.2c ?*You jump me to the sky.* (25 months) ("Help me jump to the sky.")
6.2d ?*You mad to the pig?* (25 months) ("Are you mad at the pig?")
6.2e ?*I reach my hands up* (25 months) ("I raise my hands up.")

Zach

6.3a ?*It noises*. (28 months) ("It makes a noise.")

6.3b ?*Hold me what I want*. (29 months) ("Give me what I want, by holding me so I can reach it.")

While it is regularly assumed that these types of overgeneralizations must be created on the basis of secure knowledge of the relevant English constructions, it is alternatively possible that they are produced by children who are eager to express their intended messages but are not yet familiar with more conventional alternatives. In fact, we have already seen evidence for the latter perspective when it comes to children's overextensions of the meanings of words. Recall that a child may overextend the word *dog* to apply to cows and horses, but if asked to point to *the dog*, she will reliably choose a dog over pictures of other animals (Gelman and Naigles, 1995; Kuczaj, 1982). That is, the child uses the term *dog* because no better way of expressing her meaning is accessible at the moment of speaking. Similarly, when my 20-month-old daughter asked me to *Up and down the neigh* (6.2a), she did not yet have a secure handle on the transitive construction, but she had no conventional formulation handy (e.g., *Raise and lower the horse*). Evidence for this interpretation comes from the fact that *up and down* is not a possible English verb (and *neigh* is not a noun).

Thus, we have seen that experiments that investigate children's grammatical abilities before the age of two generally offer children some type of scaffolding within the experimental context. Such scaffolding has included consistent and unambiguous exposure to a novel verb before comprehension of the same verb is tested (e.g., Yuan et al., 2012). Alternatively, children have been exposed to familiar utterance types that then can serve as an analogical model for the interpretation of the novel utterances used immediately afterwards (Conwell and Demuth, 2007; Dittmar et al., 2008). Such scaffolding appears to be crucial to eliciting above-chance performance (Gertner and Fisher, 2012).

The fact that overgeneralizations become more common when children are older, together with the apparent need for some sort of scaffolding in context, is consistent with the claim, supported in sections 6.1 and 6.2, that very young children initially have trouble clustering instances in a way that facilitates appropriate generalizations. The idea that young children's generalizations may be more tentative or partial and therefore may need more contextual support than adults' generalizations is supported by much work in nonlinguistic category formation as well (Fisher and Sloutsky, 2005; Munakata, 2001; Munakata et al., 1997; Rovee-Collier, 1997; Vygotsky, 1978).

We have so far argued that young children need time to learn how to cluster exemplars appropriately, as they do not always recognize which dimensions of similarity or difference are relevant.

This raises a different question. If children have more difficulty than adults at recognizing which dimensions are relevant, why is it that *adult* learners of second languages tend to persist in producing certain errors while child learners eventually adopt the linguistic conventions that they are exposed to? Before they can jump rope, walk to school alone, or even tie their own shoes, children routinely surpass those of us who struggle to learn a second language as adults. As our own experience tells us, even though we tend to outperform children initially (Krashen et al., 1979; Snow and Hoefnagel-Höhle, 1978), we generally fail to achieve the same level of fluency as child learners, even after years of exposure to a new language (Clahsen and Felser, 2006; DeKeysar, 2005; Hartshorne et al. 2018; Newport, 1990 ; Wray, 2002). Although there exists evidence that native-like attainment by adults learning a second language is *possible* (Bongaerts et al., 1997; Marinova-Todd et al., 2000; Moyer, 1999), it is not the norm. As discussed in the following section, there are several factors contributing to adults' difficulties in learning a second language flawlessly.

6.4 Why Adult Learners of English Are Prone to Continuing Errors

We have already discussed the fact that adult second-language (L2) learners often fail to use lexically specified constructions and idioms the way native speakers do (section 4.2). They also struggle to avoid the types of errors in example (6.4) (Bley-Vroman and Joo, 2001; Bley-Vroman and Yoshinaga, 1992; Hubbard and Hix, 1988; Inagaki, 1997; Oh, 2010):

6.4a ?"could you *recommend me some [place]* to apply?"
6.4b ?"maybe it's better to *explain me* first"
6.4c ?"have you ever *considered to go* climbing in Ecuador? We would love for you to come and discover it!"

Why do adult learners continue to have trouble avoiding overgeneralizations while children generally recover from their errors? That is, why do L2 learners continue to produce expressions like ?*Explain me this* even after years of exposure to English? There are multiple related reasons for the disparity in outcomes between child and adult learners. Speech directed at children is particularly well suited for learning in that it tends to be highly repetitive, simpler, and slower than speech addressed to adults (Fisher and Tokura, 1996; Schwab and Lew-Williams, 2016; Uther et al., 2007). Moreover, adults who attempt to learn a language only in classroom settings receive far fewer hours of exposure than native speakers receive (Clark, 2003; Leow, 1998; Nagy and Anderson, 1984; Ortega, 2014). Therefore, some L2 learners have simply not had sufficient exposure or the right kind of exposure to master the types of subtle restrictions required to avoid *explain-me-this*–type errors (Flege, 2009; Matusevych et al., 2017; Muñoz and Singleton, 2011; Ortega, 2014). In fact, at the highest proficiency levels, L2 learners do treat the *explain-me-this* type of overgeneralizations the same way native speakers do (see Robenalt and Goldberg [2016] for discussion of relevant judgments). Still, many L2 learners of English live in English-speaking countries for decades and use English as their dominant language, and yet still do not reach the highest levels of proficiency. Therefore, the reasons for age effects in language learning must extend beyond the amount and quality of the input. Two complementary reasons are discussed below.

First, adults' well-practiced first language warps aspects of the hyper-dimensional conceptual space that includes the representations required for speaking. That is, the highly practiced representations that have been used for decades by adults distort certain similarities and distinctions that are needed to speak language automatically and without conscious effort. Secondly, the fact that adults need to inhibit their well-practiced L1 (native language) in order to process an L2 appears to result in a reduced tendency to make online predictions about upcoming grammatical choices; this in turn reduces the adult learner's ability to use error-driven learning from competing constructions, in the form of statistical preemption. These ideas are unpacked below.

6.4.1 Highly Entrenched L1 Warps Representational Space

As adults, we have become highly practiced in the linguistic skills we need for the languages we already use regularly, and these skills constitute ingrained linguistic habits in which we use certain forms to express certain types of messages in certain types of contexts (chapter 4). We have implicitly learned that certain dimensions and not others are important for clustering exemplars used to express various types of messages-in-context. This is the case for sounds, words, lexically filled constructions, and abstract ASCs.

For instance, Portuguese speakers can use a single term for both lemons and limes (*limão*), so they are more likely than English speakers to incorrectly refer to lemons as *limes*. In Farsi, the third person pronoun, /*u*/, is used to refer to both males and females ("he" and "she"), and Iranian learners of English are more prone to make errors in selecting *he* or *she* than are native English speakers, who rarely make this error. L2 learners famously have trouble with formulaic language, occasionally producing mangled idioms such as *Don't try to mustard me up* or *It's just not my bowl of soup* (cf. *butter me up*; *cup of tea*) (recall section 4.2; also Kim, 2015). Spanish learners of English might inquire whether some-one *has* hunger because this the way the question is asked in Spanish (*Tienes hambre*?). In fact, errors in lexically specified constructions are more common than errors in single word choices, quite possibly because multiword utterances are less often directly translatable from one language to another than individual words are.[4] Spanish has definite determiners (*el/los*), as does English (*the*). While Spanish speakers use the plural definite determiner in "generic" contexts (***Los** perros son mamíferos*), native English speakers use bare plurals (*Dogs are mammals*) in this context. Notably, Spanish speakers who learn English as a second language are prone to using the *English* definite determiner in generic contexts (?*The dogs are fun*) (Ionin and Montrul, 2010).

These are all interference or "transfer effects" from adults' more familiar language(s) to their new language (Ambridge and Brandt, 2013; Austin et al., 2015; Bates and MacWhinney, 1981; Bley-Vroman and Joo, 2001; Ellis, 2006; Finn and Kam, 2015; Foucart and Frenck-Mestre, 2011; Goldberg, 2013; Hernandez et al., 2005; Kim and Yang, 2017; Luk and Shirai, 2009; MacWhinney, 2006; Rutherford, 1989; Sabourin et al., 2006; Selinker and Lakshmanan, 1992). Once a second language becomes highly practiced, it can have effects on a speaker's L1 as well (Dussias and Sagarra, 2007; Marian and Spivey, 2003). Children are not affected by their L1 to the same degree when learning a new language, because they have not yet achieved the same degree of fluency or routinization of their language(s) (Matusevych et al., 2017).

Once transfer effects are recognized, the question arises as to why there isn't more interference than there is. The answer to this likely stems from an observation made by Brian MacWhinney and colleagues. They point out that each language shares more and stronger connections among its own words, intonation and sound patterns, constructions, and contexts, which predictably leads to greater activation ("resonance") among patterns within a given language than across languages. For example, they note that because the Spanish word for "table," *mesa*, is richly interconnected with other Spanish words and constructions, it receives much greater activation than *table* when Spanish is being spoken (Hernandez et al., 2005, 2; see also Li et al., 2004).

Importantly, transfer effects are distinct from traditional Whorfian effects of language on nonlinguistic cognition, in which knowledge of a particular language influences our conscious perception of the world (Gentner et al., 2002; Levinson, 1996; Regier and Kay, 2009). Portuguese speakers are fully capable of distinguishing yellow lemons from green limes. Farsi speakers are as well attuned as anyone else to the distinction between men and women. It is primarily the conceptual space involved in speaking and comprehending *language* that is affected by our previous knowledge of other languages. This is a type of "thinking for speaking" that has been suggested by Slobin (1996a). Recall from section 2.1

4 Alternatively, adult learners may not amass as much knowledge of multiword exemplars as children do because adults tend to analyze phrases into words or morphemes from the start (Arnon and Christiansen, 2017; Siegelman and Arnon, 2015).

that the hyper-dimensional conceptual space in which our knowledge of language is situated does not include our entire conceptual and perceptual space, as the latter additionally encompasses our knowledge of and capacity to perceive smells, images, feelings, and experiences that are not represented by any fixed category of language. Our basic conceptual and perceptual capacity must be shared across speakers of different languages, or it would be not only difficult but impossible to learn a second or third language as adults.

L2 errors provide ample evidence for the idea that learning a language involves learning which constructions to use *in which contexts*. Intriguingly, difficulties can arise even when the same conceptual distinction is obligatory in both in L1 and L2, but is required in a wider range of contexts in L2. For example, Antón-Méndez (2010) has observed that Spanish speakers, like Farsi speakers, are more prone to make errors in selecting *he* or *she* in English than native English speakers are. Spanish *has* distinct pronouns *el* ("he") and *ella* ("she"), but these pronouns are primarily reserved for contrastive contexts; in other contexts, the subject argument is indicated by gender-neutral agreement on the verb. It seems that the tendency to conflate gender in Spanish in verb marking, coupled with the relatively infrequent use of *el* or *ella*, results in a somewhat less automatic choice of *she* versus *he* when speaking English. Similarly, in Korean, plural marking is only required in definite contexts, and Korean learners of English are faster to read and more accurate at judging *English* plural marking on definite nouns than indefinite nouns (Choi and Ionin, 2017).

How are L2 speakers' errors related to the earlier sections of this chapter (6.1 and 6.2), which reviewed evidence that adult language learners are *better* than children at identifying which dimensions are relevant for the clustering of exemplars? Importantly, the examples of adults identifying relevant factors better than children came from experiments with very limited exposure and quite simple and broad generalizations. When learning new natural languages, adults likewise outpace children initially (Krashen et al., 1979; Snow and Hoefnagel-Höhle, 1978). Adults are quite expert at recognizing broad generalizations in these circumstances. However, many of the conditioning factors required to master individual constructions in real natural languages are quite subtle (chapter 3).

Adult language learners face particular challenges when a distinction required in L2 is irrelevant in their L1, particularly if the distinction is obscure. For example, we noted that *explain* patterns with most other Latinate-sounding verbs in avoiding the double-object construction in favor of the *to*-dative (section 3.3). The cluster of Latinate verbs mutually reinforce each other, encouraging native speakers to treat novel Latinate-sounding words in the same way (Ambridge et al., 2012b; Gropen et al., 1989). L2 learners, on the other hand, are unlikely to be sensitive to the distinction between Latinate verbs (e.g., *explain*, *display*, *purchase*, *transfer*) and semantically similar Germanic verbs (e.g., *tell*, *show*, *buy*, *give*), because such distinction would be irrelevant in their more familiar language(s).

More generally, we have seen that learning a language involves associating each construction with a range of context types in which the construction is appropriately used. Adult native speakers have assigned each particular construction of their L1 to a particular range of context types, and this assignment has been reinforced and fine-tuned over decades. Once a collection of context types has been categorized together for the sake of an L1 construction, it becomes more difficult to assign an overlapping but distinct range of context types to a construction in L2. In this way, the hyper-dimensional conceptual space in which we represent our linguistic knowledge becomes somewhat warped by our L1.

The requisite idea is familiar from research in phonology. For example, English treats the sounds /r/ and /l/ as distinctive in that the difference between these two sounds distinguishes certain words (e.g., *rap* and *lap*). Korean, on the other hand, has no two words that only differ in the choice of /r/ versus /l/. Korean infants are fully capable of hearing the difference between /r/ and /l/—as they would need to be in case they happen to be raised in an English-speaking environment—but by the time they are toddlers, the ability to make use of the difference between /r/ and /l/ has markedly declined (Borden et al., 1983; Werker et al., 1981). Thus, the conceptual space used by Korean speakers to represent language deemphasizes the distinction that English speakers attend to in order to distinguish *rather* from *lather*. Other languages treat other dimensions as more important than English does. For example, many languages, including Chinese, Hindi, and Thai, contain words that only differ in how long a delay there is after stop consonant sounds (/t^h/ vs. /t/), whereas no two words are distinguished by this difference in English. Native English speakers have trouble recognizing the difference between [t^hap] and [tap]), as both sound like the same word, *tap*.

The fact that we tend to ignore distinctions that are not meaningful in our language offers a useful illustration of the fact that only our representational space of *language* is warped. That is, our underlying perceptual capacity remains intact (Carstensen, 2016; Firestone and Scholl, 2016). We remain capable of distinguishing sounds at a finer granularity than is needed for distinguishing words if the context is not a linguistic one (Werker and Tees, 1984b), or for the purposes of predicting the expected distribution of sounds (Feldman et al., 2009; Kleinschmidt and Jaeger, 2015; McMurray et al., 2003). In other words, we learn to ignore distinctions that are not relevant to the task of understanding the speaker's intention (Feldman et al., 2009), but our brains retain the ability to detect phonetic differences, as we must in order to recognize distinct accents or individual speakers.

To summarize, our ability to consciously recognize the factors that condition ASCs remains intact, but many L2 errors stem from the way our first language has essentially warped the conceptual space that represents language, and it is these representations—our network of constructions—that allow us to use language without conscious effort (see also Jäschke and Plag, 2016; LaPolla, 2016; Montrul and Ionin, 2012; Schwartz and Sprouse, 1996).

6.4.2 Reduced Tendency to Predict Grammatical Forms

Native speakers are impressively adept at generating expectations about upcoming words and grammatical forms as they comprehend language (Arnold, Eisenband, et al., 2000; Dahan et al., 2000; Sedivy et al., 1999). We take all available cues into account, in real time, and activate words before the speaker is even finished uttering them (Marslen-Wilson and Zwitserlood, 1989). For instance, we immediately anticipate food-related words as soon as we hear the word *eat* (Altmann and Kamide, 1999). If a speaker utters *uh* or *um*, we expect her to introduce a new entity rather than one that has already been mentioned, presumably because we recognize that the latter should be easy to access (Arnold et al., 2004). In a context in which multiple apples are present and someone says *the apple,* we anticipate a modifying phrase (e.g., *the apple on the napkin*) that will indicate which apple the speaker intends (Tanenhaus et al., 1995). Children have also been found to anticipate upcoming speech in their native language (Fernald et al., 2010; Kidd et al., 2011; Lew-Williams and Fernald, 2007; Lukyanenko and Fisher, 2016; Swingley et al., 1999). Our expectations become fine-tuned through the process of error-driven learning, as reviewed in chapter 5. If we expect one thing and witness another, the error signal leads to a change in the strengths of the connections that predict which constructions are used in which contexts. This in turn leads to more accurate predictions in the future.

How is this relevant to the issue of second-language learners? A number of related findings converge to indicate that L2 speakers are less likely than native speakers to predict upcoming forms during online comprehension, even when they demonstrate knowledge of the forms during production and in off-line tasks (Grüter et al., 2014; Ito et al., 2017; Kaan, 2014; Kaan et al., 2010, 2016; Lew-Williams and Fernald, 2010; Martin et al., 2013). Theres Grüter has described this as the “RAGE” hypothesis: the idea that non-native speakers have a **r**educed **a**bility to **g**enerate **e**xpectations (Grüter et al., 2014).

This idea can be illustrated by an anecdote from a Princeton undergraduate who grew up in Mexico. The student recalled how his parents got into an argument when his mother, a native Spanish speaker, gestured in the direction of the microwave and the stove below it, and asked her husband to bring her the food that was *en el, el, el,* . . . The student’s father, an American who was a fluent L2 speaker of Spanish, was flummoxed, not knowing whether his wife wanted what was in the microwave (*en el micro*) or on the stove (*en la estufa*). Notably, he did not make use of the determiner, *el*, to differentiate the two possibilities, while the student's mother thought it was obvious that the microwave was intended.

In fact, in a systematic study that measured participants' eye gaze, Lew-Williams and Fernald (2010) found that while native speakers of Spanish (like my student's mother) made use of the gender marking to predict the upcoming noun, non-native speakers, even if they were reasonably fluent (like my student's father), did not, even though they were able to use the correct genders in their own productions.

Individual differences, degree of proficiency, the relationship between L1 and L2, and task demands all play a role in the extent to which predictions about upcoming forms are made by L2 comprehenders (e.g., Borovsky et al., 2012; DeLong et al., 2012; Foucart et al., 2014; Havik et al., 2009; Hopp, 2013; Kaan, 2014; Linck et al., 2009; Martin et al., 2013; Rossi et al., 2017; Wilson, 2009). But a reduced ability to generate expectations about grammatical choices is not entirely unexpected, as L2 speakers experience greater cognitive load due to increased self-monitoring (Levelt, 1983) and the need to inhibit their L1 (Green, 1998). To the extent that L2 learners' ability to predict upcoming grammatical forms is reduced, the CENCE ME principles predict that competition-driven learning will be correspondingly reduced. In particular, if non-native speakers do not anticipate upcoming utterances to the same extent as native speakers do, they will have less opportunity to learn from predictions that are subsequently corrected.

In fact, recall that when judging novel combinations of verbs and ASCs, *native speakers* implicitly but systematically take into account whether or not there already exists a conventional means of expressing the intended message-in-context (section 5.3). In particular, unless there is some reason to flout conventional language (e.g., for memorability, emphasis, playfulness), native speakers prefer conventional combinations of verbs and ASCs over novel ones, and they are more accepting of novel combinations when no standard conventional alternative exists (e.g., Robenalt and Goldberg, 2015). Of interest is the finding that *non-native* speakers essentially appear to treat (interpretable) novel combinations of verb + ASC the same regardless of whether there exists a conventional alternative means to express the intended message-in-context (see Figure 6.1 and Robenalt and Goldberg, 2016). That is, novel combinations of verb + ASC, with or without a conventional competing alternative, were judged by L2 speakers to be less acceptable than familiar combinations, with no discernable difference between them. Only at the very highest proficiency levels did L2 judgments align with native speakers' (Robenalt and Goldberg,

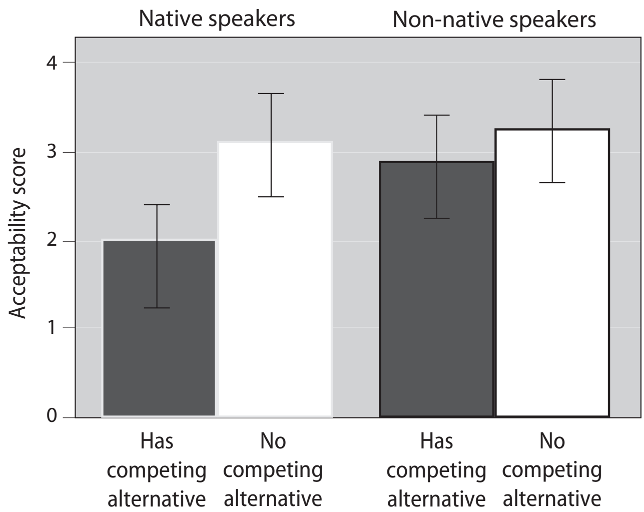


FIGURE 6.1. Unlike native speakers, L2 speakers judge sentences that have a competing alternative to be as acceptable as those without a competing alternative (Robenalt and Goldberg, 2016). Error bars represent standard error.

2016; see also Ambridge and Brandt, 2013; Kang, 2017; Tachihara and Goldberg, 2018; Zhang and Mai, 2018).

6.5 Summary

The trick to speaking a language like a native speaker is to use language in a way that conforms with the conventions used by other native speakers. Generalizations are appropriate to the extent that the required constructions have been well attested (in terms of coverage; recall chapter 4), and to the extent that there is no competing conventional form that would better match the intended message-in-context (statistical preemption; recall chapter 5).

According to the present USAGE-BASED CONSTRUCTIONIST approach, learned generalizations emerge on the basis of distributed and partially abstract exemplars that cluster together in the hyper-dimensional conceptual space in which language is represented; clusters compete with one another when more than one are simultaneously activated for the same purpose. The clustering and competition work efficiently and effectively for adult native speakers to enable them to use language in creative but constrained ways.

We can summarize the present chapter as follows. An apparent paradox in L1 research—namely, that children are both more conservative and more likely to simplify than adults—was addressed. We observed that children are not as adept at quickly discerning which dimensions are relevant in order to appropriately cluster linguistic exemplars within their high-dimensional conceptual space. When children fail to identify or make use of relevant SIMILARITIES or parallels among exemplars, they fail to generalize, and behave conservatively (section 6.1). When children fail to identify or make use of relevant DISTINCTIONS among exemplars, they oversimplify, which results in overgeneralization or "regularization" (section 6.2).

The tendency to simplify or "regularize" occurs primarily during production, when learners are required to try to use a complex system that they have not yet mastered. The available evidence suggests that learners understand their novice status, in the sense that they display an awareness, in judgment tasks in which accessibility demands are reduced, that the target system is more complex than they are capable of reproducing. This idea is familiar in the case of overextensions of word meaning: children may overgeneralize the word *dog* to label a cow, but if asked to "point to the dog," they reliably select a dog over a cow. Thus it seems that children over-rely on a more accessible option that is "good enough" if they cannot access the option that adult native speakers prefer, either because they have not learned the better option at all, or because they have not become fully fluent at accessing it. We have seen the same effects in artificial-language experiments: if two options have been witnessed, learners may over-rely on a single option, but they reliably display an awareness that both options were witnessed in forced-choice judgment tasks.

Children can be encouraged to access a more appropriate generalization when scaffolding is provided in the experimental context once they have begun to tentatively form the relevant generalizations (section 6.3). While children's reduced ability to detect which dimensions are relevantly similar or distinct leads to slower initial language learning than adults, over time, children surpass adult learners. This chapter argued that this is because, as exemplars accrue and form more well-covered clusters, children eventually discern *whichever* dimensions systematically predict how the constructions of their language are used.

Adult (L2) learners, on the other hand, are more prone to ignore subtle dimensions that are irrelevant in their L1 (section 6.3). This type of "transfer" effect is familiar from phonology, as it is widely recognized that we *learn* which phonetic distinctions to attend to and which to ignore. In the case of grammatical constructions, we can expect an adult's more dominant language(s) to become well entrenched over decades of practice. The hyper-dimensional conceptual space that contains linguistic representations (the constructicon) then becomes warped for the purposes of automatic language use. In particular, distinctions that are relevant for selecting one construction over others will become more strongly associated with that construction. That is, distinctions that are useful for predicting which constructions to produce or which to expect to witness will be weighted more heavily in the lossy representations of those constructions. Adults remain capable of perceiving dimensions that do not distinguish grammatical constructions in their own highly practiced first language(s), but making use of such distinctions can be expected to be more cognitively demanding, and will result in more errors.

Perhaps relatedly, L2 speakers appear to demonstrate a reduced ability to predict grammatical options, except at the highest proficiency levels. This may also be because of the increased cognitive demands of using a less dominant language (section 6.3). To the extent that learners fail to predict upcoming forms during comprehension, they will be less able to benefit from error-driven learning when conventional formulations are witnessed instead, since there will be no error signal. In this way, L2 speakers may benefit less from statistical preemption.

We can summarize the main points of this chapter as follows:

1. The failure to use relevant SIMILARITIES or PARALLELS among exemplars leads to conservative behavior.
2. The failure to use relevant DISTINCTIONS among exemplars leads to simplified productions.
3. Children are less adept at discerning which dimensions are relevant, but slowly accrue exemplar knowledge so that appropriate clusters are ultimately formed.
4. Highly entrenched L1 (or L2) can subtly warp the hyper-dimensional space used for language representations, resulting in transfer effects.
5. Non-native speakers appear not to anticipate upcoming grammatical options to the same extent as native speakers do, so they benefit less than native speakers from statistical preemption.

In the following chapter, I briefly review several alternative proposals aimed at addressing the *explain-me-this* puzzle. In saving this discussion until the penultimate chapter, I do not intend to underestimate their value or their influence on the field. But we will see that the present usage-based constructionist approach is ultimately more comprehensive than alternative accounts.

Chapter 7

The Roads Not Taken

This chapter addresses several alternative proposals for how to account for the partial productivity of constructions. While there is merit to each suggestion, it is argued that the alternative proposals fall short of providing a full account. Some of the proposals were mentioned briefly in earlier chapters, but they are discussed more fully below.

7.1 Is Compatibility between Verb and Construction Enough?

Armed with a recognition of the sometimes subtle semantic, phonological, and information-structure constraints on argument structure constructions, it is tempting to conclude that the *explain-me-this* puzzle can be resolved by requiring simply that any predicate (verb or adjective) that is compatible with a construction can occur in that construction (e.g., Ambridge et al., 2009; Coppock, 2009; Fisher et al., 1994; Marantz, 1997). We know that speakers are sensitive to the subtle constraints associated with constructions, but, as described below, this compatibility-based proposal begs the question of how the subtle constraints themselves are learned; it therefore does not resolve the puzzle of the limits on productivity.

An important early compatibility-based account was provided by Pinker (1989), who suggested that children create narrowly defined semantic rules to establish whether a verb could be used in a particular argument structure. For example, change-of-state verbs such as *open*, *close*, and *break* constitute a class of verbs that undergo the English causative rule (*The door opened*; *She opened the door*), whereas verbs that denote intentional or "internally caused" actions such as "coming into or going out of existence" (e.g., *disappear*, *vanish*) or emotional expression (e.g., *laugh*, *giggle*) do not undergo the rule (*He laughed*; ?*She laughed him*). By five years old, children do in fact obey this constraint (i.e., children reject causatives of novel verbs meaning "disappear" or "laugh") (Ambridge et al., 2008, 2009). Pinker seemed to acknowledge that the narrowly defined classes had to be learned, since they differ across languages, and in fact younger children (age two to three) do not seem to show the same sensitivity to narrowly defined semantic classes, presumably because they have not yet fully learned them (Brooks and Tomasello, 1999). Pinker did not offer an account of how the narrow-range rules were themselves learned.

Moreover, this account proposed narrow-class *rules* (as opposed to statistical generalizations), which would seem to predict categorical judgments without systematic gradience (Pinker, 1999). However, Pinker had observed that across-languages actions that can be more directly caused are more likely to be used in the same construction as other verbs that convey direct causation (Pinker, 1989). And, in fact, English-speaking children and adults rate transitive uses of verbs of emotional expression—both familiar and novel (e.g., ?*The funny clown laughed/giggled/meeked Lisa*)—as less acceptable than familiar or novel transitive uses of verbs that mean "disappear" (e.g., ?*The magician disappeared/vanished/tammed Lisa*) (Ambridge et al., 2008; Goldberg, 1995, 131). This type of systematic gradience is problematic for an account in which verbs are understood to either undergo or fail to undergo a categorical rule.

According to the present usage-based constructionist account, subclasses that appear in particular constructions emerge from partially overlapping representations of structured exemplars in our hyper-dimensional conceptual space. Recall that the clustering is sensitive to semantics and information-structure properties of the contexts in which constructions are used, as well as the relative frequencies of and variability among the predicates occurring in a given argument structure construction (ASC). The present proposal therefore predicts that judgments will be gradient in systematic ways (chapters 4 and 5). Appeal to coverage also predicts a role for verb frequency in judgments of relative acceptability, whereas the rule-based compatibility proposal does not (see sections 5.7, 5.8, 7.3).

To summarize, simple compatibility proposals recognize that constructions have constraints, but they do not offer an explanation of where those constraints come from. On the other hand, the CENCE ME principles predict *how* formal, semantic, and information-structure constraints on constructions emerge. They also predict that judgments of acceptability will be gradient, and will depend on coverage (chapter 4) and whether there exist other conventional ways to express the intended message (chapter 5).

7.2 Are Invisible Features or Underlying Structure Explanatory?

A proposal that is similar in key respects to the compatibility-based suggestion involves positing invisible features or an "underlying level" of representation that is stipulated in order to constrain certain verbs or other predicates from appearing in certain constructions. This sort of solution has long been popular among theoretical linguists (Lakoff, 1966; Pesetsky, 1996; Yang, 2016). However, just as we saw was the case with the compatibility-based proposal, positing diacritics or underlying structure clearly leaves us with a version of our original question: How do children learn which predicates have the requisite underlying features or underlying structure? Since the range of predicates that occurs with particular constructions varies across languages (recall section 3.7), the invisible features or structure cannot be determined by meaning alone, and *must* be learnable from the available input. That is, positing invisible features fails to address the issue from a learner's perspective. In addition, the use of diacritics or underlying structure also fails to explain the systematic gradience in judgments just discussed, because features or underlying structures are assumed to either be present or not.

7.3 Conservatism via Entrenchment?

It is tempting to think that children learn, over time, to only use verbs in ways in which those verbs have been witnessed (Braine and Brooks, 1995). We briefly touched on this proposal, CONSERVATISM VIA ENTRENCHMENT, in sections 5.2 and 5.3. It predicts that children should become progressively more reluctant to extend familiar verbs in new ways as they witness those verbs more often in other constructions; that is, as verbs become more entrenched, they should be used more conservatively and judged as less acceptable in *any* new uses. The problem with this idea is that an utterance type may not yet have occurred simply because it conveys an unusual message. We do not want to predict that such an utterance will be judged unacceptable when an appropriate context presents itself. This issue is somewhat subtle, so it is worth discussion.

Conservatism via entrenchment can be difficult to disentangle from statistical preemption because both proposals predict that judgments should become more secure and errors should become more rare on the basis of additional exposure. Both proposals are part of more general accounts in which the vast majority of constraints on constructions emerge from how those constructions have been witnessed being used. Both proposals are usage based in that they assume that statistical knowledge of verbs' distributions is retained. Both proposals acknowledge that, all other things being equal, more familiar formulations are preferred over less familiar formulations and the more familiar, the better. We referred to this type of positive evidence of occurrence as SIMPLE ENTRENCHMENT (section 4.5). That is, simple entrenchment provides positive evidence of acceptability, and all usage-based approaches agree that it is important.

As reviewed below, certain work has used "entrenchment" as evidence for *un*acceptability—that is, as "negative evidence." I have referred to this proposal as "conservatism via entrenchment" to make clear that it differs from simple entrenchment. Conservatism via entrenchment predicts that a novel combination of verb and construction should be unacceptable if the verb is well entrenched in *any* other construction(s) in the language. Yet, contrary to this prediction, many highly frequent verbs can appear in constructions that are novel for those verbs. Recall the examples from chapter 5 (5.2), repeated below in (7.1):

7.1a "I . . . coughed him [a moth] out of my mouth."
7.1b "Sarah . . . winked her way through the debates."
7.1c "She'd smiled herself an upgrade."

Each of these verbs—*cough*, *wink*, and *smile*—is very frequent (entrenched) in the intransitive construction, and only exceedingly rarely, if ever, witnessed in the various transitive constructions in (7.1). The notion that speakers are conservative via entrenchment would therefore inappropriately rule out the sentences in (7.1) as unacceptable (Goldberg, 2006, 2011b).

The novel-construction-learning studies introduced in section 5.3 provided experimental evidence against the idea that speakers are simply increasingly conservative with each verb as its frequency is increased, without taking the functions of as-yet unwitnessed combinations of verb + ASC into account. Recall that in one series of novel-construction-learning studies, adults were only conservative with novel verbs when the two constructions being learned expressed the same message and so were in competition with each other; that is, witnessing a verb in one construction was essentially not witnessing it in the other construction (Perek and Goldberg, 2015; Wonnacott et al., 2008). On the other hand, when two constructions were assigned *distinct* functions, adults were quite willing to extend verbs for use in the alternative construction in order to better express their intended messages (Perek and Goldberg, 2015, 2017; Thothathiri and Rattinger, 2016). This difference between the two conditions is predicted by statistical preemption since it takes competition in context into account, but it is not clear how to explain it by appeal to conservatism via entrenchment. Moreover, if a verb was witnessed stubbornly occurring in one construction in contexts in which the other construction would have been more appropriate, learners did avoid using the verbs productively in whichever construction better expressed their message, and this is again predicted by statistical preemption (Perek and Goldberg, 2017).

Several labs have reported a certain finding that is consistent with both statistical preemption and conservatism via entrenchment. In particular, speakers preferred to extend *lower-frequency* verbs for use in ASCs that were novel for those verbs, over nearly synonymous higher-frequency verbs. For example, Brooks et al. (1999) performed a production study with children aged three to eight. The experimenter demonstrated actions that could be named by two verbs: a more familiar (high-frequency) verb, or a near synonym with lower frequency (e.g., *disappear* vs. *vanish*). Different groups of children witnessed one of the verbs of each pair first modeled in its familiar ASC (e.g., *The box disappeared*; *The box vanished*). Subsequently, children were asked, "What did she do to the box?"—a question intended to encourage responses like ?*She disappeared it*, or ?*She vanished it*. Results demonstrated that, across all ages, children were more likely to overextend lower-frequency verbs than higher-frequency verbs; that is, they were more likely to produce *She vanished it* than *She disappeared it*. Similarly, Theakston (2004) performed a judgment task with children (ages five and eight) and adults on 12 pairs of sentences involving lower- and higher-frequency verb pairs. Again, the critical trials involved verbs used with ASCs in which they did not normally appear. The data revealed that, across all ages, sentences containing lower-frequency verbs tended to be judged as more acceptable than sentences containing higher-frequency verbs. For example, *drip* is less frequent than *spill*, and ?*I dripped the carpet with juice* was judged as more acceptable than ?*I spilled the carpet with juice*. Ambridge et al. (2008) also performed a judgment task on three pairs of novel sentences involving intransitive verbs used transitively (*fall/tumble*; *disappear/vanish*; *laugh/giggle*). They, too, found an advantage for lower-frequency over higher-frequency verbs and an even greater advantage for novel verbs (which have a prior frequency of zero).

Notice that in these experiments, the higher-frequency verbs are more frequent overall, as is relevant for conservatism via entrenchment. But the preempting alternative formulations are *also* more frequent, as is relevant for statistical preemption. That is, *disappear* is more frequent overall than *vanish*, and the preempting form, *make <something> disappear* is also more frequent than *make <something> vanish*. In fact, a search of the COCA corpus confirms that *make <something> disappear* is more frequent than *make <something> vanish* by a factor of 10. Thus, this evidence does not distinguish conservatism via entrenchment from statistical preemption. As already discussed in section 5.8, confidence of preemption increases as verbs are witnessed more often in the competing construction. Moreover, lower-frequency verbs are only judged to be more acceptable than higher-frequency verbs when used in a new ASC for that verb when there is a competing (preempting) ASC for that verb (Robenalt and Goldberg, 2015). This is predicted by statistical preemption, but not conservatism via entrenchment.

Nonetheless, several papers have argued that conservatism via entrenchment predicts correlations between corpus data and acceptability judgments better than statistical preemption (Ambridge, 2013; Ambridge et al., 2012a, 2015; Stefanowitsch, 2008), so it is worth taking a closer look. The earliest paper to my knowledge that argued for this was Stefanowitsch (2008), who referred to a version of conservatism via entrenchment as NEGATIVE ENTRENCHMENT. He suggested that learners essentially compare how often a particular verb has been witnessed occurring in a particular construction with the verb's overall frequency, and then compared that verb's distribution to the average distribution for *all* other verbs (with significance determined by Fisher's exact test or chi-square). This calculation was offered as a way to determine whether a verb was biased toward or away from a particular construction (see Ambridge et al. [2018] and Stefanowitsch and Gries [2003] for related analyses). Stefanowitsch provides the following example to explain why *say* is ill formed in the double-object construction (?*Say him something*):

> The ditransitive [double-object] construction occurs 1824 times and the verb *say* occurs 3333 times. The total number of verbs in the ICE-GB [corpus] is 136,551. If there were no particular relationship between *say* and the ditransitive construction, we would expect the combination to occur 44.5 times in the corpus . . . [but it actually occurs 0 times, which is significantly less often]. (Stefanowitsch, 2008, 520)

Thus, the proposal appeals to base rates of verbs and constructions without attention to context or semantics. As already noted, it predicts that all of the expressions in (7.1) should be unacceptable, given that both the verbs and the constructions involved are reasonably frequent while the combinations of those verbs in those constructions are quite rare. Work reviewed in section 5.7 by Robenalt and Goldberg (2015), however, found that while speakers do recognize that these sorts of sentences are novel, and therefore do not find them as acceptable as familiar formulations, verb frequency does not appear to negatively influence their acceptability and they are more acceptable than other sentences that have a readily available competing alternative. This is what statistical preemption, but not conservatism via entrenchment (negative entrenchment), predicts.

A number of studies by Ben Ambridge and colleagues have compared conservatism via entrenchment and statistical preemption in a similar way, by correlating frequencies with judgment data. This body of work is extensive, and results across the studies have been inconsistent: some studies reported that conservatism via entrenchment was important and statistical preemption was not (Ambridge, 2013; Ambridge et al., 2012b, 2015), while other, similarly designed studies, found that competition between ASCs plays an essential role as statistical preemption predicts (Ambridge, Pine, et al., 2014; Blything et al., 2014). Relevantly, the studies that failed to find an effect of statistical preemption had predicted that only verbs that *never* appeared in a given ASC (within a particular corpus) should be dispreferred in that ASC.

However, if a corpus is too small, certain verb + ASC combinations may be absent owing to chance, as small numbers are unreliable. At the same time, if the corpus is sufficiently large, examples that are dispreferred may occur occasionally (Fellbaum, 2005). The rigid requirement that there be zero occurrences of a verb in a particular construction therefore artificially defines the class of restricted verbs either too narrowly or too broadly, depending on whether the corpus is large or too small. Moreover, by comparing a verb's occurrence in two ASCs, one of which the verb never occurred in, the probability of the verb occurring in the alternative, preempting construction was artificially fixed at 1, which could not be expected to predict gradient judgment data.

New work by Ambridge et al. (2018) systematically reviews several of the conflicting results reported in earlier studies. This work employs more corpus data for the locative construction, as well more-up-to-date statistical analyses. Armed with the improved techniques, the idea that ASCs must compete in order to affect one another is made clear. In a revised analysis of the locative alternation, they find that "preemption displayed a clear effect above and beyond entrenchment (and all other predictors), reversing the null finding observed in the original study [Ambridge et al. 2012a]" (Ambridge et al. 2018).

Ambridge et al. (2018) include an analysis of the reversative *un-* prefix as well as verb + ASC alternations. Notice that acceptability judgments of verbs with the reversative prefix *un-* clearly vary: *undo* is much more acceptable than ?*unsqueeze*. Familiarity (our "simple entrenchment") increases acceptability judgments: the fact that *undo* is highly frequent undoubtedly boosts its acceptability. Semantics, too, plays a key role in judgments, as Ambridge and colleagues' (2018) results bear out: events that are easier to imagine being reversed are more acceptable than those that are not (*untie* is more acceptable than *uncut*). In addition, statistical preemption predicts that new combinations of verbs that include an *un-* prefix should be less acceptable when there already exists a different verb that expresses the intended meaning, and the authors find evidence for this as well; for example, *unlower* is judged to be particularly unacceptable because *raise* can be used instead.

Conservatism via entrenchment predicts, over and above these effects, that there should be "a negative correlation between acceptability . . . and overall verb frequency (e.g., all uses of *squeeze*, without the prefix *un-*)" (ibid., 27). And, in fact, this effect is reported: verbs prefixed with *un-* were judged less acceptable as the frequency of the bare form of the verb increased. The authors acknowledge that this is not a very intuitive finding, so it is worth taking a closer look. We know that several factors are involved, so in a new study, Goldberg et al. (unpublished MS) aimed to control for semantics, statistical preemption, and positive entrenchment, in a judgment task involving verbs with the *un-* prefix. To control for semantics, we chose pairs of higher- and lower-frequency verbs that were nearly synonymous in the sentence contexts provided to participants. As already reviewed, statistical preemption can be hard to tease apart from conservatism via entrenchment in the case of verb + ASCs, but it is more straightforward in the case of morphology. For example, any verb that could preempt a higher-frequency *un-* verb should also be available to preempt its lower-frequency counterpart, since, again, the two were designed to be nearly synonymous in the sentence contexts provided. In order to control for (simple) entrenchment, we selected pairs of *un-* verbs that did not occur even once in the 520-million-word COCA corpus, so that positive evidence for them approached zero. This was possible to do by using verbs that referred to actions that are not easily reversed (e.g., *unfire/undischarge*; *unsuggest/unpropose*), and then creating special sentence contexts that made the idea or desire to reverse each type of action semantically plausible. Sample stimuli are provided in (7.2 and 7.3):

7.2 When he had to face his victim's family, the regretful soldier sorely wished he could <unfire/undischarge> his weapon.
7.3 He realized as soon as he said it that it might backfire, but he couldn't <unsuggest/unpropose> the plan at that point.

Conservatism via entrenchment makes a clear prediction: the sentences with the lower-frequency verbs (prefixed with *un-*) should be judged more acceptable than the higher-frequency alternatives. For instance example (7.2) with *undischarge* should be more acceptable than with *unfire*, since *discharge* is less frequent (less entrenched) than *fire*. However, results showed an effect in the *opposite* direction: the sentences that contained the lower-frequency verbs were judged *less* acceptable than those with higher-frequency verbs. This finding runs directly counter to what conservatism via entrenchment predicts.

To summarize, there is broad agreement about simple entrenchment (section 4.5): speakers prefer familiar formulations, so the acceptability of frequently witnessed formulations is expected to be higher than the acceptability of less frequent formulations. The key difference between conservatism via entrenchment and statistical preemption is whether *competition in context* is necessarily taken into account as well, as statistical preemption presumes.

In many cases, statistical preemption and conservatism via entrenchment are hard to tease apart, particularly when the situations described by novel verb + ASC combinations occur regularly. In this case, failing to witness the novel verb + ASC combination correlates strongly with witnessing some other (competing) formulation. However, it is possible to determine whether speakers take the absence of occurrence as evidence for unacceptability by considering descriptions of implausible, unusual, or simply previously unwitnessed types of events. Those cases seem to undermine the predictions of conservatism via entrenchment. Relevant evidence comes from examples like those in (7.1), experimental work discussed in section 5.3, the judgment study of Robenalt and Goldberg (2015), and the new study just reviewed involving *un-* prefixed verbs. It remains possible that *both* competing and noncompeting uses of verbs matter, as argued in the review work of Ambridge et al. (2018). But there is at least widespread agreement now among usage-based researchers that statistical preemption—or competition in context—is operative.

7.4 Are "Tolerance" and "Sufficiency" Numbers Explanatory?

In *The Price of Linguistic Productivity* (*PoLP*), Charles Yang (2016) begins by assuming that our knowledge of language involves a situation very close to the traditional dichotomy between productive rules and irregular exceptions listed in memory (his "irregular rules"). Language learners are claimed to rifle through exceptions serially, in a list ordered by their relative frequency during production. For example, in order to produce the past tense of *freeze*, learners are claimed to search the ordered list: *was*, *did*, *said*, *made*, *went*, *took*, *came*, *saw*, *knew*, *got*, *gave*, . . . (90 more verbs) . . . *FROZE*! The motivation provided for this idea is the fact that people have been found to respond faster to higher-frequency words than to lower-frequency words in experimental lexical decision tasks ("is this a word or not?") (Zwitserlood, 1989). However, we know that *many* factors play a role in lexical decision tasks and in lexical access more generally. Importantly, these factors include semantic, morphological, and phonological relatedness to other items already evoked in the context, as well as word length, age of acquisition, and emotional content (Brown and Watson, 1987; Kousta et al., 2009).

In fact, all sides of the long-standing debate about whether grammar involves categorical rules have acknowledged that exceptional cases must be represented within an associative memory. That is, we do not learn *lists* of exceptions, but we instead learn richly interconnected networks (Ackerman et al., 2011; Jackendoff and Audring, 2016; Blevins et al., 2016; Booij, 2010; Burzio, 2002; Daugherty and Seidenberg, 1992; McClelland and Patterson, 2002; Pinker and Prince, 1988; Ullman, 2001). Decades of research in psycholinguistics and cognitive psychology more generally have emphasized that memory is importantly associative (for reviews, see Anderson and Bower, 2014; Hinton and Anderson, 2014). Moreover, our memory is content addressable in that we are able to access words based on the content of their representations. The USAGE-BASED CONSTRUCTIONIST approach captures this idea by relating representations along various dimensions within our hyper-dimensional conceptual space.

The fact that exceptions are represented within an associative memory predicts the existence of subregularities and "neighborhood" effects, which have long been recognized to exist (Burzio, 2002; Lakoff ,1970; Marchman, 1997; Regel et al., 2015). In particular, "exceptional" cases can give rise to productive generalizations of their own. For example, the English plural marker /-s/ is a productive generalization that applies to words with semantically plural meanings.

And yet there also exist clusters of exceptional cases that are productive. For example, we have the names of fish that are unmarked for plural (*salmon*, *halibut*, *trout*), and we also have names for legwear that are marked as plural even though they are semantically singular (*pants*, *capris*, *shorts*, *culottes*) (Williams, 1994). These are both productive subregularities. If we find a new type of fish (e.g., the *twilight zone groppo*), or a new type of legwear (*peggings*), we are likely to treat them as we do other members of these irregular subclusters (see also section 5.8; Bybee and Moder, 1983; Baayen and del Prado Martin, 2005; Marchman, 1997; Pinker, 1999).

Putting our associative memory to one side, it is worth considering the proposals in *PoLP* on their own terms. Yang's primary goal is to offer quantifiable boundary conditions between productive and nonproductive "rules." In particular, *PoLP* argues that a fully productive rule is created for the purpose of computational efficiency if and only if no more than a maximum number of exceptional cases have been witnessed *and* a sufficient number of rule-following cases have been witnessed. More specifically, *PoLP* assumes that learners keep a tally of the three numbers N, M, and e:

7.4a N = the number of types a rule potentially applies to (i.e., the size of the set of words within the domain of a potential rule).
7.4b M = the number of types (words) witnessed following a rule.
7.4c e = the number of exceptions to a rule.

A "Tolerance Principle" is claimed to provide an upper bound on the number of exceptions (e) tolerated by a productive rule:

Tolerance Principle
7.5 $e \leqslant N/\ln N$. (Yang, 2016, 64)

That is, the number of exceptional cases (e) can be no larger than the number of cases to which the rule can potentially apply over the natural log of that number ($N/\ln N$). *PoLP* does not explain how the number of potential rule-following cases, N, is to be determined. For example, there is a productive generalization in English that adds a comparative morphological suffix, -*er*, to short gradable adjectives. But speakers do not apply the -*er* ending to adjectival past participles, (e.g., ?*benter*), even though they may be short (*bent* is one syllable) and semantically gradable (e.g., things can be more or less bent). Should we amend the comparative -*er* rule so that it does not apply to adjectival past participles or should we count past participles as exceptional? And if we incorporate a restriction on past participles, what prevents us from including a very targeted additional restriction that prevents the rule from applying to *good* (?*gooder*)?

It becomes easy to see that what counts as being within the domain of a rule or an exception depends entirely on how the rule is stated.

More generally, the proposal does not address the nature, origin, or formulation of rules. Because *PoLP* assumes that any items that are retained are retained in lists instead of associative clusters, it offers no mechanism by which the domains of rules can be determined or even hypothesized. That is, without allowing for relationships among learned cases within an associative memory, it is unclear how children could learn that *-er* only applies to shorter adjectives or even that it applies to adjectives and not adverbs.

Surprisingly, with its emphasis on the maximum number of exceptions to a rule (e), *PoLP* does not address how it is that learners are able to identify exceptions as being exceptional. Chapter 5 of the current volume has emphasized the role of statistical preemption, but Yang is explicitly dismissive of such evidence, stating: "Indirect negative evidence . . . is to be avoided" (2016, 212). He instead argues that learners only know when a case follows a rule (see section 7.7; ibid., 6.1.3).

Another critical problem for the *PoLP* proposal is that it requires learners to keep a tally of rule-following cases (M). That is, a "Sufficiency Principle" requires that the number of witnessed rule-following cases must be sufficiently high, relative to the number of potential cases to which a rule can, in principle, apply (N). The Sufficiency Principle states, "Let R be a generalization over N items, of which M items are attested to follow R. R can be extended to all N items if and only iff [*sic*]:"

Sufficiency Principle

7.6 $N - M \leqslant N/\ln N.$ (ibid., 177)

Yang clarifies, "Before the positive evidence is sufficient—when M sits below the sufficiency threshold—learners lexicalize all M items and does [*sic*] not generalize beyond them" (ibid., 177). In fact, *PoLP* assumes that learners must retain the vast majority of rule-following cases. That is, "productivity only obtains when the rule-following items thoroughly overwhelm the exceptions" (ibid., 122). To be concrete, consider table 7.1, which specifies M, N, and e for rules with domains of various sizes. Notice that the prefix *pre-* can be added to any proper name or noun that can evoke a temporal onset (*pre–Harry Potter*, *pre-Carolyn*, *pre-van*). By the time a child recognizes *pre-* as a rule, she has perhaps witnessed 1000 proper names or nouns to which *pre-* could potentially be applied. The last row in table 7.1 predicts that a child must therefore witness at least *855* distinct names or nouns being used with the *pre-* prefix before she recognizes it is productive. This enormous number must be witnessed even though there are no exceptions, because the Sufficiency Principle does not

TABLE 7.1. Minimum M and maximum e allowed for a rule to be productive[a]

Number of cases a rule can potentially apply to; domain size (N)	Minimum number of cases that must be witnessed following the rule (M)[b]	Maximum number of exceptions to the rule (e)[c]
6	3	3
8	4	4
9	5	4
10	6	4
100	78	22
300	247	53
1000	855	145

[a] As predicted by the Sufficiency and Tolerance principles for rules of varying domain sizes.
[b] $M = N - N/\ln N$.
[c] $e = N/\ln N$; e is rounded to the nearest whole number.

make any reference to the number of exceptions. Surely this sets the cost of productivity unnecessarily high.

In fact, *PoLP* allows that learners may mistakenly create productive rules, leading them to make overgeneralization errors, and that they then "can still backtrack to lexicalization if the amount of positive evidence drops below the Sufficiency threshold" (ibid., 213). Notice that this requires that *all rule-following instances must be retained throughout language learning*, since M is assumed to increase as additional rule-following cases are witnessed. This aspect of the proposal undermines the stated reason for learners to create a productive rule in the first place: namely "for a more efficient organization of language, as measured in processing time, *rather than listing everything in lexical storage*" (ibid., 9).

PoLP offers an experimental study by Schuler et al. (2016) to support its claims. Those authors found that children generalized a pattern—creating a productive rule—when five cases were witnessed obeying the generalization and four cases were exceptions; moreover, children did not generalize the pattern when only three cases followed the rule and six were exceptions. But note that greater productivity in the former than the latter case is predicted by *any* theory of induction. A generalization is bound to be more robust if the ratio of cases that follow the pattern to cases that are exceptional is 5:4 rather than 3:6, other things being equal (recall figure 4.3). The Sufficiency Principle makes a much stronger prediction that was not tested. Because the number of cases that must be witnessed following a rule (M) is independent of the number of witnessed exceptions (e), children should categorically avoid generalizing a rule that potentially applies to nine cases, even if they witness four out of nine potential cases following the rule *and zero exceptions.*

Moreover, whenever e approaches its maximum, as is the case in many of the examples analyzed in *PoLP*, the proposal predicts that *all* regular (rule-following) cases *must* be learned one by one. To see this, consider the following simple argument:

7.7a Assume a rule, R, is productive.
7.7b $N \leqslant N/\ln N + M$. Sufficiency Principle (add M to both sides of [7.6])

If the maximum allowable number of exceptions has been witnessed ($e = N/\ln N$), then:

7.7c $N \leqslant e + M$. (Substitution of $e = N/\ln N$ to [7.7b])

The number of cases *potentially* following a rule (N) must be greater than or equal to the number of witnessed exceptions (e) plus the number of witnessed rule-following cases (M):

7.7d $N \geq e + M$.	Definition of N
7.7e $N = e + M$.	(7.7c) and (7.7d)
7.7f $M = N - e$.	(The number of learned rule-following cases equals the number of *all potential cases* minus the learned exceptions.)

Thus, if the maximum number of exceptions has been witnessed, in order for a rule to be productive, all cases that potentially undergo a rule have to be learned and retained individually. Surely it is not newsworthy that a generalization can apply to exactly the set of cases the learner has already learned that it applies to.

To summarize, the *PoLP* proposal assumes that learners keep a tally of potential cases that may follow a rule (N), the number of witnessed rule-following cases (M), and the number of exceptions (e). But the proposal does not address how N nor e is to be determined. In particular, by rejecting the role of associative memory and clustering algorithms, and instead assuming that item-based information is represented in lists, *PoLP* offers no proposal for how the domain of generalizations (or "rules") is to be determined, which is required for calculating N. Moreover, since statistical preemption and other forms of indirect negative evidence are rejected, *PoLP* fails to explain how learners identify exceptional cases in order to count them. Finally, the Sufficiency Principle claims that in order for a rule to be productive, the vast majority of regular cases must be witnessed and retained on an ongoing basis (in order to calculate *M*), directly undermining its stated goal of avoiding the retention of regular cases.

7.5 Are Frequencies without Function Effective?

Productivity and Reuse in Language (*PaRL*) by Tim O'Donnell (2015) addresses the issue of partial productivity from a computational perspective. The book offers a detailed tour and review of a number of approaches to morphology, making the work relevant to computational linguists as well as psycholinguists. O'Donnell acknowledges, in agreement with the present volume and a growing body of work, that words and grammatical patterns are subject to the same principles of productivity (see also Di Sciullo and Williams, 1987; Goldberg, 2006; Langacker, 1987; O'Donnell, 2015; Yang, 2016).

PaRL argues that productivity increases as the number of attested types that instantiate a pattern increases. For example, three English suffixes combine with adjectives to form nouns and they have varying degrees of productivity, from the most productive suffix, *-ness* (e.g., *crumbliness, dingbattiness*), to the much less productive *-ity* (*security*), to the unproductive suffix *-th* (*warmth*). Speakers are presumed to assign various degrees of productivity by witnessing how many different words each suffix has applied to. This is an appeal to type frequency: generalizations that occur with higher type frequency are more productive. Indeed, a construction can be identified as productive when it appears with a large number of distinct and infrequent types (Baayen and Lieber, 1991), including those that appear only once in a corpus. These cases have been referred to as *hapax legomenon*, from Greek "said (only once)." Although this idea has been appealed to predominantly within morphology, it can just as well be applied to phrasal patterns in order to capture when a construction is productive (see Barðdal, 2008; O'Donnell, 2015; Desagulier, 2015; Zeldes, 2013).

At the same time, while recognizing that a construction has appeared with a large number of infrequent verbs is a good way to demonstrate that the construction is productive, it is not a suitable way to determine whether a particular new coinage is acceptable. That is, even highly productive constructions have constraints on *which* new kinds of cases are acceptable. Thus, the *hapax*-based indication of productivity is *a result* of a construction being relatively productive, which is itself determined on a case-by-case basis by coverage and competition. By appealing to numerical counters without regard to how instances cluster within our conceptual space, we forgo the opportunity to account for the content and range of generalizations. For example, the fact that *-ness*, *-ity*, and *-th* combine with adjectives to form nouns must be stipulated. On the other hand, the CENCE ME principles allow the range of a generalization to emerge from the clustering of witnessed instances, and this is required to predict which productive instances will be acceptable. In fact, *variability* among attested instances and *similarity* of the potential coinage, along with type frequency, determine the extent to which attested cases *cover* the relevant hyper-dimensional conceptual space needed for a new coinage (chapter 4).

Variability and similarity are obviously more difficult to formalize than frequencies since they require appeal to semantics, information structure, and phonology. It is a daunting task to try to adequately formalize these varied and complex dimensions (recall chapter 3). But if we instead ignore these key dimensions, we cannot predict how we learn to use language in the ways that we do (for acknowledgement and discussion of this point, see O'Donnell [2015, 34–36]).

7.6 Are Storage and Productivity Inversely Related?

Many approaches to productivity in language assume that computation is called into service in order to avoid storage in memory. That is, it is often assumed that memory and computation stand in an inverse relationship for the sake of efficiency (e.g., O'Donnell, 2015; Yang, 2016). The usage-based constructionist approach takes a quite different perspective. Partially abstracted from experience, exemplars are retained in memory as part of a rich network of knowledge. While we are not able to *recall* individual exemplars at will, given that their representations overlap with the representations of other exemplars, our knowledge of language is formed and continually affected by them. Language is extended creatively (involving new "computations"), not in order to reduce or avoid storage in memory, but in order to express new messages in ever-changing contexts.

Importantly, the content of various types of generalizations is determined by witnessed exemplars, but not in a "blind" fashion. Instead, the relevance of particular dimensions is informed by attention and by prior learning. That is, as coverage requires, learners can and do determine which dimensions are relevant to each construction in their language. The comparative ending *-er* is conditioned by morpho-phonology (e.g., *prettier* vs. ?*beautifuller*), while the prefix *pre-* is not (e.g., *pre-dawn*, *pre-pneumonia*). Whether verbs sound Latinate or Germanic is relevant to their acceptability in the English double-object construction, but this same dimension is not relevant to the simple transitive construction (e.g., *She explained/donated/returned it*).

While we have studiously steered clear of the well-worn domain of the English past tense and other cases of inflectional morphology, the issue of whether productive "rules" replace the need for storage, or instead *emerge on the basis of storage*, has a rich history in this domain (see O'Donnell [2015] for a recent review). As acknowledged directly or indirectly by most researchers, the productivity of exceptional or irregular instances is naturally accounted for by applying *coverage* to clusters of cases: that is, irregulars are productive to the extent that there exists a well-covered cluster of similar exemplars. For instance, the irregular past tense may be applied to a novel verb *spling* to form the new past tense *splang* because there exists a cluster of cases that are irregular in just this way (*spring*/*sprang*, *stink*/*stank*, *swim*/*swam*, *sing*/*sang*, *sink*/*sank*)(e.g., Booij, 2017; Harmon and Kapatsinski, 2017; but cf. Yang, 2016). The debate has centered around the issue of how best to account for productive generalizations, or "rules."

Do regular cases require a different sort of computation that is independent of memory of previously witnessed exemplars? In appealing to a linguistic RULE, as distinct from a generalization, the following claims have been made explicitly or implied (Pinker, 1999):

1. Rules become productive suddenly.
2. Instances generated by rules are not retained in memory.
3. Rules apply categorically, so regular cases should not be affected by frequency, similarity, or neighborhoods of related cases.

The claim in (1) that learners experience a "tipping point" (Yang, 2016, 41) or "eureka moment" (Pinker, 1999, 202) at which point a categorical, productive rule is suddenly created has been challenged by more detailed analyses of the data (McClelland and Patterson, 2002). The idea that instances that are generated by rules cannot be retained in memory (2) has also fallen away (e.g., Baayen and Schreuder, 2006). In fact, before learners are able to identify a "rule," all theories recognize that a good number of regular cases must be retained. Moreover, there exist certain regular forms (e.g., *blinked*) that also must be stored because they fall within a class of productive exceptions. That is, *blink* shares key phonological similarities with the nonsense verb *spling*, and yet while speakers spontaneously suggest *splang* or *splung* for the past tense of *spling*, we recognize that the past tense of *blink* is *blinked*. It follows that regular *blinked* must be retained in memory (Pinker and Ullman, 2002).[1]

The most controversial issue, by far, has been (3): do frequencies, semantic similarity, and variability, which are recognized to affect irregulars, also affect regular cases (productive "rules")? While early findings suggested that they may not, subsequent work has provided evidence that regular forms *are* affected by phonological neighborhoods and semantics. For example, Albright and Hayes (2003) observed that virtually all verbs that end in an "f" sound (e.g., *laugh*, *cough*, *fluff*) take the regular past tense (*laughed*, *coughed*, *fluffed*), forming an

1 For modeling work that is capable of productively extending the 12-cell inflectional paradigm required for nouns in Russian on the basis of witnessed exemplars, see Janda and Tyers (forthcoming).

"island of reliability" (a well-covered area with no competitors). In a judgment task, they found higher acceptability ratings for novel verbs with this ending (e.g., *baff-ed*, *louf-ed*) than for novel verbs ending in consonants that were not part of an island of reliability (e.g., *kow-ed*, *lamm-ed*) (see also McClelland and Patterson, 2002; cf. Ullman, 2001). Likewise, Ramscar (2002) found semantic similarity at times plays a role. In particular, he found that a novel verb, *frink*, was more likely to be treated as a regular verb (*frinked*) if speakers were told its meaning was related to "blink" or "mediate," both of which take the regular past tense form. On the other hand, if speakers were told that *frink*'s meaning was related to "drink," *frink* was much more likely to be assigned an irregular past tense (*frank* or *frunk*).

It turns out that the notion of coverage from chapter 4 predicts the fact that regular generalizations are less sensitive to semantics or phonology than smaller classes of irregulars. When we say that a morphological generalization is "regular" we often *mean* that it applies generally to all instances within its domain without regard to details of semantics or phonology. The necessary key idea is that coverage determines the *content* of generalizations. In the case of the English regular past tense, for example, the phonological and semantic space of regular past tense forms is well covered over a broad area that is not strongly conditioned by phonological dimensions (beyond the /-d/ suffix itself), nor semantic dimensions (beyond the past tense meaning).

7.7 Preempted Forms Need Not Be Created

Embick and Marantz (2008) note that descriptions such as "A blocks B" and "A is in competition with B" seem to presuppose that both A and B exist. They argue that there is no evidence that irregulars such as ?*bended* or ?*intelligenter* are necessarily created by speakers (cf. *bent*, *more intelligent*). In fact, I agree. It would be very odd to suggest that unacceptable forms such as ?*bended* or ?*intelligenter* exist in the long-term memory of adult speakers only to consistently lose in a competition to their acceptable counterparts, *bent* and *more intelligent*. However, there is no need to assume this. The term "preemption" makes no assumption that the unacceptable "preempted" form has to be created or retained. Saying "*bent* preempts *bended*" does not presuppose that ?*bended* must exist. Instead, the familiar form, *bent*, is used simply because *it* is readily available and captures the intended meaning. *Any* other form (e.g., ?*bended*; ?*bond*; ?*bund*) will therefore be dispreferred.

In Embick and Marantz's (2008) framework, Distributed Morphology (DM), multimorphemic words are created on the fly by productive syntactic rules that combine morphemes, and these authors suggest that the effects of preemption can be captured by restricting the domain of certain syntactic rules. For example, in the case of the English past tense, instead of allowing the irregular past tense *hit* to block or preempt ?*hitted*, they suggest the rules in (7.8) (see also Albright and Hayes, 2003).

Tense[past][2]
Irregular
7.8a -t {Root (LEAVE); root (BEND); . . .}
7.8b -0 {Root (HIT); Root (QUIT); . . .}
Regular
7.8c -ed

This aspect of the proposal is entirely stipulative and it begs the question of how these rules are learned. That is, the DM proposal does not attempt to do what the present proposal does: it provides no account of how people *learn* either exceptions or default rules. Beyond this issue, the DM proposal faces other challenges, discussed elsewhere (Goldberg, 2006, 212–13; Williams, 2007). For example, since even "derivational" morphology is claimed to be added on the fly by rules, the DM framework is unable to account for the noncompositional meaning associated with many complex words—e.g., *derivational*, *impressionable*.

Nonetheless, Embick and Marantz (2008) make a valuable point that should be borne in mind: preempted forms do not necessarily exist in the mind of the learner or speaker. Instead, simply put, the form *ran* preempts ?*runned* because *ran* has been used reliably in contexts which would otherwise have been suitable for ?*runned.*

7.8 Witnessing Enough Data

Recall from section 5.3, there exists a class of English *a*-adjectives that resist appearing attributively as prenominal modifiers (e.g., ?*the afraid boy*, ?*the asleep man*). As reviewed in chapter 5, Boyd and Goldberg (2011) offered experimental evidence that restrictions on new (novel) adjectives are learnable on the basis of a combination of coverage and statistical preemption. It is reasonable to ask whether enough evidence for statistical preemption is available to children, and in fact Yang (2016) has argued that there is not. His argument rested on an analysis of a 4.3-million-word corpus of speech directed to three-year-olds, which found that there was some but not enough evidence of *a*-adjectives being used in preemptive contexts where a prenominal use might otherwise have

2 On the DM proposal it is technically not the full word—e.g., *hit*—that blocks a different full word, ?*hitted*, but it is instead the past tense morpheme *-0* that blocks the default rule. According to Embick and Marantz (2008), blocking doesn't exist at the level of words or phrases.

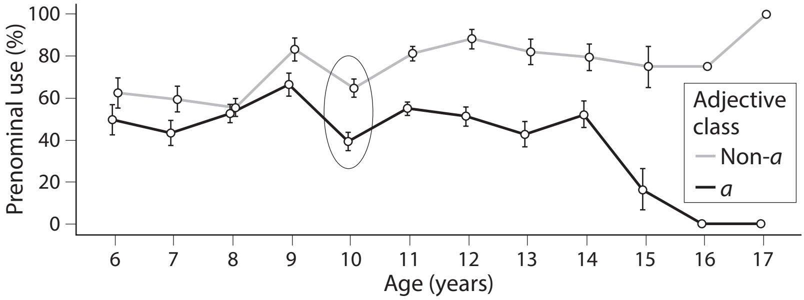


FIGURE 7.1. Percentage of prenominal uses for *a*-adjectives (*black line*) and semantically related non-*a*-adjectives (*gray line*) produced by children aged 6–17. Results based on Hao (2015) for the production task used in Boyd and Goldberg (2011).

been appropriate. One could take issue with the sorts of expressions that were assumed to be required for preemption to work (see Goldberg and Boyd, 2015), but in the present context, there is a more important point to make.

I concur, with Yang, that three-year-olds have not witnessed or absorbed sufficient evidence to reliably constrain their use of *a*-adjectives. Statistical preemption is expected to be a gradual process that can take years, since it requires that learners be able to recognize when one form might have been appropriate but another was consistently witnessed. In fact, it turns out that the restriction on *a*-adjectives is learned quite late. In her senior thesis at Princeton, Hao (2015) ran a version of Boyd and Goldberg's (2011) experimental design with children who ranged in age from 6 to 17 years. Her results demonstrated that children younger than 10 did *not* avoid using *a*-adjectives (*asleep*, *alive*, *afloat*, and *afraid*) prenominally less than other adjectives with closely related meanings (*floating*, *frightened*, *living*, and *sleepy*). Only when children are roughly 10 years old did they begin to reliably treat *a*-adjectives differently than near synonyms (see figure 7.1). Thus, children require much more language input than the amount that Yang (2016) had considered, which was roughly one year's worth. Children's appreciation of individual predicates' distributions grows over time, and they do not reach adult-like performance in the case of *a*-adjectives until they are roughly 10 years old.

7.9 Summary

Several alternative proposals for accounting for the partial productivity of morphology and grammatical constructions have been reviewed in this chapter. Each proposal has something to offer, but we have seen that each proposal either fails to address how generalizations (or "rules") are learned and constrained, or fails to allow for the creative use of language.

Whenever constructions are combined, their constraints must be compatible (section 7.1) but this does not directly address the *explain-me-this* puzzle since certain expressions are easily interpretable and yet are systematically dispreferred by native speakers (e.g., ?*Explain me this*). Simply attaching diacritics to stipulate that certain words may not occur in certain constructions begs the question of how children come to learn which diacritics to assign to which words (section 7.2). The idea that learners stray minimally from what they have witnessed has merit, but we have reviewed evidence that novel combinations of verb + ASC are reasonably acceptable when the combination expresses a novel message (section 7.3). Appeals to counting the number of witnessed cases or the number of exceptions (sections 7.4 and 7.5) without regard to the relationships among witnessed cases cannot adequately address how it is that children learn the generalizations (the content of the "rules"). Moreover, a close look at suggested "Tolerance" and "Sufficiency" principles reveals that they presume that virtually all rule-following cases and all exceptions are simply listed (section 7.4). More generally, attention to frequencies without recognizing the important role of clustering is insufficient to address the *explain-me-this* puzzle (section 7.5).

The impulse that many linguists share to minimize storage by increasing computation is ill conceived (section 7.6). Our memory for language is quite vast, as discussed in section 2.2 for words and in section 4.2 for grammatical constructions. It is computation that is expensive for our brains, not memory. Moreover, it was argued (particularly in chapter 4; also, section 7.6) that the existence of a well-covered (dense) cluster in memory facilitates the generation of new exemplars, which in turn increases the density of the cluster. That is, storage begets generalization and generalization begets storage: the two are *positively* correlated.

The current usage-based perspective anticipates that learners must witness and use language in a wide array of contexts in order to become highly skilled language users (section 7.8; see also chapter 6). The available evidence is consistent with the idea that generalizations, subregularities, and exceptions are learned on the basis of coverage and competition, in accord with the CENCE ME principles.

Chapter 8

Where We Are and What Lies Ahead

How and why do languages contain learned pairings of form and function that can be used in creative but constrained ways? We have provided support for each of the following CENCE ME principles that underlie the usage-based constructionist approach to language advocated here:[1]

A. Speakers balance the need to be **E**xpressive and **E**fficient while conforming to the conventions of their speech communities (chapter 1).
B. Our **M**emory is vast but not perfect: memory traces are retained but partially abstract ("lossy") (chapters 2, 3, and 4).
C. Lossy memories are aligned when they share relevant aspects of form and function, resulting in overlapping, emergent clusters of representations: **C**onstructions (chapters 3, 4, and 6)
D. **N**ew information is related to old information, resulting in a constructicon: a rich network of constructions (chapter 4).
E. During production, multiple constructions are activated and **C**ompete with one another to express our intended message (chapter 5).
F. During comprehension, mismatches between what is expected and what is witnessed fine-tune our network of learned constructions via **E**rror-driven learning (chapters 5 and 6)

We saw (in chapter 2) that words evoke semantically rich, structured meanings, partially abstracted from contexts of use. An appreciation of word meanings requires that our memory for how words are used must be vast, albeit partially abstracted from experience, and this observation was supported by independent

1 Recall CENCE ME ("sense me") is an anagram of the potential acronym, EEMCNCE, which would have been difficult for the reader (and the author) to recall. There is a minor irony of requiring an anagram to help us remember principles that in part refer to "our vast memory for language." But then our memory for language involves implicit, structured, overlapping representations that serve communicative functions, not arbitrary lists of meaningless and unpronounceable letter strings.

work on memory that has demonstrated that memories are detailed and sensitive to contextual information, while not being wholly veridical, meaning memory traces are somewhat abstracted away from our direct experience, as they involve lossy compression.

Most words, especially frequent words, are used in a cluster of conventional, related senses, because we regularly employ familiar words in new ways; doing so lessens the speaker's effort and increases the chance of successful comprehension by the listener. The widespread existence of polysemy—words with multiple distinct but related meanings—entails that memory traces for previous uses of a word cannot be deleted; that is, witnessing one sense cannot provide evidence that a different sense is not possible. Any misinterpretations of word meaning must instead be corrected by the creation of a stronger link between a different word and those contexts that had been associated with the misinterpretation.

The interpretations of utterances do not come from words alone but also from the way in which words and phrases are combined. Chapter 3 introduced argument structure constructions (ASCs), which convey types of events, or "who did what to whom." ASCs are conditioned by factors involving form, meaning, information structure, and social context. The lessons drawn from word meanings apply equally well to abstract grammatical ASCs. That is, in chapter 4 we saw that just as word usages form clusters of overlapping representations in our hyper-dimensional conceptual space, so, too, do usage events of ASCs. We regularly use both words and phrasal patterns in creative—partially productive—ways. We need to do this, because we regularly have new things to say, and the contexts we encounter are always changing. A productive use of a construction is supported to the extent that the potential coinage falls within a densely covered existing cluster of cases that exemplifies the construction.

Native-like fluency in production and comprehension requires a knowledge base of constructional clusters that have become sufficiently dense through experience to afford users the appropriate linguistic tools in a variety of contexts. Children (and L2 learners) are more likely to overextend words initially; for example, they may use *ball* to refer to a button. Likewise, they sometimes produce grammatical overgeneralization errors such as ?*I'm gonna cover a screen over you* (Bowerman, 1988), or ?*Explain me this*. Note that such overextensions or overgeneralizations are only considered errors because there exists a more conventional alternative way to express the intended message (in these cases, *button*; *cover you with something*; *Explain something to me*, respectively). When there is no available (combination of) conventional constructions to express an intended message-in-context, speakers must extend their available constructions in new ways. That is, when there is no conventional formulation, speakers make do with (combinations of) representations that are good enough.

An important role for competition and error-driven learning was detailed in chapter 5, offering a way for learners to overcome overgeneralizations and learn the constraints on words and constructions. As a preferred alternative becomes more familiar through repeated exposure, it will become easier to access than the dispreferred formulation through the process of statistical preemption. That is, words that express closely related meanings or grammatical constructions that have closely related functions in discourse are in competition with one another. When multiple representations are simultaneously activated in a given context to express the same (aspects of) the intended message, the representation with the greatest strength wins, and any other representations become progressively dissociated with features of that context. This accounts for why conventional formulations become easier to access. The reason we *judge* novel formulations to be "wrong," "inappropriate," or non-native-like when there exists a more appropriate way to express the particular meaning or function is that we want to speak like "our people." That is, shared language conventions signal that one is a member of the group. Using *ball* to refer to a button or saying ?*Explain me this* is considered "wrong" by native speakers because this is simply not how other native speakers speak.

Of course, we may intentionally flout the conventions of our language, if our purpose in doing so is to say something memorable, playful, or to draw special attention to some aspect of the constructions we choose to use. The last type of purpose can be illustrated by the transitive use of *disappear* as used in:

8.1 "four days after the military coup, they had disappeared her husband"

This has become a new conventional way to specifically emphasize that dissidents or rebels have been intentionally and directly made to disappear. It was originally an innovation but likely gained currency because the transitive causative emphasizes the culpability of the actors and the lack of control of the disappeared people in a way that the alternative does not (cf. *They made her husband disappear*).

Chapter 6 addressed several apparent paradoxes within language acquisition research. Children have been argued to be both more conservative (less likely to generalize) *and* overeager generalizers, when compared with adults. This apparent paradox was resolved by recognizing that children are less adept at quickly recognizing (or guessing) which dimensions are relevant for aligning exemplars within the hyper-dimensional conceptual space that maintains our long-term knowledge of language. When learners fail to identify or access the relevant SIMILARITIES or parallels among exemplars that are required to support the appropriate generalization, they behave conservatively. When learners fail to identify or access the relevant DISTINCTIONS among witnessed exemplars required to cleave one cluster into two or more distinct clusters, they overuse whichever option is more accessible when they speak (thereby "regularizing"). Competition between possible means of expressing a particular message leads to a preference for familiar formulations over novel formulations, whenever there exists a familiar formulation that is appropriate to express a particular message. If learners are unable to access the more familiar formulation (because it has not yet been well learned or is not sufficiently accessible at the moment of speaking), they produce formulations that native speakers consider unacceptable. Children eventually recover from these errors by taking advantage of error-driven implicit feedback, in which witnessing a more conventional formulation instead of any other formulation they may have anticipated serves to strengthen the former at the expense of the latter.

Adult learners of a second language often continue to produce errors of the *explain-me-this* type for interrelated reasons. Adults' first language has become well entrenched due to decades of practice. The hyper-dimensional conceptual space that contains their L1 representations has become warped for the purposes of automatic language use. While L2 learners are typically able to perceive the distinctions or similarities needed to align exemplars appropriately in L2, attending to those dimensions is more cognitively demanding than using the dimensions that have been highly practiced in their L1. Secondly, insofar as L2 learners are less able to anticipate upcoming grammatical forms, also possibly because doing so is cognitively demanding, they have less opportunity to benefit from error-driven learning.

Chapter 7 reviewed several alternative proposals for addressing the partial productivity of constructions. These included the suggestion that that speakers only stray very minimally from what they have witnessed ("conservatism via entrenchment"); that positing invisible syntactic diacritics or underlying structures could somehow address the issue from the learner's perspective; and that proposing a fixed upper bound on the proportion of exceptions and fixed lower bound on the proportion of rule-following cases could predict how and when generalizations are productive. While aspects of these proposals were useful, none of them attempted to address both the creative potential and restrictions at the same time. We have argued that creativity and constraints are flip sides of the same coin: creativity is encouraged to the extent that the ad hoc dynamic category required to include a new coinage is sufficiently well attested, and constraints arise from competition with alternative possible ways of expressing the intended message in the given context.

It is important to bear in mind that we do not generalize in a "blind" fashion. Instead, the relevance of particular dimensions is informed by attentional biases, predictive success (mutual information), and by prior learning. Also important are our immediate goals. For instance, infants learn that certain phonetic contrasts are predictable insofar as they do not distinguish one word from another. But the same phonetic distinctions may remain useful for distinguishing Boston from British accents, or for identifying the voices of particular people. When we pick up a cup we take into account its weight and density, but if we are deciding which objects are cups and which are bowls, these are not the dimensions that matter. In choosing lettuce, different shades of green distinguish romaine from spinach, but when we are navigating an intersection, any green light will do. In a parallel way, distinct dimensions are relevant to different grammatical constructions, depending on the functional purpose of the construction.

The emphasis in this book has been on argument structure constructions (ASCs)—constructions that determine "who did what to whom." ASCs are relatively simple grammatical constructions that involve an array of conditioning factors (chapters 3 and 4). As discussed in section 4.4, predicates (verbs and adjectives) are particularly relevant to their constraints. But, importantly, other dimensions are relevant for other constructions. For instance, determiners are highly relevant to the form and function of noun phrase constructions; *wh*- words and auxiliary verbs are highly relevant to the function of question constructions (e.g., *who* vs. *how come*; *aren't* vs. *isn't*) (Rowland et al., 2003). Learners need to discern which dimensions are relevant for each construction in their language.

Grammatical constructions typically combine with other constructions (recall section 3.8). Many grammatical constructions are distinguished primarily by the way they package information in discourse. These constructions combine with the ASCs focused on in this volume. For example, a simple transitive ASC can be combined with a cleft construction, a left-dislocation construction, or a topicalization construction, as in (8.2):

8.2a	It was a giraffe that the mouse saw.	(it-cleft)
8.2b	The giraffe, the mouse saw it.	(left-dislocation)
8.2c	The giraffe the mouse saw.	(topicalization)

Each of these constructions is associated with unique information-structure properties. For example, in an analysis of the Switchboard corpus of spoken language, Gregory and Michaelis (2001) document that the fronted noun phrase in the left-dislocation construction (8.2b) tends to be new to the discourse, and continues to be used as a topic of the discourse after the sentence ends. The fronted noun phrase in the "topicalization" construction (8.2c) displays the opposite tendency: it tends to have been previously mentioned without persisting as a topic in the following discourse. Thus, the left-dislocation construction is topic establishing, whereas the topicalization construction tends to be used for moribund topics (Gregory and Michaelis, 2001).

Other constructions convey various attitudes or emotions, including incredulity (8.3) (Akmajian, 1984; Lambrecht, 1990) or skepticism (8.4) (Gordon and Lakoff, 1971):

8.3 <accusative NounPhrase>, <NounPhrase>?
e.g., "Him, a man of the cloth?"
8.4 Why <VerbPhrase>?
e.g., "Why watch the performance on video?"

The NICE-OF-YOU construction in (8.5) is used to convey a judgment about how an action reflects on the agent of that action:

8.5 (It is) <adjective> [of <someone>] (<*to*-verb-phrase>) NICE-OF-YOU construction
e.g., "It is nice of you to defend the honor of those who deserve it."

This construction is not used to simply describe a person objectively (8.6a), nor can it be used to capture an evaluation of an inanimate entity (8.6b). The construction is only compatible with adjectives that can be interpreted as evaluative. For example, in (8.6c) *big* must be interpreted to mean "generous," it cannot simply mean "large" (Herbst and Goldberg, unpublished MS):

8.6a ?It's tall of you to reach the top shelf.
8.6b ?It's helpful of the dishwasher to save water.
8.6c It's big of you to reach the top shelf. ≈ "It's generous of you."

There are hundreds of grammatical constructions, thousands of idioms and conventional phrases, and tens of thousands of words (Culicover, 1999; Jackendoff, 2002; Sinclair, 1991; Wray, 2002). These collectively form our network of interrelated knowledge of language, our CONSTRUCTICON (Desagulier, 2015; Goldberg, 1995; Hoffman and Trousdale, 2013; Langacker, 1987). The constructicon is highly structured, just as the lexicon has long been recognized to be, yet the constructicon is much larger (see also Jackendoff and Audring, 2016).

Linguists and psychologists often study sentences in isolation, which may be akin to studying animals in separate cages in a zoo. I have been guilty of this in much of my own work. Our understanding of language will undoubtedly benefit from more focus on language in its natural habitat: conversation (e.g., Du Bois et al., 2003; Hilpert, 2017; Thompson and Hopper, 2001). In fact, the human tendency to cooperate plays a key role in how our complex system of language emerges via cultural evolution (Botha and Knight, 2009; De Boer et al., 2012; Ellis and Larsen-Freeman, 2009; Richerson and Christiansen, 2013;

Steels, 2005; Tomasello, 2009). Cooperation is most evident when people use language to communicate with one another in conversations.

Language provides a constrained and discrete system that offers us a window into our even more general, creative, but constrained system of general-purpose knowledge. It allows us to teach and learn, dream and imagine, and reflect and reason in ways that are uniquely human. Moreover, a focus on the functions and distributions of constructions offers important insights about how individual constructions emerge and evolve over time (e.g., Barðdal et al., 2015; Mauri and Sansò, 2011; Traugott, 2015; Traugott and Trousdale, 2013).

There is a growing synergy among linguists, psychologists, anthropologists, and computer scientists, and so it is a very exciting time for research on language. This volume only scratches the surface of what we have already learned, as is evident in the copious list of references.

There also remains much more work to do. While many subtle differences and relationships among particular constructions have been documented across a range of languages, this volume has focused primarily on English. Clearly, more attention to typological work is needed to better understand the oftentimes subtle functional differences in less studied languages; this of course requires great diligence on the part of researchers as well as an open mind. We can't rely only on published grammars because authors are often biased by the languages they speak. For example, forms may be described as "subjects" or as "adjectives" because that is how the meanings would be translated into English. We need to respect the subtle and complex ways in which languages may differ from one another (see, e.g., Croft, 2001; Delancey, 2012; Fried and Östman, 2004; LaPolla, 2017).

Researchers need to be well versed in information structure and lexical semantics, or our analyses of particular words and constructions and how they combine will suffer. The need is particularly acute in modeling work, which has gained well-deserved attention, but which currently lacks the means to capture relational meanings or the myriad ways in which words and grammatical constructions combine to give rise to interpretations in context. At the same time, modeling work has begun to tackle problems that many of us never imagined it would a decade ago, so it is an opportune time to investigate how computational models can capture information about constructions beyond the level of individual words (e.g., Dunn, 2017; Fitz et al., 2011; Kádár et al., 2017; Tsao and Wible, 2013; Van Trijp and Steels, 2012).

This book is only intended to whet the appetite of those who are eager for an approach to language that is responsive to what we know about memory, categorization, and learning. Far from answering all of the relevant questions, my goal here has been modest. I hope to have avoided as much unnecessary jargon as possible in order to encourage students, teachers, and researchers to focus on memory, learning, and the function of language as a means of communication, as we continue to study how grammatical constructions are constructed.

References

Abbot-Smith, K., E. Lieven, and M. Tomasello. 2008. "Graded Representations in the Acquisition of English and German Transitive Constructions." *Cognitive Development* 23 (1): 48–66.

Abbot-Smith, K., C. Rowland, F. Chang, J. Pine, and H. Ferguson. 2017. "Two and Three Year Old Children Use a Incremental First-NP-as-Agent Bias to Process Transitive Sentences." *PLoS One* 12 (10): e0186129.

Ackerman, F., G. T. Stump, and G. Webelhuth. 2011. "Lexicalism, Periphrasis, and Implicative Morphology." In *Non-transformational Syntax: Formal and Explicit Models of Grammar*, edited by R. Borsley and K. Börjars, 325–58. Hoboken, NJ: Wiley.

Adelman, J. S., S. G. Brown, and J. F. Quesada. 2006. "Contextual Diversity, not Word Frequency, Determines Word-Naming and Lexical Decision Times." *Psychological Science* 17 (9): 814–23.

Aha, D. W., D. Kibler, and M. K. Albert. 1991. "Instance-Based Learning Algorithms." *Machine Learning* 6 (1): 37–66.

Ahrens, K. V. 1995. "The Mental Representation of Verbs." PhD thesis, University of California.

Aĭkhenvald, A. Y. 2000. *Classifiers: A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.

Aĭkhenvald, A. Y., and R.M.W. Dixon, eds. 2006. *Serial Verb Constructions: A Cross-Linguistic Typology*. Vol. 2. Oxford: Oxford University Press.

Akhtar, N. 1999. "Acquiring Basic Word Order: Evidence for Data-Driven Learning of Syntactic Structure." *Journal of Child Language* 26 (2): 339–56.

Akhtar, N., M. Carpenter, and M. Tomasello 1996. "The Role of Discourse Novelty in Early Word Learning." *Child Development* 67(2): 635–45.

Akhtar, N., and M. Tomasello. 1997. "Young Children's Productivity with Word Order and Verb Morphology." *Developmental Psychology* 33 (6): 952.

Akmajian, A. 1984. "Sentence Types and the Form-Function Fit." *Natural Language and Linguistic Theory* 2 (1): 1–23.

Albright, A., and B. Hayes. 2003. "Rules vs. Analogy in English Past Tenses: A Computational/Experimental Study." *Cognition* 90 (2): 119–61.

Alishahi, A., and S. Stevenson. 2008. "A Computational Model of Early Argument Structure Acquisition." *Cognitive Science* 32(5): 789–834.

Allen, K., F. Pereira, M. Botvinick, and A. E. Goldberg. 2012. "Distinguishing Grammatical Constructions with fMRI Pattern Analysis." Brain and Language 123 (3): 174–82.

Alsina, A., J. Bresnan, and P. Sells. 1997. *Complex Predicates*. Stanford, CA: Center for the Study of Language and Information.

Alsina, A., and S. A. Mchombo. 1993. "Object Asymmetries and the Chichewa Applicative Construction." *Theoretical Aspects of Bantu Grammar* 1: 17–45.

Altmann, G. T., and Y. Kamide. 1999. "Incremental Interpretation at Verbs: Restricting the Domain of Subsequent Reference." *Cognition* 73 (3): 247–64.

Ambridge, B. 2013. "How Do Children Restrict Their Linguistic Generalizations? An (Un-) Grammaticality Judgment Study." *Cognitive Science* 37 (3): 508–43.

Ambridge, B., L. Barak, E. Wonnacott, C. Bannard, & G. Sala. (2018). "Effects of Both Preemption and Entrenchment in the Retreat from Verb Overgeneralization Errors: Four Reanalyses, an Extended Replication, and a Meta-Analytic Synthesis." *Collabra Psychology*, 4(1).

Ambridge, B., A. Bidgood, K. E. Twomey, J. M. Pine, C. F. Rowland, and D. Freudenthal. 2015. "Preemption versus Entrenchment: Towards a Construction-General Solution to the Problem of the Retreat from Verb Argument Structure Overgeneralization." *PLoS One* 10 (4): e0123723.

Ambridge, B., and R. P. Blything. 2016. "A Connectionist Model of the Retreat from Verb Argument Structure Overgeneralization." *Journal of Child Language* 43 (6): 1245–76.

Ambridge, B., and S. Brandt. 2013. "Lisa Filled Water into the Cup: The Roles of Entrenchment, Pre-emption and Verb Semantics in German Speakers' Acquisition of English Locatives." *Zeitschrift für Anglistik und Amerikanistik* 61 (3): 245–63.

Ambridge, B., and A. E. Goldberg. 2008. "The Island Status of Clausal Complements: Evidence in Favor of an Information Structure Explanation." *Cognitive Linguistics* 19 (3): 357–89.

Ambridge, B., and E. Lieven. 2015. "A Constructivist Account of Child Language Acquisition." In *The Handbook of Language Emergence*, edited by B. MacWhinney and W. O'Grady, 478–510. Hoboken, NJ: Wiley.

Ambridge, B., C. H. Noble, and E. V. Lieven. 2014. "The Semantics of the Transitive Causative Construction: Evidence from a Forced-Choice Pointing Study with Adults and Children." *Cognitive Linguistics* 25 (2): 293–311.

Ambridge, B., J. M. Pine, and C. F. Rowland. 2012a. "Semantics versus Statistics in the Retreat from Locative Overgeneralization Errors." *Cognition* 123 (2): 260–79.

Ambridge, B., J. M. Pine, C. F. Rowland, and F. Chang. 2012b. "The Roles of Verb Semantics, Entrenchment and Morphophonology in the Retreat from Dative Argument Structure Overgeneralization Errors." *Language* 88 (1): 45–81.

Ambridge, B., J. M. Pine, C. F. Rowland, D. Freudenthal, and F. Chang. 2014. "Avoiding Dative Overgeneralisation Errors: Semantics, Statistics or Both?" *Language, Cognition and Neuroscience* 29 (2): 218–43.

Ambridge, B., J. M. Pine, C. F. Rowland, R. F. Jones, and V. Clark. 2009. "A Semantics-Based Approach to the 'No Negative Evidence' Problem." *Cognitive Science* 33(7): 1301–16.

Ambridge, B., J. M. Pine, C. F. Rowland, and C. R. Young. 2008. "The Effect of Verb Semantic Class and Verb Frequency (Entrenchment) on Children's and Adults' Graded Judgments of Argument-Structure Overgeneralization Errors." *Cognition* 106 (1): 87–129.

Ameka, F. K. 2006. "Ewe Serial Verb Constructions in Their Grammatical Context." In *Serial Verb Constructions: A Cross-Linguistic Typology*, edited by A. I. Aĭkhenvald and R.M.W. Dixon, 124–43. Chicago: Oxford University Press.

Ameka, F. K., and J. Essegbey. 2013. "Serialising Languages: Satellite-Framed, Verb-Framed or Neither." *Ghana Journal of Linguistics* 21: 19–38.

Anderson, J. R., and G. H. Bower. 2014. *Human Associative Memory*. Hove, UK: Psychology Press.

Anderson, M. C., E. L. Bjork, and R. A. Bjork. 2000. "Retrieval-Induced Forgetting: Evidence for a Recall-Specific Mechanism." *Psychonomic Bulletin and Review* 7 (3): 522–30.

Anderson, M. C., C. Green, and K. C. McCulloch. 2000. "Similarity and Inhibition in Long-Term Memory: Evidence for a Two-Factor Theory." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26 (5): 1141.

Anderson, M. C., and B. A. Spellman. 1995. "On the Status of Inhibitory Mechanisms in Cognition: Memory Retrieval as a Model Case." *Psychological Review* 102 (1): 68.

Anderson, S. R. 1971. "On the Role of Deep Structure in Semantic Interpretation." *Foundations of Language* 7 (3): 387–96.

Antón-Méndez, I. 2010. "Gender Bender: Gender Errors in L2 Pronoun Production." *Journal of Psycholinguistic Research* 39 (2): 119–39.

Aravind, A., J. de Villiers, A. Pace, H. Valentine, R. Golinkoff, K. Hirsch-Pasek, A. Iglesias, and M. S. Wilson. 2018. "Fast Mapping Word Meanings across Trials: Young Children Forget All but Their First Guess." *Cognition* 177: 177–88.

Arbib, M. A. 2017. "Dorsal and Ventral Streams in the Evolution of the Language-ready Brain: Linking Language to the World." Journal of Neurolinguistics 43 (B): 228–53.

Ariel, M. 1991. "The Function of Accessibility in a Theory of Grammar." *Journal of Pragmatics* 16 (5): 443–63.

Arnold, J. E., J. G. Eisenband, S. Brown-Schmidt, and J. C. Trueswell. 2000. "The Rapid Use of Gender Information: Evidence of the Time Course of Pronoun Resolution from Eyetracking." *Cognition* 76 (1): 13–26.

Arnold, J. E., M. K. Tanenhaus, R. J. Altmann, and M. Fagnano. 2004. "The Old and Thee, uh, New: Disfluency and Reference Resolution." *Psychological Science* 15 (9): 578–82.

Arnold, J. E., T. Wasow, A. Losongco, and R. Ginstrom, R. 2000. "Heaviness vs. Newness: The Effects of Structural Complexity and Discourse Status on Constituent Ordering." *Language* 76: 28–55.

Arnon, I., and M. H. Christiansen. 2017. "The Role of Multiword Building Blocks in Explaining L1–L2 Differences." *Topics in Cognitive Science* 9 (3): 621–36.

Arnon, I., and E. V. Clark. 2011. "Why Brush Your Teeth Is Better than Teeth: Children's Word Production Is Facilitated in Familiar Sentence-Frames." *Language Learning and Development* 7 (2): 107–29.

Arnon, I., S. M. McCauley, and M. H. Christiansen. 2017. "Digging Up the Building Blocks of Language: Age-of-Acquisition Effects for Multiword Phrases." *Journal of Memory and Language* 92: 265–80.

Arnon, I., and N. Snider. 2010. "More than Words: Frequency Effects for Multi-word Phrases." *Journal of Memory and Language* 62 (1): 67–82.

Aronoff, M. 1976. *Morphology in Generative Grammar*. Cambridge, MA: MIT Press.

Aronoff, M., and M. Lindsay. 2016. "Competition and the Lexicon." *Livelli di Analisi e fenomeni di interfaccia. Atti del XLVII congresso internazionale della società di linguistica Italiana* 47: 39–52.

Artola, A., S. Brocher, and W. Singer. 1990. "Different Voltage-Dependent Thresholds for Inducing Long-Term Depression and Long-Term Potentiation in Slices of Rat Visual Cortex." *Nature* 347 (6288): 69.

Arunachalam, S., E. Escovar, M. A. Hansen, and S. R. Waxman. 2013. "Out of Sight, but Not Out of Mind: 21-Month-Olds Use Syntactic Information to Learn Verbs Even in the Absence of a Corresponding Event." *Language and Cognitive Processes* 28 (4): 417–25.

Arunachalam, S., and S. R. Waxman. 2010. "Meaning from Syntax: Evidence from 2-Year-Olds." *Cognition* 114 (3): 442–46.

Atkinson R. C., and R. M. Shiffrin. 1968. "Human Memory: A Proposed System and its Control Processes." In *The Psychology of Learning and Motivation: Advances in Research and Theory*, vol. 2, edited by K. W. Spence and J. T. Spence, 89–195. New York: Academic Press.

Austin, G., N. Pongpairoj, and D. Trenkic. 2015. "Structural Competition in Second Language Production: Towards a Constraint-Satisfaction Model." *Language Learning* 65 (3): 689–722.

Austin, J. L. 1962. *How to Do Things with Words*. Oxford: Clarendon.

Baayen, R. H. 2007. "Storage and Computation in the Mental Lexicon." In *The Mental Lexicon: Core Perspectives*, edited by G. Jarema and G. Libben, 81–104. Bingley, UK: Emerald.

Baayen, R. H., and F. M. del Prado Martin. 2005. "Semantic Density and Past-Tense Formation in Three Germanic Languages." *Language* 81 (3): 666–98.

Baayen, R. H., P. Hendrix, and M. Ramscar. 2013. "Sidestepping the Combinatorial Explosion: An Explanation of N-Gram Frequency Effects Based on Naive Discriminative Learning." *Language and Speech* 56 (3): 329–47.

Baayen, R. H., and R. Lieber. 1991. "Productivity and English Derivation: A Corpus-Based Study." *Linguistics* 29 (5): 801–44.

Baayen, R. H., and R. Schreuder. 2006. "Morphological Processing." In *Encyclopedia of Cognitive Science*, edited by L. Nadel, 131–54. Hoboken, NJ: Wiley.

Baker, C. L. 1979. "Syntactic Theory and the Projection Problem." *Linguistic Inquiry* 10 (4): 533–81.

Baker, M. C. 1997. "Thematic Roles and Syntactic Structure." In *Elements of Grammar*, edited by L. Haegeman, 73–137. Dordrecht, Netherlands: Springer.

Baldwin, D. A., and M. Tomasello. 1998. "Word Learning: A Window on Early Pragmatic Understanding." In *The Proceedings of the Twenty-Ninth Annual Child Language Research Forum*, edited by E. V. Clark, 3–23. Chicago: Center for the Study of Language and Information.

Balota, D. A., M. Pilotti, and M. J. Cortese. 2001. "Subjective Frequency Estimates for 2,938 Monosyllabic Words." *Memory and Cognition* 29 (4): 639–47.

Bannard, C., and D. Matthews. 2008. "Stored Word Sequences in Language Learning: The Effect of Familiarity on Children's Repetition of Four-Word Combinations." *Psychological Science* 19 (3): 241–48.

Bar, M. 2004. "Visual Objects in Context." *Nature Reviews Neuroscience* 5 (8): 617–29.

Barak, L., A. Fazly, and S. Stevenson. 2014. "Gradual Acquisition of Mental State Meaning: A Computational Investigation." *Proceedings of the Annual Meeting of the Cognitive Science Society* 36: 1886–91.

Barak, L., and A. E. Goldberg. 2017. "Modeling the Partial Productivity of Constructions." In *Proceedings of the American Association of Artificial Intelligence (AAAI) Symposium on Computational Construction Grammar and Natural Language Understanding*, Technical Report SS-17-02, 131–38 (Stanford, CA: AAAI).

Barak, L., A. E. Goldberg, and S. Stevenson. 2016. "Comparing Computational Cognitive Models of Generalization in a Language Acquisition Task." *Empirical Methods in Natural Language Processing* 2016: 96–106.

Bardi, L., L. Regolin, and F. Simion. 2011. "Biological Motion Preference in Humans at Birth: Role of Dynamic and Configural Properties." *Developmental Science* 14 (2): 353–59.

Barðdal, J. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. Vol. 8. Amsterdam: John Benjamins.

Barðdal, J., E. Smirnova, L. Sommerer, and S. Gildea, eds. 2015. *Diachronic Construction Grammar*. Constructional Approaches to Language 18. Amsterdam: John Benjamins.

Bartlett, F. C. 1932. *Remembering*. Oxford: Oxford University Press.

Bates, E., and B. MacWhinney. 1981. "Second-Language Acquisition from a Functionalist Perspective: Pragmatic, Semantic, and Perceptual Strategies." *Annals of the New York Academy of Sciences* 379 (1): 190–214.

———. 1987. "Competition, Variation, and Language Learning." In *Mechanisms of Language Acquisition*, edited by B. MacWhinney, 157–93. Hillsdale, NJ: Lawrence Erlbaum Associates.

Beck, D. M., and S. Kastner,. 2009. "Top-Down and Bottom-Up Mechanisms in Biasing Competition in the Human Brain." *Vision Research* 49 (10): 1154–65.

Beekhuizen, B., R. Bod, A. Fazly, S. Stevenson, and A. Verhagen. 2014. "A Usage-Based Model of Early Grammatical Development." *Annual Meeting of the Association for Computational Linguistics* 52: 46–54.

Bencini, G.M.L., and A. E. Goldberg. 2000. "The Contribution of Argument Structure Constructions to Sentence Meaning." *Journal of Memory and Language* 43 (4): 640–51.

Bergen, B. K. 2012. *Louder than Words: The New Science of How the Mind Makes Meaning*. New York: Basic Books.

Bergen, B. K., and N. Chang. 2005. "Embodied Construction Grammar in Simulation-Based Language Understanding." In *Construction Grammars: Cognitive Grounding and Theoretical Extensions*, vol. 3, edited by J. O. Östman and M. Fried, 147–90. Amsterdam: John Benjamins.

Bermel, N., and L. Knittl. 2012. "Corpus Frequency and Acceptability Judgments: A Study of Morphosyntactic Variants in Czech." *Corpus Linguistics and Linguistic Theory* 8 (2): 241–75.

Berwick, R. C., and A. S. Weinberg. 1986. *The Grammatical Basis of Linguistic Performance: Language Use and Acquisition*. Cambridge, MA: MIT Press.

Bickel, B. 2003. "Referential Density in Discourse and Syntactic Typology." *Language* 79 (4): 708–36.

Biederman, I., R. J. Mezzanotte, and J. C. Rabinowitz. 1982. Scene Perception: Detecting and Judging Objects Undergoing Relational Violations." *Cognitive Psychology* 14 (2): 143–77.

Birner, B. J., and G. Ward. 1998. *Information Status and Noncanonical Word Order in English*. Studies in Language Companion Series 40. Amsterdam: John Benjamins.

Blevins, J. P., F. Ackerman, R. Malouf, and M. Ramscar. 2016. "Morphology as an Adaptive Discriminative System." In *Morphological Metatheory*, edited by D. Siddiqi and H. Harley, 271–302. Amsterdam: John Benjamins.

Bley-Vroman, R. and H. R. Joo. 2001. "The Acquisition and Interpretation of English Locative Constructions by Native Speakers of Korean." *Studies in Second Language Acquisition* 23 (2): 207–19.

Bley-Vroman, R., and N. Yoshinaga. 1992. "Broad and Narrow Constraints on the English Dative Alternation: Some Fundamental Differences between Native Speakers and Foreign Language Learners." *University of Hawai'i Working Papers in ESL* 11 (1): 157–99.

Bloom, L. 1975. *Language Development*. Chicago: University of Chicago Press.

Bloom, P. 2000. *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press.

Blything, R. P., B. Ambridge, and E. V. Lieven. 2014. "Children Use Statistics and Semantics in the Retreat from Overgeneralization." *PLoS One* 9 (10): e110009.

Boas, F. 1911. "Introduction." In *Handbook of American Indian Languages*. Vol. 1. Bureau of American Ethnology bulletin 40. Washington, DC: Bureau of American Ethnology.

Boas, H. C. 2003. *A Constructional Approach to Resultatives*. Stanford, CA: Center for the Study of Language and Information.

———. 2014. "Lexical and Phrasal Approaches to Argument Structure: Two Sides of the Same Coin." *Theoretical Linguistics* 40 (1–2): 89–112.

Bod, R. 2009. "From Exemplar to Grammar: A Probabilistic Analogy-Based Model of Language Learning." *Cognitive Science* 33 (5): 752–93.

Bohannon, J. N., and L. B. Stanowicz. 1988. "The Issue of Negative Evidence: Adult Responses to Children's Language Errors." *Developmental Psychology* 5: 684–89.

Bongaerts, T., C. Van Summeren, B. Planken, and E. Schils. 1997. "Age and Ultimate Attainment in the Pronunciation of a Foreign Language." *Studies in Second Language Acquisition* 19 (4): 447–65.

Bonial, C. N. 2014. "Take a Look at This! Form, Function and Productivity of English Light Verb Constructions." PhD thesis, University of Colorado, Boulder.

Booij, G. E. 2002. *The Morphology of Dutch*. Oxford: Oxford University Press.

———. 2010. “Construction Morphology.” *Linguistics and Language Compass* 4 (7): 543–55.

———. 2017. “The Construction of Words.” In *The Cambridge Handbook of Cognitive Linguistics*, edited by B. Dancygier, chap. 15. Cambridge, UK: Cambridge University Press.

Borden, G., A. Gerber, and G. Milsark, G. 1983. “Production and Perception of the /r/-/l/ Contrast in Korean Adults Learning English.” *Language Learning* 33 (4): 499–526.

Borovsky, A., J. Elman, and A. Fernald. 2012. “Knowing a Lot for One’s Age: Vocabulary Skill and Not Age Is Associated with the Timecourse of Incremental Sentence Interpretation in Children and Adults.” *Journal of Experimental Child Psychology* 112 (4): 417–36.

Borovsky, A., M. Kutas, and J. Elman. 2010. “Learning to Use Words: Event-Related Potentials Index Single-Shot Contextual Word Learning.” *Cognition* 116 (2): 289–96.

Botha, R., and C. Knight, eds. 2009. *The Cradle of Language*. Studies in the Evolution of Language 12. Oxford: Oxford University Press.

Bowerman, M. 1982. “Reorganizational Processes in Lexical and Syntactic Development.” In *Language Acquisition: The State of the Art*, edited by E. Wanner and L. Gleitman, 319–46. New York: Academic Press.

———. 1988. “The ‘No Negative Evidence’ Problem: How Do Children Avoid Constructing an Overly General Grammar?” In *Explaining Language Universals*, edited by J. A. Hawkins, 73–101. Oxford: Blackwell.

Bowerman, M., and S. Choi. 2001. “Shaping Meanings for Language: Universal and Language-Specific in the Acquisition of Semantic Categories.” In *Language Acquisition and Conceptual Development*, edited by M. Bowerman and S. C. Levinson, 475–511. Cambridge, UK: Cambridge University Press.

Bowerman, M., and S. C. Levinson, eds. 2001. *Language Acquisition and Conceptual Development*. Language, Culture and Cognition 3. Cambridge, UK: Cambridge University Press.

Boyd, J. K., and A. E. Goldberg. 2009. “Input Effects within a Constructionist Framework.” *Modern Language Journal* 93 (3): 418–29.

———. 2011. “Learning What Not to Say: The Role of Statistical Preemption and Categorization in *a*-Adjective Production.” *Language* 87 (1): 55–83.

———. 2012. “Young Children Fail to Fully Generalize a Novel Argument Structure Construction when Exposed to the Same Input as Older Learners.” *Journal of Child Language* 39 (3): 457–81.

Boyd, J. K., E. A. Gottschalk, and A. E. Goldberg. 2009. “Linking Rule Acquisition in Novel Phrasal Constructions.” *Language Learning* 59 (s1): 64–89.

Boyd, R., and P. J. Richerson. 1988. *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.

Bradlow, A. R., and T. Bent. 2002. “The Clear Speech Effect for Non-native Listeners. *Journal of the Acoustical Society of America* 112 (1): 272–84.

Brady, T. F., T. Konkle, G. A. Alvarez, and A. Oliva. 2008. “Visual Long-Term Memory Has a Massive Storage Capacity for Object Details.” *Proceedings of the National Academy of Sciences of the USA* 105 (38): 14325–29.

Braine, M.D.S., and M. Bowerman. 1976. “Children’s First Word Combinations.” *Monographs of the Society for Research in Child Development* 41 (1): 1–104.

Braine, M.D.S., R. E. Brody, S. M. Fisch, M. J. Weisberger, and M. Blum. 1990. “Can Children Use a Verb without Exposure to Its Argument Structure?” *Journal of Child Language* 17 (2): 313–42.

Braine, M.D.S., and P. J. Brooks. 1995. “Verb Argument Structure and the Problem of Avoiding an Overgeneral Grammar.” In *Beyond Names for Things: Young Children’s Acquisition* *of Verbs*, edited by M. Tomasello and W. E. Merriman, 353–76. Hillsdale, NJ: Lawrence Erlbaum Associates.

Brainerd, C. J., and A. H. Mojardin. 1998. "Children's and Adults' Spontaneous False Memories: Long-Term Persistence and Mere-Testing Effects." *Child Development* 69 (5): 1361–77.

Brainerd, C. J., V. F. Reyna, and S. J. Ceci. 2008. "Developmental Reversals in False Memory: A Review of Data and Theory." *Psychological Bulletin* 134 (3): 343.

Brainerd, C. J., V. F. Reyna, and T. J. Forrest. 2002. "Are Young Children Susceptible to the False-Memory Illusion?" *Child Development* 73 (5): 1363–77.

Bresnan, J., A. Cueni, T. Nikitina, and R. H. Baayan. 2007. "Predicting the Dative Alternation." In *Cognitive Foundations of Interpretation*, edited by G. Boume, I. Kraemer, and J. Zwarts, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Bresnan, J., and M. Ford. 2010. "Predicting Syntax: Processing Dative Constructions in American and Australian Varieties of English." *Language* 86 (1): 168–213.

Briscoe, E. J. 1998. "Language as a Complex Adaptive System: Co-evolution of Language and of the Language Acquisition Device." In *Proceedings of Eighth Computational Linguistics in the Netherlands Conference*, 3–40. Amsterdam: Rodopi.

Broadbent, D. E. 1967. "Word-Frequency Effect and Response Bias." *Psychological Review* 74 (1): 1–15.

Brocher, A., J. Koenig, G. Mauner, and S. Foraker. 2017. "About Sharing and Commitment: The Retrieval of Biased and Balanced Irregular Polysemes." *Language, Cognition and Neuroscience* 33 (4): 443–66.

Brooks, P. J., and M. Tomasello. 1999. "How Children Constrain Their Argument Structure Constructions." *Language* 75: 720–38.

Brooks, P. J., M. Tomasello, K. Dodson, and L. B. Lewis. 1999. "Young Children's Overgeneralizations with Fixed Transitivity Verbs." *Child Development* 70 (6): 1325–37.

Brooks, P. J., and O. Zizak. 2002. "Does Preemption Help Children Learn Verb Transitivity?" *Journal of Child Language* 29 (4): 759–81.

Brown, G. D., and F. L. Watson. 1987. "First In, First Out: Word Learning Age and Spoken Word Frequency as Predictors of Word Familiarity and Word Naming Latency." *Memory and Cognition* 15 (3): 208–16.

Bryant, J. E. 2008. "Best-Fit Constructional Analysis." PhD dissertation. University of California, Berkeley.

Brysbaert, M., and B. New. 2009. "Moving beyond Kučera and Francis: A Critical Evaluation of Current Word Frequency Norms and the Introduction of a New and Improved Word Frequency Measure for American English." *Behavior Research Methods* 41 (4): 977–90.

Burzio, L. 2002. "Missing Players: Phonology and the Past-Tense Debate." *Lingua* 112 (3): 157–99.

Buz, E., T. F. Jaeger, and M. K. Tanenhaus. 2014. "Contextual Confusability Leads to Targeted Hyperarticulation." *Proceedings of the Annual Meeting of the Cognitive Science Society* 36: 1970–75.

Bybee, J. 1985. *Morphology: A Study of the Relation between Meaning and Form*. Typological Studies in Language 9. Amsterdam: John Benjamins.

———. 1995. "Regular Morphology and the Lexicon." *Language and Cognitive Processes* 10 (5): 425–55.

———. 2002. "Phonological Evidence for Exemplar Storage of Multiword Sequences." *Studies in Second Language Acquisition* 24 (2): 215–21.

———. 2003. “Cognitive Processes in Grammaticalization.” In *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*, edited by M. Tomasello, vol. 2, 145–67. Mahwah, NJ: Lawrence Erlbaum Associates.

———. 2010. *Language, Usage and Cognition*. Cambridge, UK: Cambridge University Press.

———. 2015. *Language Change*. Cambridge, UK: Cambridge University Press.

Bybee, J. L., and D. Eddington. 2006. “A Usage-Based Approach to Spanish Verbs of ‘Becoming.’” *Language* 82 (2): 323–55.

Bybee, J. L., J. Haiman, and S. A. Thompson, eds. 1997. *Essays on Language Function and Language Type: Dedicated to T. Givón*. Amsterdam: John Benjamins.

Bybee, J., and C. L. Moder. 1983. “Morphological Classes as Natural Categories.” *Language* 59: 251–70.

Bybee, J. L., and D. I. Slobin. 1982. “Rules and Schemas in the Development and Use of the English Past Tense.” *Language* 58 (2): 265–89.

Cacciari, C., N. Bolognini, I. Senna, M. C. Pellicciari, C. Miniussi, and C. Papagno. 2011. “Literal, Fictive and Metaphorical Motion Sentences Preserve the Motion Component of the Verb: A TMS Study.” *Brain and Language* 119 (3): 149–57.

Carey, S., and E. Bartlett. 1978. “Acquiring a Single New Word.” *Papers and Reports on Child Development* 15: 17–29.

Carpenter, M., N. Akhtar, and M. Tomasello. 1998. “Fourteen- through 18-Month-Old Infants Differentially Imitate Intentional and Accidental Actions.” *Infant Behavior and Development* 21 (2): 315–30.

Carstensen, A. 2016. “Universals and Variation in Language and Thought: Concepts, Communication, and Semantic Structure.” PhD thesis, University of California, Berkeley.

Casenhiser, D., and A. E. Goldberg. 2005. “Fast Mapping between a Phrasal Form and Meaning.” *Developmental Science* 8 (6): 500–508.

Chafe, W. L. 1976. “Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View.” In *Subject and Topic*, edited by C. N. Li, 25–55. New York: Academic Press.

Chang, F., G. S. Dell, and K. Bock. 2006. “Becoming Syntactic.” *Psychological Review* 113 (2): 234.

Chaves, R. P. 2012. “On the Grammar of Extraction and Coordination.” *Natural Language and Linguistic Theory* 30 (2): 465–512.

Childers, J. B., and M. Tomasello. 2002. “Two-Year-Olds Learn Novel Nouns, Verbs, and Conventional Actions from Massed or Distributed Exposures.” *Developmental Psychology* 38 (6): 967.

Choi, S. H., and T. Ionin. 2017. “Processing Plural Marking in the Second Language: Evidence for Transfer.” Poster presented at the CUNY Sentence Processing Conference, MIT, 29 March–1 April 2017.

Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton.

Chouinard, M. M., and E. V. Clark. 2003. “Adult Reformulations of Child Errors as Negative Evidence.” *Journal of Child Language* 30 (3): 637–69.

Christiansen, M. H., and N. Chater. 2016. *Creating Language: Integrating Evolution, Acquisition, and Processing*. Cambridge, MA: MIT Press.

Chung, T.T.R., and P. Gordon. 1998. “The Acquisition of Chinese Dative Constructions.” In *BUCLD 22: Proceedings of the 22nd Boston University Conference on Language Development*, edited by A. Greenhill, M. Hughes, and H. Littlefield, 109–20. Cambridge, MA: Boston University.

Citron, F. M., and A. E. Goldberg. 2014. “Metaphorical Sentences Are More Emotionally Engaging than Their Literal Counterparts.” *Journal of Cognitive Neuroscience* 26 (11): 2585–95.

Clahsen, H., and C. Felser. 2006. "Grammatical Processing in Language Learners." *Applied Psycholinguistics* 27: 3–42.

Clark, E. V. 1973. "What's in a Word? On the Child's Acquisition of Semantics in His First Language." In *Cognitive Development and the Acquisition of Language*, edited by T. Moore, 65–110. New York: Academic Press.

———. 1987. The Principle of Contrast: A Constraint on Language Acquisition." In *Mechanisms of Language Acquisition*, edited by B. MacWhinney, 1–33. Hillsdale, NJ: Lawrence Erlbaum Associates.

———. 1995. *The Lexicon in Acquisition*. Cambridge Studies in Linguistics 65. Cambridge, UK: Cambridge University Press.

———. 2003. "Critical Periods, Time, and Practice." *University of Pennsylvania Working Papers in Linguistics* 9 (2): 5.

Clark, E. V., and H. H. Clark. 1979. "When Nouns Surface as Verbs." *Language* 55 (4): 767–811.

Clark, H. H. 1996. *Using Language*. Cambridge, UK: Cambridge University Press.

Clark, H. H., and S. E. Haviland. 1977. "Comprehension and the Given-New Contract." In *Discourse Production and Comprehension*, edited by R. Freedle, 1–40. Santa Barbara, CA: Praeger.

Clausner, T. C., and W. Croft. 1997. "Productivity and Schematicity in Metaphors." *Cognitive Science* 21 (3): 247–82.

Clifton, C., Jr., L. Frazier, and C. Connine. 1984. "Lexical Expectations in Sentence Comprehension." *Journal of Verbal Learning and Verbal Behavior* 23 (6): 696–708.

Colleman, T., and B. De Clerck. 2011. "Constructional Semantics on the Move: On Semantic Specialization in the English Double Object Construction." *Cognitive Linguistics* 22 (1): 183–209.

Collins, P. 1995. "The Indirect Object Construction in English: An Informational Approach." *Linguistics* 33: 35–49.

Conwell, E., and K. Demuth. 2007. "Early Syntactic Productivity: Evidence from Dative Shift." *Cognition* 103: 163–79.

Copestake, A., and T. Briscoe. 1995. "Semi-productive Polysemy and Sense Extension." *Journal of Semantics* 12 (1): 15–67.

Coppock, E. 2009. "The Logical and Empirical Foundations of Baker's Paradox." PhD thesis, Stanford University.

Croft, W. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective.* Oxford: Oxford University Press.

———. 2003. "Lexical Rules vs. Constructions: A False Dichotomy. Motivation in Language." In *Motivation in Language: Studies in Honor of Gunter Radden*, edited by H. Cuyckens, T. Berg, R. Dirven, and K.-U. Panther, 49–68. Amsterdam: John Benjamins.

Croft, W., and D. A. Cruse. 2004. *Cognitive Linguistics*. Cambridge, UK: Cambridge University Press.

Culbertson, J., and E. L. Newport. 2017. "Innovation of Word Order Harmony across Development." *Open Mind* 1 (2): 91–100.

Culicover, P. W. 1999. *Syntactic Nuts: Hard Cases, Syntactic Theory, and Language Acquisition*. Oxford: Oxford University Press.

Culicover, P. W., and R. S. Jackendoff. 2005. *Simpler Syntax*. Oxford: Oxford University Press.

Dąbrowska, E. 2014. "Recycling Utterances: A Speaker's Guide to Sentence Processing." *Cognitive Linguistics* 25 (4): 617–53.

Daelemans, W., and A. van den Bosch. 2005. *Memory-Based Language Processing*. Cambridge, UK: Cambridge University Press.

Dahan, D., D. Swingley, M. K. Tanenhaus, and J. S. Magnuson. 2000. "Linguistic Gender and Spoken-Word Recognition in French." *Journal of Memory and Language* 42 (4): 465–80.
Darwin, C. (1859) 1993. *The Origin of Species by Means of Natural Selection*. Modern Library edition. New York: Random House.
Daugherty, K., and M. Seidenberg. 1992. "Rules or Connections? The Past Tense Revisited." *Proceedings of the Annual Meeting of the Cognitive Science Society*14: 259–64.
Dautriche, I. 2015. "Weaving an Ambiguous Lexicon." PhD thesis. University of Edinburgh.
Davies, M. 2004. *BYU-BNC*. Based on the British National Corpus from Oxford University Press. https://corpus.byu.edu/bnc/.
———. 2009. "The 385+ Million Word Corpus of Contemporary American English (1990–2008+): Design, Architecture, and Linguistic Insights." *International Journal of Corpus Linguistics* 14 (2): 159–90.
———. 2010. *The Corpus of Historical American English (COHA): 400 Million Words, 1810–2009*. https://corpus.byu.edu/coha/.
———. (2013) Corpus of News on the Web (NOW): 3+ billion words from 20 countries, updated every day. Available online at https://corpus.byu.edu/now/.
De Boer, B., W. Sandler, and S. Kirby. 2012. "New Perspectives on Duality of Patterning: Introduction to the Special Issue." *Language and Cognition* 4 (4): 251.
DeKeysar, R. M. 2005. "What Makes Learning Second Language Grammar Difficult? A Review of Issues." *Language Learning* 55: 1–25.
Delancey, S. 2012. "Still Mirative after All These Years." *Linguistic Typology* 16 (3): 529–64.
Delbecque, N., J. van der Auwera, and D. Geeraerts, eds. 2005. *Perspectives on Variation: Sociolinguistic, Historical, Comparative*. Berlin: Walter de Gruyter.
DeLong, K. A, D. M. Groppe, T. P. Urbach, and M. Kutas. 2012. "Thinking Ahead or Not? Natural Aging and Anticipation during Reading." *Brain and Language* 121 (3): 226–39.
Demetras, M. J., K. N. Post, and C. E. Snow. 1986. "Feedback to First Language Learners: The Role of Repetitions and Clarification Questions." *Journal of Child Language* 13 (2): 275–92.
Desagulier, G. 2015. "A Lesson from Associative Learning: Asymmetry and Productivity in Multiple-Slot Constructions." *Corpus Linguistics and Linguistic Theory* 12 (2). doi:10.1515/cllt-2015-0012.
Desai, R. H., J. R. Binder, L. L. Conant, Q. R. Mano, and M. S. Seidenberg. 2011. "The Neural Career of Sensory-Motor Metaphors." *Journal of Cognitive Neuroscience* 23 (9): 2376–86.
Diesendruck, G., and L. Markson. 2001. "Children's Avoidance of Lexical Overlap: A Pragmatic Account." *Developmental Psychology* 3(5): 630.
DiSciullo, A. M., and E. Williams. 1987. *On the Definition of Word*. Cambridge, MA: MIT Press.
Dittmar, M., K. Abbot-Smith, E. Lieven, and M. Tomasello. 2008. "German Children's Comprehension of Word Order and Case Marking in Causative Sentences." *Child Development* 79 (4): 1152–67.
Dixon, R.M.W. 1977. "Where Have All the Adjectives Gone?" *Studies in Language* 1 (1): 19–80.
———. 1994. *Ergativity*. Cambridge, UK: Cambridge University Press.
———. 2000. "A Typology of Causatives: Form, Syntax and Meaning." In *Changing Valency: Case Studies in Transitivity*, edited by R.M.W. Dixon and A. Y. Aĭkhenvald, 30–83. Cambridge, UK: Cambridge University Press.
Dowty, D. 1991. "Thematic Proto-roles and Argument Selection." *Language* 67 (3): 547–619.
Dryer, M. S. 1986. "Primary Objects, Secondary Objects, and Antidative." *Language* 62 (4): 808–45.
Du Bois, J. W. 1987. "The Discourse Basis of Ergativity." *Language* 63 (4): 805–55.

Du Bois, J. W., L. E. Kumpf, and W. J. Ashby, eds. 2003. *Preferred Argument Structure: Grammar as Architecture for Function*. Studies in Discourse and Grammar 14. Amsterdam: John Benjamins.

Dunn, J. 2017. "Computational Learning of Construction Grammars." *Language and Cognition* 9 (2): 254–92.

Durkin, K., and J. Manning. 1989. "Polysemy and the Subjective Lexicon: Semantic Relatedness and the Salience of Intraword Senses." *Journal of Psycholinguistic Research* 18 (6): 577–612.

Dussias, P. E., and N. Sagarra. 2007. "The Effect of Exposure on Syntactic Parsing in Spanish–English Bilinguals." *Bilingualism: Language and Cognition* 10 (1): 101–16.

Eberhard, K. M., M. J. Spivey-Knowlton, J. C. Sedivy, and M. K. Tanenhaus. 1995. "Eye Movements as a Window into Real-Time Spoken Language Comprehension in Natural Contexts." *Journal of Psycholinguistic Research* 24 (6): 409–36.

Eckert, P. 2012. "Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation." *Annual Review of Anthropology* 41: 87–100.

Ellis, N. C. 2006. "Selective Attention and Transfer Phenomena in L2 Acquisition: Contingency, Cue Competition, Salience, Interference, Overshadowing, Blocking, and Perceptual Learning." *Applied Linguistics* 27 (2): 164–94.

———. 2008. "The Periphery and the Heart of Language." In *Phraseology: An Interdisciplinary Perspective*, edited by S. Granger and F. Meunier, 1–13. Amsterdam: John Benjamins.

Ellis, N. C., and F. Ferreira-Junior. 2009. "Construction Learning as a Function of Frequency, Frequency Distribution, and Function." *Modern Language Journal* 93 (3): 370–85.

Ellis, N. C., and D. Larsen-Freeman. 2009. *Language as a Complex Adaptive System*. Hoboken, NJ: Wiley.

Ellis, N. C., R. Simpson-Vlach, and C. Maynard. 2008. "Formulaic Language in Native and Second-Language Speakers: Psycholinguistics, Corpus Linguistics, and TESOL." *TESOL Quarterly* 41: 375–96.

Elman, J. L. 1990. "Finding Structure in Time." *Cognitive Science* 14 (2): 179–211.

———. 2004. "An Alternative View of the Mental Lexicon." *Trends in Cognitive Sciences* 8 (7): 301–6.

———. 2009. "On the Meaning of Words and Dinosaur Bones: Lexical Knowledge without a Lexicon." *Cognitive Science* 33 (4): 547–82.

Emberson, L. L., A. M. Boldin, J. E. Riccio, R. Guillet, and R. N. Aslin. 2017. "Deficits in Top–Down Sensory Prediction in Infants at Risk Due to Premature Birth." *Current Biology* 27 (3): 431–36.

Emberson, L. L., C. Mazzei, and A. E. Goldberg. Unpublished MS. "The Blowfish Effect: Subordinate Categories are Inferred from Atypical Exemplars of a Basic Level Category."

Embick, D., and A. Marantz. 2008. "Architecture and Blocking." *Linguistic Inquiry* 39 (1): 1–53.

Enfield, N. J. 2002. "Functions of 'Give' and 'Take' in Lao Complex Predicates." In *Collected Papers on Southeast Asian and Pacific Languages*, edited by R. S. Bauer, 13–36. Canberra: Pacific Linguistics.

Erman, B., and B. Warren. 2000. "The Idiom Principle and the Open Choice Principle." *Text* 20 (1): 29–62.

Erteschik-Shir, N. 1979. "Discourse Constraints on Dative Movement." In *Syntax and Semantics*, vol. 12, *Discourse and Syntax*, edited by T. Givón, 441–67. New York: Academic Press.

———. 2007. *Information Structure: The Syntax-Discourse Interface*. Vol. 3. Oxford: Oxford University Press.

Essegbey, J. 1999. “Inherent Complement Verbs Revisited: Towards an Understanding of Argument Structure in Ewe.” PhD thesis, Radboud University Nijmegen.

Evans, N. 1997. “Role or Cast? Noun Incorporation and Complex Predicates in Mayali.” In *Complex Predicates*, edited by J. Bresnan, P. Sells, and A. Alsina, 397–430. Stanford, CA: Center for the Study of Language and Information.

Everett, D. L. 2005. “Cultural Constraints on Grammar and Cognition in Piraha.” *Current Anthropology* 46 (4): 621–46.

Fagot, J., and R. K. Thompson. 2011. “Generalized Relational Matching by Guinea Baboons (*Papio papio*) in Two-by-Two-Item Analogy Problems.” *Psychological Science* 22 (10): 1304–9.

Falkenhainer, B., K. D. Forbus, and D. Gentner. 1989. “The Structure-Mapping Engine: Algorithm and Examples.” *Artificial Intelligence* 41 (1): 1–63.

Family, N. 2006. *Explorations of Semantic Space: The Case of Light Verb Constructions in Persian*. Paris: Ecole des Hautes Etudes en Sciences Sociales.

Farrar, M. J. 1992. “Negative Evidence and Grammatical Morpheme Acquisition.” *Developmental Psychology* 28 (1): 90.

Farroni, T., M. H. Johnson, E. Menon, L. Zulian, D. Faraguna, and G. Csibra. 2005. “Newborns’ Preference for Face-Relevant Stimuli: Effects of Contrast Polarity.” *Proceedings of the National Academy of Sciences of the USA* 102 (47): 17245–50.

Fausey, C. M., S. Jayaraman, and L. B. Smith. 2016. “From Faces to Hands: Changing Visual Input in the First Two Years.” *Cognition* 152: 101–7.

Feldman, N. H., T. L. Griffiths, and J. L. Morgan. 2009. “The Influence of Categories on Perception: Explaining the Perceptual Magnet Effect as Optimal Statistical Inference.” *Psychological Review* 116 (4): 752.

Fellbaum, C. 2005. “Examining the Constraints on the Benefactive Alternation by Using the World Wide Web as a Corpus.” In *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*, edited by S. Kepser and M. Reis, 209–40. Berlin: Walter de Gruyter.

———. 1998. *WordNet*. Hoboken, NJ: Blackwell.

Ferman, S., and A. Kami. 2010. “No Childhood Advantage in the Acquisition of Skill in Using an Artificial Language Rule.” *PLoS One* 5: e12648.

Fernald, A., J. P. Pinto, D. Swingley, A. Weinbergy, and G. W. McRoberts. 1998. “Rapid Gains in Speed of Verbal Processing by Infants in the 2nd Year.” *Psychological Science* 9 (3): 228–31.

Fernald, A., K. Thorpe, and V. A. Marchman. 2010. “Blue Car, Red Car: Developing Efficiency in Online Interpretation of Adjective–Noun Phrases.” *Cognitive Psychology* 60 (3): 190–217.

Ferreira, F., and J. M. Henderson. 1990. “Use of Verb Information in Syntactic Parsing: Evidence from Eye Movements and Word-by-Word Self-Paced Reading.” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16 (4): 555.

Fillmore, C. J. 1968. “The Case for Case.” In *Universals in Linguistic Theory*, edited by E. Bach and R. T. Harms, 1–25. London: Holt, Rinehart, and Winston.

———. 1975. “An Alternative to Checklist Theories of Meaning.” *Berkeley Linguistic Society Proceedings* 1: 123–31.

———. 1976. “Frame Semantics and the Nature of Language.” *Annals of the New York Academy of Sciences* 280 (1): 20–32.

———. 1977. “Topics in Lexical Semantics.” In *Current Issues in Linguistic Theory*, edited by R. W. Cole, 76–138. Bloomington, IN: Indiana University Press.

———. 1982. “Frame Semantics.” In *Linguistics in the Morning Calm*, edited by the Linguistic Society of Korea, 111–38. Seoul: Hanshin.

———. 1984. "Lexical Semantics and Text Semantics." In *New Directions in Linguistics and Semiotics*, edited by J. E. Copeland, 123–47. Amsterdam: John Benjamins.

Finn, A. S., and C.L.H. Kam. 2015. "Why Segmentation Matters: Experience-Driven Segmentation Errors Impair 'Morpheme' Learning." *Journal of Experimental Psychology: Language, Memory, and Cognition* 41 (5): 1560.

Firestone, C., and B. J. Scholl. 2016. "Cognition Does Not Affect Perception: Evaluating the Evidence for 'Top-Down' Effects." *Behavioral and Brain Sciences* 39: 1–77.

Fisher, A. V., and V. M. Sloutsky. 2005. "When Induction Meets Memory: Evidence for Gradual Transition from Similarity-Based to Category-Based Induction." *Child Development* 76 (3): 583–97.

Fisher, C. L., Y. Gertner, R. M. Scott, and S. Yuan. 2010. "Syntactic Bootstrapping." *Wiley Interdisciplinary Reviews: Cognitive Science* 1 (2): 143–49.

Fisher, C. L., D. G. Hall, S. Rakowitz, and L. Gleitman. 1994. "When It Is Better to Receive Than to Give: Syntactic and Conceptual Constraints on Vocabulary Growth." *Lingua* 92: 333–75.

Fisher, C. L, and H. Tokura. 1996. "Acoustic Cues to Grammatical Structure in Infant Directed Speech: Cross-Linguistic Evidence." *Child Development* 67: 3192–218.

Fitz, H., F. Chang, and M. H. Christiansen. 2011. "A Connectionist Account of the Acquisition and Processing of Relative Clauses." In *The Acquisition of Relative Clauses*, edited by E. Kidd, 39–60. Amsterdam: John Benjamins.

Flege, J. E. 2009. "Give Input a Chance." In *Input Matters in SLA*, edited by T. Piske and M. Young-Scholten, 175–90. Bristol, UK: Multilingual Matters.

Flemming, T. M., R. K. Thompson, and J. Fagot. 2013. "Baboons, Like Humans, Solve Analogy by Categorical Abstraction of Relations." *Animal Cognition* 16 (3): 519–24.

Floyd, S., and A. E. Goldberg. Unpublished MS. "Children Make Use of Relationships across Meanings in Word Learning." Princeton University.

Fodor, J. A. 1975. *The Language of Thought*. Cambridge, MA: Harvard University Press.

Fodor, J. A., M. F. Garrett, E. C. Walker, and C. H. Parkes. 1980. "Against Definitions." *Cognition* 8 (3): 263–367.

Foley, W. A. 2008. "The Place of Philippine Languages in a Typology of Voice Systems." In *Voice and Grammatical Relations in Austronesian Languages*, edited by P. K. Austin and S. Musgrave, 22–44. Stanford, CA: Center for the Study of Language and Information.

Foley, W. A., and R. D. Van Valin Jr. 1984. *Functional Syntax and Universal Grammar*. Cambridge Studies in Linguistics 38. Cambridge, UK: Cambridge University Press.

Foraker, S., T. Regier, N. Khetarpal, A. Perfors, and J. Tenenbaum. 2009. "Indirect Evidence and the Poverty of the Stimulus: The Case of Anaphoric One." *Cognitive Science* 33 (2): 287–300.

Foucart, A., and C. Frenck-Mestre. 2011. "Grammatical Gender Processing in L2: Electrophysiological Evidence of the Effect of L1–L2 Syntactic Similarity." *Bilingualism: Language and Cognition* 14 (3): 379–99.

Foucart, A., C. D. Martin, E. Moreno, and A. Costa. 2014. "Can Bilinguals See It Coming? Word Anticipation in L2 Reading." *Journal of Experimental Psychology: Learning, Memory and Cognition* 40 (5): 1461–69.

François, A. 2008. "Semantic Maps and the Typology of Colexification." In *From Polysemy to Semantic Change: Towards a Typology of Lexical Semantic Associations*, edited by M. Vanhove, 163–215. Studies in Language Companion Series 106. Amsterdam: John Benjamins.

Frankland, S. M., and J. D. Greene. 2015. "An Architecture for Encoding Sentence Meaning in Left Mid-Superior Temporal Cortex." *Proceedings of the National Academy of Sciences of the USA* 112 (37): 11732–37.

Fried, M. 1995. "Grammatical Subject and Its Role in the Grammar of Case Languages." PhD thesis, University of California, Berkeley.
Fried, M., and J. O. Östman, eds. 2004. *Construction Grammar in a Cross-Language Perspective*. Vol. 2. Amsterdam: John Benjamins.
Friederici, A. D., A. Hahne, and A. Mecklinger. 1996. "Temporal Structure of Syntactic Parsing: Early and Late Event-Related Brain Potential Effects." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22 (5): 1219.
Futrell, R., L. Stearns, D. L. Everett, S. T. Piantadosi, and E. Gibson. 2016. "A Corpus Investigation of Syntactic Embedding in Pirahã." *PLoS One* 11 (3): e0145289.
Gahl, S., and S. M. Garnsey. 2004. "Knowledge of Grammar, Knowledge of Usage: Syntactic Probabilities Affect Pronunciation Variation." *Language* 80 (4): 748–75.
Gahl, S., and A. C. Yu. 2006. "Introduction to the Special Issue on Exemplar-Based Models in Linguistics." *Linguistic Review* 23 (3): 213–16.
Garnsey, S. M., N. J. Pearlmutter, E. Myers, and M. A. Lotocky. 1997. "The Contributions of Verb Bias and Plausibility to the Comprehension of Temporarily Ambiguous Sentences." *Journal of Memory and Language* 37 (1): 58–93.
Gaskell, M. G., and N. Dumay. 2003. "Lexical Competition and the Acquisition of Novel Words." *Cognition* 89 (2): 105–32.
Gauger, H. M. 1973. *Die Anfänge der Synonymik: Girard (1718) und Roubaud (1785); ein Beitrag zur Geschichte der lexikalischen Semantik*. Vol. 39. Tübingen: G. Narr.
Gause, G. F. 1934. "Experimental Analysis of Vito Volterra's Mathematical Theory of the Struggle for Existence." *Science* 79 (2036): 16–17.
Geeraerts, D. 1993. "Vagueness's Puzzles, Polysemy's Vagaries." *Cognitive Linguistics* 4 (3): 223–72.
Gelman, S. A., and L. Naigles. 1995. "Overextensions in Comprehension and Production Revisited: Preferential-Looking in a Study of Dog, Cat, and Cow." *Journal of Child Language* 22 (1): 19–46.
Gentner, D. 1983. "Structure-Mapping: A Theoretical Framework for Analogy." *Cognitive Science* 7 (2): 155–70.
———. 2003. "Why We're So Smart." In *Language in Mind: Advances in the Study of Language and Thought*, edited by D. Gentner and S. Goldin-Meadow, 195–235. Cambridge, MA: MIT Press.
Gentner, D., and K. D. Forbus. 2011. "Computational Models of Analogy." *Wiley Interdisciplinary Reviews: Cognitive Science* 2 (3): 266–76.
Gentner, D., M. Imai, and L. Boroditsky. 2002. "As Time Goes By: Evidence for Two Systems in Processing Space→Time Metaphors." *Language and Cognitive Processes* 17 (5): 537–65.
Gentner, D., and A. B. Markman. 1997. "Structure Mapping in Analogy and Similarity." *American Psychologist* 52 (1): 45.
Gershkoff-Stowe, L. 2001. "The Course of Children's Naming Errors in Early Word Learning." *Journal of Cognition and Development* 2 (2): 131–55.
Gertner, Y., and C. Fisher. 2012. "Predicted Errors in Children's Early Sentence Comprehension." *Cognition* 124 (1): 85–94.
Gertner, Y., C. Fisher, and J. Eisengart. 2006. "Learning Words and Rules: Abstract Knowledge of Word Order in Early Sentence Comprehension." *Psychological Science* 17 (8): 684–91.
Gibson, E., S. T. Piantadosi, K. Brink, L. Bergen, E. Lim, and R. Saxe. 2013. "A Noisy-Channel Account of Crosslinguistic Word-Order Variation." *Psychological Science* 24 (7): 1079–88.
Givón, T. 1979. *On Understanding Grammar*. Cambridge, MA: Academic Press.

———. 1984. "Direct Object and Dative Shifting: Semantic and Pragmatic Case." In *Objects: Towards a Theory of Grammatical Relations*, edited by F. Plank, 151–82. Cambridge, MA: Academic Press.

Gleitman, L. 1990. "The Structural Sources of Verb Meanings." *Language Acquisition* 1 (1): 3–55.

Godden, D. R., and A. D. Baddeley. 1975. "Context-Dependent Memory in Two Natural Environments: On Land and Underwater." *British Journal of Psychology* 66: 325–31.

Goldberg, A. E. 1992a. "Argument Structure Constructions." PhD thesis, University of California, Berkeley.

———. 1992b. "The Inherent Semantics of Argument Structure: The Case of the English Ditransitive Construction." *Cognitive Linguistics* 3 (1): 37–74.

———. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago, IL: Chicago University Press.

———. 1997. "The Relationships between Verbs and Constructions." In *Lexical and Syntactical Constructions and the Construction of Meaning*, edited by M. Verspoor, K. D. Lee, and E. Sweetser, 383–98. Amsterdam: John Benjamins.

———. 1999. "The Emergence of Argument Structure Semantics." In *The Emergence of Language*, edited by B. MacWhinney, 197–212. Mahwah, NJ: Lawrence Erlbaum Associates.

———. 2002. "Surface Generalizations: An Alternative to Alternations." *Cognitive Linguistics* 13 (4): 327–56.

———. 2003. "Constructions: A New Theoretical Approach to Language." *Trends in Cognitive Sciences* 7 (5): 219–24.

———. 2004. "Pragmatics and Argument Structure." In *The Handbook of Pragmatics*, edited by L. R. Horn and G. Ward, 427–43. Malden, MA: Blackwell.

———. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

———. 2010. "Verbs, Constructions, and Semantic Frames." In *Syntax, Lexical Semantics and Event Structure*, edited by M. Rappaport Hovav, E. Doron, and I. Sichel, 39–58. Oxford: Oxford University Press.

———. 2011a. "Are *A*-Adjectives Like Afraid Prepositional Phrases Underlying and Does It Matter from a Learnability Perspective?" Unpublished MS, Princeton University.

———. 2011b. "Corpus Evidence of the Viability of Statistical Preemption." *Cognitive Linguistics* 22: 131–53.

———. 2013. "Constructionist Approaches." In *The Oxford Handbook of Construction Grammar*, edited by T. Hoffmann and G. Trousdale, 15–31. Oxford: Oxford University Press.

———. 2014. "Fitting a Slim Dime between the Verb Template and Argument Structure Construction Approaches." *Theoretical Linguistics* 40 (1–2): 113–35.

———. 2015. "Compositionality." In *The Routledge Handbook of Semantics*, edited by N. Reimer, 419–34. London: Routledge.

———. 2016. "Tuning in to the English Verb-Particle Construction." In *Approaches to Complex Predicates*, Syntax and Semantics 41, edited by L. Nash and P. Samavilian, 50–65. Leiden: Brill.

Goldberg, A. E., L. Barak, and K. Tachihara. Unpublished MS. "A Closer Look at Whether Conservatism via Entrenchment Plays a Role in un- Prefixation." Princeton University.

Goldberg, A. E., and J. K. Boyd. 2015. "*A*-Adjectives, Statistical Preemption, and the Evidence: Reply to Yang." *Language* 91 (4): e184–97.

Goldberg, A. E., D. M. Casenhiser, and N. Sethuraman. 2004. "Learning Argument Structure Generalizations." *Cognitive Linguistics* 15 (3): 289–316.

———. 2005. “The Role of Prediction in Construction-Learning.” *Journal of Child Language* 32 (2): 407–26.

Goldberg, A. E., and R. Jackendoff. 2004. “The English Resultative as a Family of Constructions.” *Language* 80 (3): 532–68.

Goldberg, A. E., and L. A. Michaelis. 2017. “One among Many: Anaphoric One and Its Relationship with Numeral One.” *Cognitive Science* 41 (s2): 233–58.

Goldberg, A. E., and F. Perek. Forthcoming. “Ellipsis in Construction Grammar.” In *Handbook of Ellipsis*, edited by J. van Craenenbroeck and T. Temmerman. Oxford University Press.

Goldberg, A. E., and J. van der Auwera. 2012. “This Is to Count as a Construction.” *Folia Linguistica* 46 (1): 109–32.

Goldinger, S. D. 1998. “Echoes of Echoes? An Episodic Theory of Lexical Access.” *Psychological Review* 105 (2): 251.

Goldwater, M. B. 2017. “Grammatical Constructions as Relational Categories.” *Topics in Cognitive Science* 9 (3): 776–99.

Goldwater, M. B., and A. B. Markman. 2009. “Constructional Sources of Implicit Agents in Sentence Comprehension.” *Cognitive Linguistics* 20 (4): 675–702.

Gordon, D., and G. Lakoff. 1971. “Conversational Postulates.” *Papers from the Regional Meeting of the Chicago Linguistic Society* 7: 63–84.

Grant, P. R., and B. R. Grant. 2002. “Unpredictable Evolution in a 30-Year Study of Darwin’s Finches.” *Science* 296 (5568): 707–11.

Green, D. W. 1998. “Mental Control of the Bilingual Lexico-semantic System.” *Bilingualism: Language and Cognition*, 1 (2): 67–81.

Green, G. M. 1974. *Semantics and Syntactic Regularity*. Bloomington, IN: Indiana University Press.

Gregory, M. L., and L. A. Michaelis. 2001. “Topicalization and Left-Dislocation: A Functional Opposition Revisited.” *Journal of Pragmatics* 33 (11): 1665–706.

Grice, H. P. 1975. “Logic and Conversation.” In *Syntax and Semantics*, vol. 3, *Speech Acts*, edited by P. Cole and J. L. Morgan, 41–58. New York: Academic Press.

Gries, S. T. 2003. *Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement*. London: A. and C. Black.

Gries, S. T., and S. Wulff. 2005. “Do Foreign Language Learners also Have Constructions?” *Annual Review of Cognitive Linguistics* 3 (1): 182–200.

Gropen, J., S. Pinker, M. Hollander, and R. Goldberg. 1991. “Syntax and Semantics in the Acquisition of Locative Verbs.” *Journal of Child Language* 18 (1): 115–51.

Gropen, J., S. Pinker, M. Hollander, R. Goldberg, and R. Wilson. 1989. “The Learnability and Acquisition of the Dative Alternation in English.” *Language* 65 (2): 203–57.

Grüter, T., N. Hurtado, V. A. Marchman, and A. Fernald. 2014. “Language Exposure and Online Processing Efficiency in Bilingual Development.” *Input and Experience in Bilingual Development* 13: 15.

Gundel, J. K. 1985. “‘Shared knowledge’ and topicality.” *Journal of Pragmatics* 9 (1): 83–107.

Gurevich, O., M. A. Johnson, and A. E. Goldberg. 2010. “Incidental Verbatim Memory for Language.” *Language and Cognition* 2 (1): 45–78.

Gweon, H., P. Shafto, and L. Schulz. 2014. “Children Consider Prior Knowledge and the Cost of Information Both in Learning from and Teaching Others.” *Proceedings of the Annual Meeting of the Cognitive Science Society* 36: 565–70.

Hagoort, P., C. Brown, and J. Groothusen. 1993. “The Syntactic Positive Shift (SPS) as an ERP Measure of Syntactic Processing.” *Language and Cognitive Processes* 8 (4): 439–83.

Haiman, J. 1983. “Iconic and Economic Motivation.” *Language* 59: 781–819.

———, ed. 1985. *Iconicity in Syntax: Proceedings of a Symposium on Iconicity in Syntax, Stanford, June 24–6, 1983*. Typological Studies in Language 6. Amsterdam: John Benjamins.

Hall, M. L., R. I. Mayberry, and V. S. Ferreira. 2013. "Cognitive Constraints on Constituent Order: Evidence from Elicited Pantomime." *Cognition* 129 (1): 1–17.

Halliday, M., C. M. Matthiessen, and C. Matthiessen. 2014. *An Introduction to Functional Grammar*. London: Routledge.

Hao, J. 2015. "Abstraction versus Restriction in Syntactic Learning: An Examination of Children's Acquisition of the *a*-Adjective Restriction." Senior thesis, Princeton University.

Hare, M., J. L. Elman, T. Tabaczynski, and K. McRae. 2009. "The Wind Chilled the Spectators, but the Wine Just Chilled: Sense, Structure, and Sentence Comprehension." *Cognitive Science* 33 (4): 610–28.

Harmon, Z., and V. Kapatsinski. 2017. "Putting Old Tools to Novel Uses: The Role of Form Accessibility in Semantic Extension." *Cognitive Psychology* 98: 22–44.

Hart, B., and T. R. Risley. 1995. *Meaningful Differences in the Everyday Experience of Young American Children.* Baltimore, MD: Paul H. Brookes.

Hartshorne, J. K., J. B. Tenenbaum, & S. Pinker. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, 177(April), 263–277. https://doi.org/10.1016/j.cognition.2018.04.007

Haspelmath, M., ed. 2001. *Language Typology and Language Universals: An International Handbook*. Berlin: Walter de Gruyter.

———. 2010. "Comparative Concepts and Descriptive Categories in Crosslinguistic Studies." *Language* 86 (3): 663–87.

———. 2015. "Ditransitive constructions." *Annual Review of Linguistics* 1 (1): 19–41.

Hauk, O., I. Johnsrude, and F. Pulvermüller. 2004. "Somatotopic Representation of Action Words in Human Motor and Premotor Cortex." *Neuron* 41 (2): 301–7.

Havel, V. 1980. *Memorandum*. New York: Grove Press.

Havik, E., L. Roberts, R. van Hout, R. Schreuder, and M. Haverkort. 2009. "Processing Subject-Object Ambiguities in the L2: A Self-Paced Reading Study with German L2 Learners of Dutch." *Language Learning* 59: 73–112.

Hay, J. B., and J. Bresnan. 2006. "Spoken Syntax: The Phonetics of Giving a Hand in New Zealand English." *Linguistic Review* 23 (3): 321–49.

Herbst, T. 2014. "The Valency Approach to Argument Structure Constructions." In *Constructions, Collocations, Patterns*, edited by T. Herbst, H.-J. Schmid, and S. Faulhaber, 159–207. Berlin: Walter de Gruyter.

Herbst, T., and A. E. Goldberg. Unpublished MS. "The Nice-of-You Construction and its Fragments." Erlangen University.

Herbst, T., D. Heath, I. F. Roe, and D. Götz. 2004. *A Valency Dictionary of English*. Berlin: Mouton de Gruyter.

Herbst, T., and S. Schüller. 2008. *Introduction to Syntactic Analysis: A Valency Approach*. Tübingen, Germany: Narr.

Hernandez, A., P. Li, and B. MacWhinney. 2005. "The Emergence of Competing Modules in Bilingualism." *Trends in Cognitive Sciences* 9 (5): 220–25.

Heylen, K., T. Wielfaert, D. Speelman, and D. Geeraerts. 2015. "Monitoring Polysemy: Word Space Models as a Tool for Large-Scale Lexical Semantic Analysis." *Lingua* 157: 153–72.

Hilpert, M. 2017. "Historical Sociolinguistics and Construction Grammar." In *Exploring Future Paths for Historical Sociolinguistics*, Advances in Historical Sociolinguistics 7, edited by T. Säily, A. Nurmi, M. Palander-Collin, and A. Auer, 217–38. Amsterdam: John Benjamins.

Hinton, G. E., and J. A. Anderson, eds. 2014. *Parallel Models of Associative Memory*. Updated edition. New York: Psychology Press.
Hintzman, D. L. 1988. "Judgments of Frequency and Recognition Memory in a Multiple-Trace Memory Model." *Psychological Review* 95 (4): 528.
Hirsh-Pasek, K., R. Treiman, and M. Schneiderman. 1984. "Brown and Hanlon Revisited: Mothers' Sensitivity to Ungrammatical Forms." *Journal of Child Language* 2: 81–88.
Hoffmann, S., and J. Mukherjee. 2007. "Ditransitive Verbs in Indian English and British English: A Corpus-Linguistic Study." *Arbeiten aus Anglistik und Amerikanistik* 32 (1): 5–24.
Hoffmann, T., and G. Trousdale, eds. 2013. *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press.
Hopp, H. 2013. "Grammatical Gender in Adult L2 Acquisition: Relations between Lexical and Syntactic Variability." *Second Language Research* 29: 33–56.
Hopper, P. J., and S. A. Thompson. 1980. "Transitivity in Grammar and Discourse." *Language* 56 (2): 251–99.
Hopper, P. J., and E. C. Traugott. 2003. *Grammaticalization*. Cambridge, UK: Cambridge University Press.
Horner, V., and A. Whiten. 2005. "Causal Knowledge and Imitation/Emulation Switching in Chimpanzees (*Pan troglodytes*) and Children (*Homo sapiens*)." *Animal Cognition* 8 (3): 164–81.
Horst, J. S., and L. K. Samuelson. 2008. "Fast Mapping but Poor Retention by 24-Month-Old Infants." *Infancy* 13 (2): 128–57.
Hovav, M. R., and B. Levin. 1998. "Building Verb Meanings." In *The Projection of Arguments: Lexical and Compositional Factors*, edited by M. Butt and W. Geuder, 97–134. Stanford, CA: Center for the Study of Language and Information.
———. 2008. "The English Dative Alternation: The Case for Verb Sensitivity." *Journal of Linguistics* 44 (1): 129.
Hubbard, P. L., and D. Hix. 1988. "Where Vocabulary Meets Grammar: Verb Subcategorization Errors in ESL Writers." *CATESOL Journal* 1: 89–100.
Huddleston, R., and G. K. Pullum. 2002. *The Cambridge Grammar of English Language*. Cambridge, UK: Cambridge University Press.
Hughes, A., and P. Trudgill, P. 1996. *English Accents and Dialects*. London: Arnold.
Huth, A. G., W. A. de Heer, T. L. Griffiths, F. E. Theunissen, and J. L. Gallant. 2016. "Natural Speech Reveals the Semantic Maps that Tile the Human Cerebral Cortex." *Nature* 532: 453–58.
Huttenlocher, J. 1974. "The Origins of Language Comprehension." In *Theories in Cognitive Psychology: The Loyola Symposium*, edited by R. L. Solso, 331–68. Oxford: Lawrence Erlbaum Associates.
Hwang, J. D., 2014. "Identification and Representation of Caused Motion Constructions." PhD thesis, University of Colorado Boulder.
Imai, M., D. Gentner, and N. Uchida. 1994. "Children's Theories of Word Meaning: The Role of Shape Similarity in Early Acquisition." *Cognitive Development* 9 (1): 45–75.
Imai, M., E. Haryu, and H. Okada. 2005. "Mapping Novel Nouns and Verbs onto Dynamic Action Events: Are Verb Meanings Easier to Learn than Noun Meanings for Japanese Children?" *Child Development* 76: 340–55.
Inagaki, S. 1997. "Japanese and Chinese Learners' Acquisition of the Narrow-Range Rules for the Dative Alternation in English." *Language Learning* 47 (4): 637–69.
Ionin, T., and S. Montrul. 2010. "The Role of L1 Transfer in the Interpretation of Articles with Definite Plurals in L2 English." *Language Learning* 60 (4): 877–925.

Israel, M. 1996. "The Way Constructions Grow." In *Conceptual Structure, Discourse and Language*, edited by A. E. Goldberg, 217–30. Stanford, CA: Center for the Study of Language and Information.

Ito, A., A. E. Martin, and M. S. Nieuwland. 2017. "On Predicting Form and Meaning in a Second Language." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43 (4): 635.

Jackendoff, R. 1990. *Semantic Structures*. Cambridge, MA: MIT Press.

———. 1992. *Semantic Structures*. Current Studies in Linguistics 18. Cambridge, MA: MIT Press.

———. 1997. "Twistin' the Night Away." *Language* 73 (3): 534–59.

———. 2002. *Foundations of Language.* Oxford: Oxford University Press.

Jackendoff, R., and J. Audring. 2016. "Morphological Schemas." *Mental Lexicon* 11 (3): 467–93.

Jaeger, T. F. 2010. "Redundancy and Reduction: Speakers Manage Syntactic Information Density." *Cognitive Psychology* 61 (1): 23–62.

Jaeger, T. F. and R. Levy. 2006. "Speakers Optimize Information Density through Syntactic Reduction." Advances in Neural Information Processing Systems 19: 849–56.

Janda, L. A. 1990. "The Radial Network of a Grammatical Category: Its Genesis and Dynamic Structure." *Cognitive Linguistics* 1 (3): 269–88.

———. 1996. *Back from the Brink: A Study of How Relic Forms in Languages Serve as Source Material for Analogical Extension*. LINCOM Studies in Slavic Linguistics 1. Munich: Lincom Europa.

Janda, L. A., and F. Tyers, F. Forthcoming. "Less Is More: Why All Paradigms Are Defective, and Why That Is a Good Thing." *International Journal of Corpus Linguistics*.

Jäschke, K., and I. Plag. 2016. "The Dative Alternation in German-English Interlanguage." *Studies in Second Language Acquisition* 38 (3): 485–521.

Jescheniak, J. D., and W. J. Levelt. 1994. "Word Frequency Effects in Speech Production: Retrieval of Syntactic Information and of Phonological Form." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20 (4): 824.

Jin, K. S., and C. Fisher. 2014. "Early Evidence for Syntactic Bootstrapping: 15-Month-Olds Use Sentence Structure in Verb Learning." In *BUCLD 38 Online Proceedings Supplement*, edited by W. Orman and M. J. Valleau. Boston, MA: Cascadilla Press. http://www.bu.edu/bucld/files/2014/04/jin.pdf.

Johns, B. T., M. Dye, and M. N. Jones. 2016. "The Influence of Contextual Diversity on Word Learning." *Psychonomic Bulletin and Review* 23 (4): 1214–20.

Johnson, K. 2006. "Resonance in an Exemplar-Based Lexicon: The Emergence of Social Identity and Phonology." *Journal of Phonetics* 34 (4): 485–99.

Johnson, M. A., and A. E. Goldberg. 2013. "Evidence for Automatic Accessing of Constructional Meaning: Jabberwocky Sentences Prime Associated Verbs." *Language and Cognitive Processes* 28 (10): 1439–52.

Johnson, M. A., N. Turk-Browne, and A. E. Goldberg. 2013. "Prediction is essential to language processing and development. Comment on Pickering and Garrod." *Brain and Behavioral Science* 36 (4): 360–61.

Kaan, E. 2014. "Predictive Sentence Processing in L2 and L1: What Is Different?" *Linguistic Approaches to Bilingualism* 4 (2): 257–82.

Kaan, E., A. Dallas, and N. K. Wijnen. 2010. "Syntactic Predictions in Second-Language Sentence Processing." In *Structure Preserved: Festschrift in the Honor of Jan Koster*, edited by J.-W. Zwart and M. de Vries, 208–13. Amsterdam: John Benjamins.

Kaan, E., J. Kirkham, and F. Wijnen. 2016. "Prediction and Integration in Native and Second-Language Processing of Elliptical Structures." *Bilingualism* 19 (1): 1.

Kachergis, G., C. Yu, and R. M. Shiffrin. 2017. "A Bootstrapping Model of Frequency and Context Effects in Word Learning." *Cognitive Science* 41 (3): 590–622.

Kádár, A., G. Chrupała, and A. Alishahi. 2017. "Representation of Linguistic Form and Function in Recurrent Neural Networks." *Computational Linguistics* 43 (4): 761–80.

Kako, E. 2006. "The Semantics of Syntactic Frames." *Language and Cognitive Processes* 21 (5): 562–75.

Kalyan, S. 2012. "Similarity in Linguistic Categorization: The Importance of Necessary Properties." *Cognitive Linguistics* 23 (3): 539–54.

Kam, C.L.H. 2009. "More than Words: Adults Learn Probabilities over Categories and Relationships between Them." *Language Learning and Development* 5 (2): 115–45.

Kam, C.L.H., and E. L. Newport. 2005. "Regularizing Unpredictable Variation: The Roles of Adult and Child Learners in Language Formation and Change." *Language Learning and Development* 1 (2): 151–95.

———. 2009. "Getting It Right by Getting It Wrong: When Learners Change Languages." *Cognitive Psychology* 59 (1): 30-66.

Kaminski, J., J. Call, and J. Fischer. 2004. "Word Learning in a Domestic Dog: Evidence for 'Fast Mapping.'" *Science* 304 (5677): 1682–83.

Kang, J. 2017. "Statistical Preemption and Korean Learners of English." *Seoul National University Working Papers in English Language and Linguistics* 15: 40–52.

Kapatsinski, V. 2018. *Changing Minds Changing Tools: From Learning Theory to Language Acquisition to Language Change*. Cambridge, MA: MIT Press.

Kaschak, M. P. 2006. "What This Construction Needs Is Generalized." *Memory and Cognition* 34 (2): 368–79.

Kaschak, M. P., and A. Glenberg. 2000. "Constructing Meaning: The Role of Affordances and Grammatical Constructions in Sentence Comprehension." *Journal of Memory and Language* 43 (3): 508–29.

Kidd, C., S. T. Piantadosi, and R, N. Aslin. 2012. "The Goldilocks Effect: Human Infants Allocate Attention to Visual Sequences that are Neither Too Simple Nor Too Complex." *PLoS One* 7 (5): e36399.

Kidd, C., K. S. White, and R. N. Aslin. 2011. "Toddlers Use Speech Disfluencies to Predict Speakers' Referential Intentions." *Developmental Science* 14 (4): 925–34.

Kim, C. 2015. "L2 Learners' Recognition of Unfamiliar Idioms Composed of Familiar Words." *Language Awareness* 25 (1–2): 89–109.

Kim, G., J. Lewis-Peacock, K. Norman, and N. B. Turk-Browne. 2014. "Pruning of Memories by Context-Based Prediction Error." *Proceedings of the National Academy of Sciences of the USA* 111 (24): 8997–9002.

Kim, J. B., and P. Sells. 2013. "The Korean Sluicing: As a Family of Constructions." *Studies in Generative Grammar* 23 (1): 103–30.

Kim, R., and H.-K. Yang. 2017. "Why Do Nonnative English Learners Perform L2 Statistical Preemption Less than Native Counterparts?: The Role of Different Repertoires for L1 and L2 Constructions." Poster session at the 42nd Annual Boston University Conference on Language Development, Boston, 3–5 November 2017.

King, J. W., and M. Kutas. 1995. "A Brain Potential Whose Latency Indexes the Length and Frequency of Words." *CRL Newsletter* 10 (2): 1–9.

Kiparsky, P. 1982. "Lexical Morphology and Phonology." In *Linguistics in the Morning Calm*, edited by I. S. Yang. Seoul: Hanshin.

Kirby, S., H. Cornish, and K. Smith. 2008. "Cumulative Cultural Evolution in the Laboratory: An Experimental Approach to the Origins of Structure in Human Language." *Proceedings of the National Academy of Sciences of the USA* 105 (31): 10681–86.

Klein, D. E., and G. L. Murphy, G. 2001. "The Representation of Polysemous Words." *Journal of Memory and Language* 45 (2): 259–82.

Kleinschmidt, D. F., and T. F. Jaeger. 2011. "A Bayesian Belief Updating Model of Phonetic Recalibration and Selective Adaptation." *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics* 2: 10–19.

———. 2015. "Robust Speech Perception: Recognize the Familiar, Generalize to the Similar, and Adapt to the Novel." *Psychological Review* 122 (2): 148.

Koenig, J., and K. Michelson. 2017. "Invariance in Argument Realization: The Case of Iroquoian." *Language* 91 (1): 1–47.

Korsgaard, C. M., with G. A. Cohen, R. Geuss, T. Nagel, and B. Williams. 1996. *The Sources of Normativity*. Cambridge, UK: Cambridge University Press.

Kotovsky, L., and D. Gentner. 1996. "Comparison and Categorization in the Development of Relational Similarity." *Child Development* 67 (6): 2797–822.

Kousta, S. T., D. P. Vinson, and G. Vigliocco. 2009. "Emotion Words, Regardless of Polarity, Have a Processing Advantage over Neutral Words." *Cognition* 112 (3): 473–81.

Krashen, S. D., M. A. Long, and R. Scarcella. 1979. "Age, Rate, and Eventual Attainment in Second Language Acquisition." *TESOL Quarterly* 13: 573–82.

Kruschke, J. K. 1992. "ALCOVE: An Exemplar-Based Connectionist Model of Category Learning." *Psychological Review* 99 (1): 22.

Kuczaj, S. A. 1982. "Young Children's Overextensions of Object Words in Comprehension and/or Production: Support for a Prototype Theory of Early Object Word Meaning." *First Language* 3 (8): 93–105.

Kuhl, P. K., K. A. Williams, F. Lacerda, K. N. Stevens, and B. Lindblom. 1992. "Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age." *Science* 255 (5044): 606–8.

Kutas, M., and S. A. Hillyard. 1984. "Brain Potentials during Reading Reflect Word Expectancy and Semantic Association." *Nature* 307 (5947): 161–63.

Labov, W. 2011. *Principles of Linguistic Change*. Vol. 3, *Cognitive and Cultural Factors*. Hoboken, NJ: Wiley.

Lacey, S., R. Stilla, and K. Sathian. 2012. "Metaphorically Feeling: Comprehending Textural Metaphors Activates Somatosensory Cortex." *Brain and Language* 120 (3): 416–21.

Lakoff, G. 1966. "Irregularity in Syntax." PhD thesis, Indiana University.

———. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

Lambrecht, K. 1990. "'What, Me Worry?': 'Mad Magazine Sentences' Revisited." In *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society*, edited by K. Hall, J.-P. Koenig, M. Meacham, S. Reinman, and L. A. Sutton, 215–28. Berkeley, CA: Berkeley Linguistics Society.

———. 1994. *Information Structure and Sentence Form*. Cambridge, UK: Cambridge University Press.

Landau, B., and L. R. Gleitman. 1985. *Language and Experience: Evidence from a Blind Child*. Cambridge, MA: Harvard University Press.

Landau, B., L. B. Smith, and S. S. Jones. 1988. "The Importance of Shape in Early Lexical Learning." *Cognitive Development* 3 (3): 299–321.

Landauer, T. K. 2006. *Latent Semantic Analysis*. Hoboken, NJ: Wiley.

Langacker, R. W. 1987. *Foundations of Cognitive Grammar*. Vol. 1, *Theoretical Prerequisites*. Stanford, CA: Stanford University Press.

———. 1988. "A Usage-Based Model." *Topics in Cognitive Linguistics* 50: 127–63.

———. 2008. *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.

LaPolla, R. J. 1994. "Parallel Grammaticalizations in Tibeto-Burman Languages: Evidence of Sapir's 'Drift.'" *Linguistics of the Tibeto-Burman Area* 17 (1): 61–80.

———. 2009. "Chinese as a Topic-Comment (Not Topic-Prominent and Not SVO) Language." In *Studies of Chinese Linguistics: Functional Approaches*, edited by J. Xing, 9–22. Hong Kong: Hong Kong University Press.

———. 2016. "On Categorization: Stick to the Facts of the Languages." *Linguistic Typology* 20 (2): 365–75.

———. 2017. "Once Again on Methodology and Argumentation in Linguistics." *Linguistics of the Tibeto-Burman Area* 39 (2): 282–97.

Lebani, G. E., and A. Lenci. 2016. "'Beware the Jabberwock, Dear Reader!': Testing the Distributional Reality of Constructional Semantics." In *CogALex@COLING: Proceedings of the Workshop on Cognitive Aspects of the Lexicon*, Osaka, Japan, 11–17 December 2016, 8–18.

Leech, G. 2003. "Modality on the Move: The English Modal Auxiliaries 1961–1992." *Topics in English Linguistics* 44: 223–40.

Leow, R. P. 1998. "The Effects of Amount and Type of Exposure on Adult Learners' L2 Development in SLA." *Modern Language Journal* 82 (1): 49–68.

Levelt, W. J. 1983. "Monitoring and Self-Repair in Speech." *Cognition* 14 (1): 41–104.

Levin, B. 1993. *English Verb Classes and Alternations*. Chicago: Chicago University Press.

———. 2008. "Dative Verbs: A Crosslinguistic Perspective." *Lingvisticæ Investigationes* 31 (2): 285–312.

Levinson, S. C. 1983. *Pragmatics*. Cambridge, UK: Cambridge University Press.

———. 1996. "Relativity in Spatial Conception and Description." In *Rethinking Linguistic Relativity*, edited by J. J. Gumperz and S. C. Levinson, 177–202. Cambridge, UK: Cambridge University Press.

Levy, R. 2008. "A Noisy-Channel Model of Rational Human Sentence Comprehension under Uncertain Input." In *Proceedings of the 13th Conference on Empirical Methods in Natural Language Processing*, 234–43. Vancouver: Association for Computational Linguistics.

Levy, R., K. Bicknell, T. Slattery, and K. Rayner. 2009. "Eye Movement Evidence that Readers Maintain and Act on Uncertainty about Past Linguistic Input." *Proceedings of the National Academy of Sciences of the USA* 106 (50): 21086–90.

Lewis, M. 2008. "The Idiom Principle in L2 English: Assessing Elusive Formulaic Sequences as Indicators of Idiomaticity, Fluency, and Proficiency." PhD thesis, Uppsala University, English Institute.

Lew-Williams, C., and A. Fernald. 2007. "Young Children Learning Spanish Make Rapid Use of Grammatical Gender in Spoken Word Recognition." *Psychological Science* 33: 193–98.

———. 2010. "Real-Time Processing of Gender-Marked Articles by Native and Non-native Spanish Speakers." *Journal of Memory and Language* 63 (4): 447–64.

Li, P., I. Farkas, and B. MacWhinney. 2004. "Early Lexical Development in a Self-Organizing Neural Network." *Neural Networks* 17: 1345–62.

Liang, J. 2002. "Verbs and Argument Structure Constructions." Masters thesis, Gunangdong University of Foreign Studies.

Lidz, J., H. Gleitman, and L. Gleitman. 2003. "Understanding How Input Matters: Verb Learning and the Footprint of Universal Grammar." *Cognition* 87 (3): 151–78.

Lieven, E., J. Pine, and G. Baldwin. 1997. "Lexically-Based Learning and Early Grammatical Development." *Journal of Child Language* 24 (1): 187–220.

Light, L. L., and L. Carter-Sobell. 1970. "Effects of Changed Semantic Context on Recognition Memory." *Journal of Verbal Learning and Verbal Behavior* 9: 1–11.

Linck, J. A., J. F. Kroll, and G. Sunderman. 2009. "Losing Access to the Native Language while Immersed in a Second Language: Evidence for the Role of Inhibition in Second-Language Learning." *Psychological Science* 20: 1507–15.

Linzen, T., and T. F. Jaeger. 2014. "Investigating the Role of Entropy in Sentence Processing." In *CMCL@ACL: Proceedings of the 2014 Workshop on Cognitive Modeling and Computational Linguistics*, 10–18. Baltimore, MD: Association of Computational Linguistics.

Liu, M., M. C. Machado, G. Tesauro, and M. Campbell. 2017. "The Eigenoption–Critic Framework." Cornell University Library, *arXiv*, submitted 11 December 2017. http://arxiv.org/abs/1712.04065.

Long, R. B. 1961. *The Sentence and Its Parts: A Grammar of Contemporary English*. Chicago, IL: University of Chicago Press.

Luk, Z.P.S., and Y. Shirai. 2009. "Is the Acquisition Order of Grammatical Morphemes Impervious to L1 Knowledge? Evidence from the Acquisition of Plural -s, Articles, and Possessive 's." *Language Learning* 59 (4): 721–54.

Lukyanenko, C., and C. Fisher. 2016. "Where Are the Cookies? Two-and Three-Year-Olds Use Number-Marked Verbs to Anticipate Upcoming Nouns." *Cognition* 146: 349–70.

MacDonald, M. C. 2013. "How Language Production Shapes Language Form and Comprehension." *Frontiers in Psychology* 4 (April): 226.

MacDonald, M. C., N. J. Pearlmutter, and M. S. Seidenberg. 1994. "The Lexical Nature of Syntactic Ambiguity Resolution." *Psychological Review* 101 (4): 676.

MacWhinney, B. 1982. "Basic Syntactic Processes." *Language Acquisition* 1: 73–136.

———. 2006. "Emergent Fossilizaton." In *Studies of Fossilization in Second Language Acquisition*, vol. 14, edited by Z. Han and T. Odlin, 134–56. Buffalo, NY: Multilingual Matters.

Madlener, K. 2016. "Input optimization." In *Experience Counts: Frequency Effects in Language*, edited by H. Behrens and S. Pfänder, 133–174. Berlin: Walter de Gruyter.

Manela-Arnold, E., J. L. Evans, and J. A. Coady. 2010. "Explaining Lexical–Semantic Deficits in Specific Language Impairment: The Role of Phonological Similarity, Phonological Working Memory, and Lexical Competition." *Journal of Speech, Language, and Hearing Research* 53 (6): 1742–56.

Marantz, A. 1997. "No Escape from Syntax: Don't Try Morphological Analysis in the Privacy of Your Own Lexicon." *University of Pennsylvania Working Papers in Linguistics* 4 (2): 14.

Marchman, V. A. 1997. "Children's Productivity in the English Past Tense: The Role of Frequency, Phonology, and Neighborhood Structure." *Cognitive Science* 21 (3): 283–304.

Marcotte, J. 2005. "Causative Alternation Errors in Child Language Acquisition." PhD thesis, Stanford University.

Marcus, G. F., S. Vijayan, S. B. Rao, and P. M. Vishton. 1999. "Rule Learning by Seven-Month-Old Infants." *Science* 283 (5398): 77–80.

Marian, V., and M. Spivey. 2003. "Competing Activation in Bilingual Language Processing: Within-and Between-Language Competition." *Bilingualism: Language and Cognition* 6 (2): 97–115.

Marinova-Todd, S.H., D. B. Marshall, and C. E. Snow. 2000. "Three Misconceptions about Age and L2 Learning." *TESOL Quarterly* 34 (1): 9–34.

Markman, A. B., and D. Gentner. 1993. "Structural Alignment during Similarity Comparisons." *Cognitive Psychology* 25 (4): 431–67.

———. 2001. "Thinking." *Annual Review of Psychology* 52 (1): 223–47.

Markman, E. M. 1989. *Categorization and Naming in Children: Problems of Induction*. Cambridge, MA: MIT Press.

Markman, E. M., and J. E. Hutchinson. 1984. "Children's Sensitivity to Constraints on Word Meaning: Taxonomic versus Thematic Relations." *Cognitive Psychology* 16 (1): 1–27.

Markman, E. M., and G. F. Wachtel. 1988. "Children's Use of Mutual Exclusivity to Constrain the Meanings of Words." *Cognitive Psychology* 20 (2): 121–57.

Marshall, L., and J. Born. 2007. "The Contribution of Sleep to Hippocampus-Dependent Memory Consolidation." *Trends in Cognitive Sciences* 11 (10): 442–50.

Marslen-Wilson, W., and P. Zwitserlood. 1989. "Accessing Spoken Words: The Importance of Word Onsets." *Journal of Experimental Psychology: Human Perception and Performance* 15 (3): 576.

Martin, C. D., G. Thierry, J. Kuipers, B. Boutonnet, A. Foucart, and A. Costa. 2013. "Bilinguals Reading in Their Second Language Do Not Predict Upcoming Words as Native Readers Do." *Journal of Memory and Language* 69 (4): 574–88.

Matthews, D., and C. Bannard. 2010. "Children's Production of Unfamiliar Word Sequences Is Predicted by Positional Variability and Latent Classes in a Large Sample of Child-Directed Speech." *Cognitive Science* 34 (3): 465–88.

Matusevych, Y., A. Alishahi, and A. Backus. 2017. "The Impact of First and Second Language Exposure on Learning Second Language Constructions." *Bilingualism: Language and Cognition* 20 (1): 128–49.

Mauri, C., and A. Sansò. 2011. "How Directive Constructions Emerge: Grammaticalization, Constructionalization, Cooptation." *Journal of Pragmatics* 43 (14): 3489–521.

McCawley, J. 1978. "Conversational Implicature and the Lexicon." In *Syntax and Semantics*, vol. 9, *Pragmatics*, edited by Peter Cole, 245–59. New York: Academic Press.

McClelland, J. L., and J. L. Elman. 1986. "The TRACE Model of Speech Perception." *Cognitive Psychology* 18 (1): 1–86.

McClelland, J. L., and K. Patterson. 2002. "Rules or Connections in Past-Tense Inflections: What Does the Evidence Rule Out?" *Trends in Cognitive Sciences* 6 (11): 465–72.

McDonald, S. A., and R. C. Shillcock. 2001. "Rethinking the Word Frequency Effect: The Neglected Role of Distributional Information in Lexical Processing." *Language and Speech* 44 (3): 295–322.

McDonough, K., and Y. Kim. 2009. "Syntactic Priming, Type Frequency, and EFL Learners' Production of Wh-Questions." *Modern Language Journal* 93 (3): 386–98.

McDonough, K., and T. Nekrasova-Becker. 2014. "Comparing the Effect of Skewed and Balanced Input on English as a Foreign Language Learners' Comprehension of the Double-Object Dative Construction." *Applied Psycholinguistics* 35 (2): 419–42.

McGuigan, N., J. Makinson, and A. Whiten. 2011. "From Over-Imitation to Super-Copying: Adults Imitate Causally Irrelevant Aspects of Tool Use with Higher Fidelity than Young Children." *British Journal of Psychology* 102 (1): 1–18.

McMurray, B., M. K. Tanenhaus, R. N. Aslin, and M. J. Spivey. 2003. "Probabilistic Constraint Satisfaction at the Lexical/Phonetic Interface: Evidence for Gradient Effects of Within-Category VOT on Lexical Access." *Journal of Psycholinguistic Research* 32 (1): 77–97.

McRae, K., M. J. Spivey-Knowlton, and M. K. Tanenhaus. 1998. "Modeling the Influence of Thematic Fit (and Other Constraints) in On-Line Sentence Comprehension." *Journal of Memory and Language* 38 (3): 283–312.

Mervis, C. B., R. M. Golinkoff, and J. Bertrand. 1994. "Two-Year-Olds Readily Learn Multiple Labels for the Same Basic-Level Category." *Child Development* 65 (4): 1163–77.

Meyer, D. E., and R. W. Schvaneveldt. 1971. "Facilitation in Recognizing Pairs of Words: Evidence of a Dependence between Retrieval Operations." *Journal of Experimental Psychology* 90 (2): 227.

Michaelis, L. A. 2003. "Word Meaning, Sentence Meaning, and Syntactic Meaning." In *Cognitive Approaches to Lexical Semantics*, edited by H. Cuyckens, R. Dirven, and J. R. Taylor, 163–209. Berlin: Mouton de Gruyter.

———. 2005. "Entity and Event Coercion in a Symbolic Theory of Syntax." In *Construction Grammars: Cognitive Grounding and Theoretical Extensions*, edited by J.-O. Östman and M. Fried, 45–88. Amsterdam: John Benjamins.

Michaelis, L. A., and K. Lambrecht. 1996. "Toward a Construction-Based Theory of Language Function: The Case of Nominal Extraposition." *Language* 72 (2): 215–47.

Mithun, M. 1984. "The Evolution of Noun Incorporation." *Language* 60 (4): 847–94.

———. 1986. "On the Nature of Noun Incorporation." *Language* 62 (1): 32–37.

Mondorf, B. 2014. "(Apparently) Competing Motivations in Morpho-syntactic Variation." In: *Competing Motivations in Grammar and Usage*, edited by E. A. Moravcsik, A. Malchukov, and B. MacWhinney, 209–28. Oxford: Oxford University Press.

Montag, J. L., K. Matsuki, J. Y. Kim, and M. C. MacDonald. 2017. "Language Specific and Language General Motivations of Production Choices: A Multi-Clause and Multi-Language Investigation." *Collabra: Psychology* 3 (1): 1–22.

Montgomery, M., and J. S. Hall. 2004. *Dictionary of Smoky Mountain English*. Knoxville, TN: University of Tennessee Press.

Montrul, S., and T. Ionin. 2012. "Dominant Language Transfer in Spanish Heritage Speakers and Second Language Learners in the Interpretation of Definite Articles." *Modern Language Journal* 96 (1): 70–94.

Moon, C., R. P. Cooper, and W. P. Fifer. 1993. "Two-Day-Olds Prefer Their Native Language." *Infant Behavior and Development* 16 (4): 495–500.

Morgan, J. L., K. M. Bonamo, and L. L. Travis. 1995. "Negative Evidence on Negative Evidence." *Developmental Psychology* 31 (2): 180.

Moyer, A. 1999. "Ultimate Attainment in L2 Phonology." *Studies in Second Language Acquisition* 21 (1): 81–108.

Müller, S. 2002. *Complex Predicates: Verbal Complexes, Resultative Constructions, and Particle Verbs in German*. Studies in Constraint-Based Lexicalism 13. Stanford, CA: Center for the Study of Language and Information.

———. 2007. "Phrasal or Lexical Constructions: Some Comments on Underspecification of Constituent Order, Compositionality, and Control. In *Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar*, edited by S. Müller, 373–93. Stanford, CA: Center for the Study of Language and Information.

Müller, S., and S. Wechsler. 2014. "Lexical Approaches to Argument Structure." *Theoretical Linguistics* 40 (1–2): 1–76.

Munakata, Y. 2001. Graded representations in behavioral dissociations. Trends in cognitive sciences, 5(7), 309-315.

Munakata, Y., J. L. McClelland, M. H., Johnson, and R. S. Siegler. 1997. "Rethinking Infant Knowledge: Toward an Adaptive Process Account of Successes and Failures in Object Permanence Tasks." *Psychological Review* 104 (4): 686.

Muñoz, C., and D. Singleton, 2011. "A Critical Review of Age-Related Research on L2 Ultimate Attainment." *Language Teaching* 44 (1): 1–35.

Münte, T. F., H. J. Heinze, M. Matzke, B. M. Wieringa, and S. Johannes. 1998. "Brain Potentials and Syntactic Violations Revisited: No Evidence for Specificity of the Syntactic Positive Shift." *Neuropsychologia* 36 (3): 217–26.

Murphy, G. L. 2002. *The Big Book of Concepts*. Cambridge, MA: Bradford Books.

Murphy, G. L., and E. J. Wisniewski. 1989. "Categorizing Objects in Isolation and in Scenes: What a Superordinate Is Good For. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15 (4): 572.

Murray, T. E., T. C. Frazer, and B. L. Simon. 1996. "Need + Past Participle in American English." *American Speech* 71 (3): 255–71.

Nagy, W. E., and R. C. Anderson. 1984. "How Many Words Are There in Printed School English?" *Reading Research Quarterly* 19 (3): 304–30.

Naigles, L., A. Fowler, and A. Helm. 1992. "Developmental Shifts in the Construction of Verb Meanings." *Cognitive Development* 7(4): 403–27.

Naigles, L., H. Gleitman, and L. Gleitman. 1993. "Syntactic Bootstrapping and Verb Acquisition." In *Language and Cognition: A Developmental Perspective*, edited by E. Dromi, 104–40. Norwood, NJ: Ablex.

Naigles, L., E. Hoff, and D. Vear. 2009. "Flexibility in Early Verb Use: Evidence from a Multiple-N Diary Study." *Monographs of the Society for Research in Child Development* 74 (2): vii–112.

Nelson, A. B., and R. M. Shiffrin. 2013. "The Co-evolution of Knowledge and Event Memory." *Psychological Review* 120 (2): 356–94.

Newman, E. L., and K. Norman. 2010. "Moderate Excitation Leads to Weakening of Perceptual Representations." *Cerebral Cortex* 20 (11): 2760–70.

Newport, E. L. 1990. "Maturational Constraints on Language Learning." *Cognitive Science* 14 (1): 11–28.

Noble, C. H., C. F. Rowland, and J. M. Pine. 2011. "Comprehension of Argument Structure and Semantic Roles: Evidence from English-Learning Children and the Forced-Choice Pointing Paradigm. *Cognitive Science* 35 (5): 963–82.

Norman, D. A., and D. E. Rumelhart, eds. 1975. *Explorations in Cognition*. San Francisco: Freeman.

Norman, K. A., E. L. Newman, and G. Detre. 2007. "A Neural Network Model of Retrieval-Induced Forgetting." *Psychological Review* 114 (4): 887–953.

Nosofsky, R. M. 1986. "Attention, Similarity, and the Identification–Categorization Relationship." *Journal of Experimental Psychology: General* 115 (1): 39.

Núñez-Peña, M. I., and M. L. Honrubia-Serrano. 2004. "P600 Related to Rule Violation in an Arithmetic Task." *Cognitive Brain Research* 18 (2): 130–41.

O'Donnell, T. 2015. *Productivity and Reuse in Language.* Cambridge, MA: MIT Press.

Oehrle, R. T. 1976. "The Grammatical Status of the English Dative Alternation." PhD thesis, MIT.

Oh, E. 2010. "Recovery from First-Language Transfer: The Second Language Acquisition of English Double Objects by Korean Speakers." *Second Language Research* 26 (3): 407–39.

Oldfield, R. C., and A. Wingfield. 1965. "Response Latencies in Naming Objects." *Quarterly Journal of Experimental Psychology* 17 (4): 273–81.

Ortega, L. 2014. *Understanding Second Language Acquisition*. Abingdon, UK: Routledge.

Osherson, D. N., E. E. Smith, O. Wilkie, A. Lopez, and E. Shafir. 1990. "Category-Based Induction." *Psychological Review* 97 (2): 185.

Osterhout, L., and P. J. Holcomb. 1992. "Event-Related Brain Potentials Elicited by Syntactic Anomaly." *Journal of Memory and Language* 31 (6): 785–806.

Östman, J. O., and G. Trousdale. 2013. "Dialects, Discourse, and Construction Grammar." In *The Oxford Handbook of Construction Grammar*, edited by T. Hoffmann and G. Trousdale, 476–90. Oxford: Oxford University Press.

Palmer, G. B. 1996. *Toward a Theory of Cultural Linguistics*. Austin, TX: University of Texas Press.

Palmeri, T. J., S. D. Goldinger, and D. B. Pisoni. 1993. "Episodic Encoding of Voice Attributes and Recognition Memory for Spoken Words." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19 (2): 309.

Parisien, C., and S. Stevenson. 2010. "Learning Verb Alternations in a Usage-Based Bayesian Model." *Proceedings of the Annual Meeting of the Cognitive Science Society* 32: 540–45.

Partee, B. H. 1970. *On the Requirement that Transformations Preserve Meaning*. [Bloomington, IN]: Indiana University Linguistics Circle.

Paul, H. 1888. *Principles of the History of Language*. London: Swan Sonnenschein, Lowrey.

Pawley, A., and F. H. Syder. 1983. "Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency." In *Language and Communication*, edited by J. C. Richards and R. W. Schmidt, 191–225. London: Longman.

Peng, R. 2016. "The Integration of Exemplars and Prior Knowledge in the Extension of Schematic Constructions: Evidence from Chinese Emerge-Hide Construction." *Language Sciences* 56: 1–29.

Perek, F. 2016. "Using Distributional Semantics to Study Syntactic Productivity in Diachrony: A Case Study." *Linguistics* 54 (1): 149–88.

———. Unpublished MS. "Productivity and Schematicity in Constructional Change." University of Birmingham, UK.

Perek, F., and A. E. Goldberg. 2015. "Generalizing beyond the Input: The Functions of the Constructions Matter." *Journal of Memory and Language* 84: 108–27.

———. 2017. "Linguistic Generalization on the Basis of Function and Constraints on the Basis of Statistical Preemption." *Cognition* 168: 276–93.

Perfors, A., J. B. Tenenbaum, and E. Wonnacott. 2010. "Variability, Negative Evidence, and the Acquisition of Verb Argument Constructions." *Journal of Child Language* 37 (3): 607–42.

Pesetsky, D. M. 1996. *Zero Syntax: Experiencers and Cascades*. Cambridge, MA: MIT Press.

Piantadosi, S. T., H. Tily, and E. Gibson. 2011. "Word Lengths are Optimized for Efficient Communication." *Proceedings of the National Academy of Sciences of the USA* 108 (9): 3526–29.

———. 2012. "The Communicative Function of Ambiguity in Language." *Cognition* 122 (3): 280–91.

Pickering, M. J., and S. Garrod. 2013. "An Integrated Theory of Language Production and Comprehension." *Behavioral and Brain Sciences* 36 (4): 329–47.

Pierrehumbert, J. 2001. "Stochastic Phonology." *Glot International* 5 (6): 195–207.

———. 2002. "Word-Specific Phonetics." *Laboratory Phonology* 7: 101–39.

Pinker, S. 1987. "The Bootstrapping Problem in Language Acquisition." In *Mechanisms of Language Acquisition*, edited by B. MacWhinney, 399–441. Hillsdale, NJ: Lawrence Erlbaum Associates.

———. 1989. *Learnability and Cognition*. Cambridge, MA: MIT Press.

———. 1999. *Words and Rules: The Ingredients of Language.* New York: Hatchet Book Group.

Pinker, S., and A. Prince. 1988. "On Language and Connectionism: Analysis of a Parallel Distributed Processing Model of Language Acquisition." *Cognition* 28 (1): 73–193.

Pinker, S., and M. T. Ullman. 2002. "The Past-Tense Debate: The Past and Future of the Past Tense." *Trends in Cognitive Sciences* 6 (11): 456–63.

Plunkett, K., and V. Marchman. 1991. "U-Shaped Learning and Frequency Effects in a Multi-layered Perception: Implications for Child Language Acquisition." *Cognition* 38 (1): 43–102.

———. 1993. "From Rote Learning to System Building: Acquiring Verb Morphology in Children and Connectionist Nets." *Cognition* 48 (1): 21–69.

Potter, C. 2016. "Learning and Generalization of Multiple Structures by Children and Adults." PhD thesis, University of Madison Psychology Department.

Potter, M. C., and L. Lombardi. 1990. "Regeneration in the Short-Term Recall of Sentences." *Journal of Memory and Language* 29 (6): 633–54.

Poznan, L., L. R. Gleitman, and J. Trueswell. 2016. "Semantic Ambiguity and Syntactic Bootstrapping: The Case of Conjoined-Subject Intransitive Sentences." *Language Learning and Development* 12 (1): 14–41.

Pulvermüller, F., Y. Shtyrov, and R. Ilmoniemi, R. 2005. "Brain Signatures of Meaning Access in Action Word Recognition." *Journal of Cognitive Neuroscience* 17 (6): 884–92.

Pustejovsky, J., ed. 2012. *Semantics and the Lexicon*. Studies in Linguistics and Philosophy 49. Dordrecht, Netherlands: Springer Science and Business Media.

Quine, W. V. 1960. "Carnap and Logical Truth." *Synthese* 12 (4): 350–74.

Ramscar, M. 2002. "The Role of Meaning in Inflection: Why the Past Tense Does Not Require a Rule." *Cognitive Psychology* 45 (1): 45–94.

Ratcliff, R., P. Gomez, and G. McKoon. 2004. "A Diffusion Model Account of the Lexical Decision Task." *Psychological Review* 111 (1): 159.

Raviv, L., and I. Arnon. 2017. "Differences between Children and Adults in the Emergence of Linguistic Structure." Poster presented at the CUNY Conference on Human Sentence Processing, MIT, Cambridge, MA, 30 March–1 April 2017.

Rayner, K., and S. A. Duffy. 1986. "Lexical Complexity and Fixation Times in Reading: Effects of Word Frequency, Verb Complexity, and Lexical Ambiguity." *Memory and Cognition* 14 (3): 191–201.

Rayner, K., and C. J. Springer. 1986. "Graphemic and Semantic Similarity Effects in the Picture: Word Interference Task. *British Journal of Psychology* 77 (2): 207–22.

Reali, F., and M. H. Christiansen. 2007. "Word Chunk Frequencies Affect the Processing of Pronominal Object-Relative Clauses." *Quarterly Journal of Experimental Psychology* 60 (2): 161–70.

Reali, F., and T. L. Griffiths. 2009. "The Evolution of Frequency Distributions: Relating Regularization to Inductive Biases through Iterated Learning." *Cognition* 111 (3): 317–28.

Regel, S., A. Opitz, G. Müller, and A. D. Friederici. 2015. "The Past Tense Debate Revisited: Electrophysiological Evidence for Subregularities of Irregular Verb Inflection." *Journal of Cognitive Neuroscience* 27 (9): 1870–85.

Regier, T., and P. Kay. 2009. "Language, Thought, and Color: Whorf Was Half Right." *Trends in Cognitive Sciences* 13 (10): 439–46.

Rescorla, R. A., and A. R. Wagner. 1972. "A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement." In *Classical Conditioning II: Current Theory and Research*, edited by A. H. Black and W. F. Prokasy, 64–99. New York: Appleton-Century-Crofts.

Rhodes, M., S. A. German, and D. Brinkman. 2010. "Children's Attention to Sample Composition in Learning, Teaching and Discovery." *Developmental Science* 13 (3): 421–29.

Richerson, P. J., and M. H. Christiansen, eds. 2013. *Cultural Evolution: Society, Technology, Language, and Religion*. Cambridge, MA: MIT Press.

Rickford, J. R. 1997. "Prior Creolization of African-American Vernacular English?: Sociohistorical and Textual Evidence from the 17th and 18th Centuries." *Journal of Sociolinguistics* 1 (3): 315–36.

Robenalt, C., and A. E. Goldberg. 2015. "Judgment Evidence for Statistical Preemption: It Is Relatively Better to *Vanish* than to *Disappear* a Rabbit, but a Lifeguard Can Equally Well *Backstroke* or *Swim* Children to Shore." *Cognitive Linguistics* 26 (3): 467–503.

———. 2016. "Nonnative Speakers Do Not Take Competing Alternative Expressions into Account the Way Native Speakers Do." *Language Learning* 66 (1): 60–93.

Rossi, E., M. Diaz, J. F. Kroll, and P. E. Dussias. 2017 "Late Bilinguals Are Sensitive to Unique Aspects of Second Language Processing: Evidence from Clitic Pronouns Word-Order." *Frontiers in Psychology* 8: 342.

Rothschild, G., E. Eban, and L. M. Frank. 2017. “A Cortical–Hippocampal–Cortical Loop of Information Processing during Memory Consolidation.” *Nature Neuroscience* 20 (2): 251.
Rovee-Collier, C. 1997. “Dissociations in Infant Memory: Rethinking the Development of Implicit and Explicit Memory.” *Psychological Review* 104 (3): 467.
Rowland, C. F., and C. L. Noble. 2010. “The Role of Syntactic Structure in Children’s Sentence Comprehension: Evidence from the Dative.” *Language Learning and Development* 7 (1): 55–75.
Rowland, C. F., and J. M. Pine. 2000. “Subject–Auxiliary Inversion Errors and Wh-Question Acquisition: ‘What Children Do Know?’” *Journal of Child Language* 27 (1): 157–81.
Rowland, C. F., J. M. Pine, E. V. Lieven, and A. L. Theakston. 2003. “Determinants of Acquisition Order in Wh-Questions: Re-evaluating the Role of Caregiver Speech.” *Journal of Child Language* 30 (3): 609–35.
Russek, E. M., I. Momennejad, M. M. Botvinick, S. J. Gershman, and N. D. Daw. 2017. “Predictive Representations Can Link Model-Based Reinforcement Learning to Model-Free Mechanisms.” *PLoS Computational Biology* 13 (9): e1005768.
Rutherford, W. E. 1989. “Preemption and the Learning of L2 Grammars.” *Studies in Second Language Acquisition* 11 (4): 441–57.
Sabourin, L., L. A. Stowe, and G. J. de Haan. 2006. “Transfer Effects in Learning a Second Language Grammatical Gender System.” *Second Language Research* 22 (1): 1–29.
Sachs, J. S. 1967. “Recognition Memory for Syntactic and Semantic Aspects of Connected Discourse.” *Perception and Psychophysics* 2 (9): 437–42.
Saffran, J. R., S. D. Pollak, R. L. Seibel, and A. Shkolnik. 2007. “Dog Is a Dog Is a Dog: Infant Rule Learning Is Not Specific to Language.” *Cognition* 105 (3): 669–80.
Samara, A., K. Smith, H. Brown, and E. Wonnacott. 2017. “Acquiring Variation in an Artificial Language: Children and Adults Are Sensitive to Socially Conditioned Linguistic Variation.” *Cognitive Psychology* 94: 85–114.
Savage, C., E. Lieven, A. Theakston, and M. Tomasello. 2003. “Testing the Abstractness of Children’s Linguistic Representations: Lexical and Structural Priming of Syntactic Constructions in Young Children.” *Developmental Science* 6 (5): 557–67.
Saxton, M. 1997. “The Contrast Theory of Negative Input.” *Journal of Child Language* 24 (1): 139–61.
Schlichting, M. L., and A. R. Preston. 2016. “Hippocampal-Medial Prefrontal Circuit Supports Memory Updating during Learning and Post-Encoding Rest.” *Neurobiology of Learning and Memory* 134: 91–106.
Schmid, H. J., ed. 2017. *Entrenchment, Memory and Automaticity: The Psychology of Linguistic Knowledge and Language Learning*. Berlin: Mouton de Gruyter.
Schuler, K., C. Yang, and E. Newport. 2016. “Testing the Tolerance Principle: Children Form Productive Rules when It Is More Computationally Efficient to Do So.” *Proceedings of the Annual Meeting of the Cognitive Science Society* 38: 2321–26.
Schultz, W., P. Dayan, and P. R. Montague. 1997. “A Neural Substrate of Prediction and Reward.” *Science* 275 (5306): 1593–99.
Schwab, J. F., and C. Lew-Williams. 2016. “Language Learning, Socioeconomic Status, and Child-Directed Speech.” *Wiley Interdisciplinary Reviews: Cognitive Science* 7 (4): 264–75.
Schwab, J. F., C. Lew-Williams, and A. E. Goldberg. 2018. “When Regularization Gets It Wrong: Children Over-Simplify Language Input Only in Production.” *Journal of Child Language*, 21 February 2018. doi:10.1017/S0305000918000041.
Schwartz, B. D., and R. A. Sprouse. 1996. “L2 Cognitive States and the Full Transfer/Full Access Model.” *Second Language Research* 12 (1): 40–72.

Sedivy, J. C., M. K. Tanenhaus, C. G. Chambers, and G. N. Carlson. 1999. "Achieving Incremental Semantic Interpretation through Contextual Representation." *Cognition* 71 (2): 109–47.
Selinker, L., and U. Lakshmanan. 1992. "Language Transfer and Fossilization: The Multiple Effects Principle." In *Language Transfer in Language Learning*, rev. ed., edited by S. M. Gass and L. Selinker, 197–216. Amsterdam: John Benjamins.
Sethuraman, N., A. E. Goldberg, and J. Goodman. 1997. "Using the Semantics Associated with Syntactic Frames for Interpretation without the Aid of Non-linguistic Context." In *Proceedings of the Twenty-eighth Annual Child Language Research Forum*, edited by E. Clark, 283–94. Stanford, CA: Center for the Study of Language and Information.
Seuren, P. A. 1990. "Serial Verb Constructions." In *When Verbs Collide: Papers from the Ohio State Mini-Conference on Serial Verbs*, edited by B. D. Joseph and A. M. Zwicky, 14–32. Columbus, OH: Ohio State University Department of Linguistics.
Shibatani, M. 1976. *The Grammar of Causative Constructions*. Syntax and Semantics 6. New York: Academic Press.
Siegelman, N., and I. Arnon. 2015. "The Advantage of Starting Big: Learning from Unsegmented Input Facilitates Mastery of Grammatical Gender in an Artificial Language." *Journal of Memory and Language* 85: 60–75.
Siewierska, A., and W. B. Hollmann. 2007. "Ditransitive Clauses in English with Special Reference to Lancashire Dialect." In *Structural-Functional Studies in English Grammar*, edited by M. Hannay and G. J. Steen, 83–102. Amsterdam: John Benjamins.
Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
Singleton, J. L., and E. L. Newport. 2004. "When Learners Surpass Their Models: The Acquisition of American Sign Language from Inconsistent Input." *Cognitive Psychology* 49 (4): 370–407.
Siskind, J. M. 1996. "A Computational Study of Cross-Situational Techniques for Learning Word-to-Meaning Mappings." *Cognition* 61 (1–2): 39–91.
Slobin, D. I. 1977. "Language Change in Childhood and in History." In *Language Learning and Thought: Perspectives in Neurolinguistics and Psycholinguistics*, edited by J. Macnamara, 185–214. New York: Academic Press.
———. 1996a. "From 'Thought and Language' to 'Thinking for Speaking.'" In *Rethinking Linguistic Relativity*, edited by J. J. Gumperz and S. C. Levinson, 70–96. Cambridge, UK: Cambridge University Press.
———. 1996b. "Two Ways to Travel: Verbs of Motion in English and Spanish." In *Grammatical Constructions: Their Form and Meaning*, edited by M. Shibatani and S. A. Thompson, 195–217. Oxford: Oxford University Press.
———. 2004. "The Many Ways to Search for a Frog: Linguistic Typology and the Expression of Motion Events." In *Relating Events in Narrative*, vol. 2, *Typological and Contextual Perspectives*, edited by S. Strömqvist and L. Verhoven, 219–57. Mahwah, NJ: Lawrence Erlbaum Associates.
Smith, K., A. Perfors, O. Fehér, A. Samara, K. Swoboda, and E. Wonnacott. 2017. "Language Learning, Language Use and the Evolution of Linguistic Variation." *Philosophical Transactions of the Royal Society of London, B* 372 (1711): 20160051.
Smith, K., and E. Wonnacott. 2010. "Eliminating Unpredictable Variation through Iterated Learning." *Cognition* 116 (3): 444–49.
Smith, S. M. 1979. "Remembering in and out of Context." *Journal of Experimental Psychology: Human Learning and Memory* 5 (5): 460–71.
Snow, C. E., and M. Hoefnagel-Höhle. 1978. "The Critical Period for Language Acquisition: Evidence from Second Language Learning." *Child Development* 49 (4): 1114–28.

Soja, N. N., S. Carey, and E. S. Spelke. 1991. "Ontological Categories Guide Young Children's Inductions of Word Meaning: Object Terms and Substance Terms." *Cognition* 38 (2): 179–211.

Stallings, L. M., M. C. MacDonald, and P. G. O'Seaghdha. 1998. "Phrasal Ordering Constraints in Sentence Production: Phrase Length and Verb Disposition in Heavy-NP Shift." *Journal of Memory and Language* 417: 392–417.

Standing, L. 1973. "Learning 10,000 Pictures." *Quarterly Journal of Experimental Psychology* 25 (2): 207–22.

Stearns, L. 2012. "Watermelon, Honeydew, and Antelope: An ERP Study of Semantically Anomalous but Phonologically Expected Words in Sentences." Honors thesis collection, Wellesley College.

Steels, L. 2005. "The Emergence and Evolution of Linguistic Structure: From Lexical to Grammatical Communication Systems." *Connection Science* 17 (3–4): 213–30.

Stefanowitsch, A. 2008. "Negative Entrenchment: A Usage-Based Approach to Negative Evidence." *Cognitive Linguistics* 19 (3): 513–31.

Stefanowitsch, A., and S. T. Gries. 2003. "Collostructions: Investigating the Interaction of Words and Constructions." *International Journal of Corpus Linguistics* 8 (2): 209–43.

———. 2009. "Corpora and Grammar." In *Corpus Linguistics: An International Handbook*, vol. 2, edited by A. Lüdeling and M. Kytö, 933–51. Berlin: Walter de Gruyter.

Stephens, G. J., L. J. Silbert, and U. Hasson. 2010. "Speaker–Listener Neural Coupling Underlies Successful Communication." *Proceedings of the National Academy of Sciences of the USA* 107 (32): 14425–30.

Stevens, J. S., L. R. Gleitman, J. C. Trueswell, and C. Yang. 2017. "The Pursuit of Word Meanings." *Cognitive Science* 41 (s4): 638–76.

Stickgold, R. 2005. "Sleep-Dependent Memory Consolidation." *Nature* 437 (7063): 1272–78.

Storm, B. C., and B. J. Levy. 2012. "A Progress Report on the Inhibitory Account of Retrieval-Induced Forgetting." *Memory and Cognition* 40 (6): 827–43.

Strapp, C. M., D. M. Bleakney, A. L. Helmick, and H. M. Tonkovich. 2008. "Developmental Differences in the Effects of Negative and Positive Evidence." *First Language* 28 (1): 35–53.

Sung, M. C., and H. K. Yang. 2016. "Effects of Construction Centered Instruction on Korean Students' Learning of English Transitive Resultative Constructions." 56. In *Applied Construction Grammar*, Applications of Cognitive Linguistics 32, edited by S. de Knop and G. Gilquin, 89–114. Berlin: Mouton de Gruyter.

Suttle, L., and A. E. Goldberg. 2011. "The Partial Productivity of Constructions as Induction." *Linguistics* 49 (6): 1237–69.

Sutton, R. S., and A. G. Barto. 1998. *Reinforcement Learning: An Introduction*. Vol. 1. Cambridge, MA: MIT Press.

Swingley, D., J. P. Pinto, and A. Fernald. 1999. "Continuous Processing in Word Recognition at 24 Months." *Cognition* 71 (2): 73–108.

Tachihara, K., and A. E. Goldberg. 2018. "L2 Speakers Know what They've Heard but They Don't Take Competing Alternatives into Account the Same Way as L1 Speakers Do." Poster presented at CUNY Human Sentence Processing Conference, Davis, CA, 15–17 March 2018.

Tagliamonte, S. A., and A. D'Arcy. 2009. "Peaks beyond Phonology: Adolescence, Incrementation, and Language Change." *Language* 85 (1): 58–108.

Talmy, L. 1985. "Lexicalization Patterns: Semantic Structure in Lexical Forms." *Language Typology and Syntactic Description* 3 (99): 36–149.

———. 2003. *Toward a Cognitive Semantics*. Vol. 1. Cambridge, MA: MIT Press.

Tanenhaus, M. K., M. J. Spivey-Knowlton, K. M. Eberhard, and J. C. Sedivy. 1995. "Integration of Visual and Linguistic Information in Spoken Language Comprehension." *Science* 268 (5217): 1632–34.

Tenny, C. 1998. "Psych Verbs and Verbal Passives in Pittsburghese." *Linguistics* 36: 591–98.

Tesnière, L. 1953. *Esquisse d'une syntaxe structurale*. Paris: Klincksieck.

Theakston, A. L. 2004. "The Role of Entrenchment in Children's and Adults' Performance on Grammaticality Judgment Tasks." *Cognitive Development* 19: 15–34.

Theakston, A. L., P. Ibbotson, D. Freudenthal, E. V. Lieven, and M. Tomasello. 2015. "Productivity of Noun Slots in Verb Frames." *Cognitive Science* 39 (6): 1369–95.

Theakston, A. L., E. V. Lieven, J. M. Pine, and C. F. Rowland. 2001. "The Role of Performance Limitations in the Acquisition of Verb-Argument Structure: An Alternative Account." *Journal of Child Language* 28 (1): 127–52.

Thompson, S. A. 1990. "Information Flow and Dative Shift in English Discourse." In *Development and Diversity*, edited by J. A. Edmondson, C. Feagin, and P. Mühlhäusler, 239–53. Dallas, TX: Summer Institute of Linguistics.

———. 1995. "The Iconicity of 'Dative Shift' in English: Considerations from Information Flow in Discourse." In *Syntactic Iconicity and Linguistic Freezes: The Human Dimension*, edited by M. E. Landsberg, 155–75. Berlin: Mouton de Gruyter.

Thompson, S. A., and P. J. Hopper. 2001. "Transitivity, Clause Structure, and Argument Structure: Evidence from Conversation." In *Frequency and the Emergence of Linguistic Structure*, Typological Studies in Language 45, edited by J. Bybee and P. Hopper, 27–60. Amsterdam: John Benjamins.

Thothathiri, M., and M. G. Rattinger. 2016. "Acquiring and Producing Sentences: Whether Learners Use Verb-Specific or Verb-General Information Depends on Cue Validity." *Frontiers in Psychology* 7: 404.

Tily, H., S. Gahl, I. Arnon, N. Snider, A. Kothari, and J. Bresnan. 2009. "Syntactic Probabilities Affect Pronunciation Variation in Spontaneous Speech." *Language and Cognition* 1 (2): 147–65.

Tomasello, M. 1992. *First Verbs: A Case Study of Early Grammatical Development*. Cambridge, UK: Cambridge University Press.

———. 2000. "Do Young Children Have Adult Syntactic Competence?" *Cognition* 74 (3): 209–53.

———. 2001. "Perceiving Intentions and Learning Words in the Second Year of Life." In *Language Acquisition and Conceptual Development*, edited by M. Bowerman and S. Levinson, 132–58. Cambridge, UK: Cambridge University Press.

———. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

———. 2009. *Why We Cooperate.* Cambridge, MA: MIT Press.

———. 2016. "The Ontogeny of Cultural Learning." *Current Opinion in Psychology* 8: 1–4.

Tomasello, M., N. Akhtar, K. Dodson, and L. Rekau. 1997. "Differential Productivity in Young Children's Use of Nouns and Verbs." *Journal of Child Language* 24 (2): 373–87.

Tomasello, M., and M. E. Barton. 1994. "Learning Words in Nonostensive Contexts." *Developmental Psychology* 30 (5): 639.

Tomasello, M., and K. Haberl. 2003. "Understanding Attention: 12- and 18-Month-Olds Know What Is New for Other Persons." *Developmental Psychology* 39 (5), 906–12.

Traugott, E. C. 2015. "Toward a Coherent Account of Grammatical Constructionalization." In *Diachronic Construction Grammar*, edited by J. Barðdal, E. Smirnova, L. Sommerer, and S. Gildea, 51–80. Amsterdam: John Benjamins.

Traugott, E. C., and G. Trousdale. 2013. *Constructionalization and Constructional Changes.* Vol. 6. Oxford: Oxford University Press.

Tremblay, A., B. Derwing, G. Libben, and G. Westbury. 2011. "Processing Advantages of Lexical Bundles: Evidence from Self-Paced Reading and Sentence Recall Tasks." *Language Learning* 61: 569–613.

Trudgill, P. 2011. *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford: Oxford University Press.

Trueswell, J. C., T. N. Medina, A. Hafri, and L. R. Gleitman. 2013. "Propose but Verify: Fast Mapping Meets Cross-Situational Word Learning." *Cognitive Psychology* 66: 126–56.

Trueswell, J. C., M. K. Tanenhaus, and C. Kello. 1993. "Verb-Specific Constraints in Sentence Processing: Separating Effects of Lexical Preference from Garden-Paths." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19 (3): 528.

Tsao, N., and D. Wible. 2013. "Word Similarity Using Constructions as Contextual Features." In *Proceedings of the Joint Symposium on Semantic Processing: Textual Inference and Structures in Corpora*, edited by A. Lavelli and O. Popescu, 51–59. Trento, Italy: SIGLEX.

Tuggy, D. 1993. "Ambiguity, Polysemy and Vaguess." *Cognitive Linguistics* 4 (3): 273–90.

Tummeltshammer, K., and D. Amso. 2017. "Top-Down Contextual Knowledge Guides Visual Attention in Infancy." *Developmental Science*, 26 October 2017. doi:10.1111/desc.12599.

Ullman, M. T. 2001. "The Declarative/Procedural Model of Lexicon and Grammar." *Journal of Psycholinguistic Research* 30 (1): 37–69.

Uther, M., M. A. Knoll, and D. Burnham. 2007. "Do You Speak E-NG-LI-SH? A Comparison of Foreigner- and Infant-Directed Speech." *Speech Communication* 49 (1): 2–7.

Vaidya, C. J., M. Zhao, J. E. Desmond, and J. D. Gabrieli. 2002. "Evidence for Cortical Encoding Specificity in Episodic Memory: Memory-Induced Re-activation of Picture Processing Areas." *Neuropsychologia* 40 (12): 2136–43.

Van de Meerendonk, N., H. H. Kolk, D. J. Chwilla, and C.T.W. Vissers. 2009. "Monitoring in Language Perception." *Language and Linguistics Compass* 3 (5): 1211–24.

Van Trijp, R., and L. Steels. 2012. "Multilevel Alignment Maintains Language Systematicity." *Advances in Complex Systems* 15 (03n04): 1250039.

Von Humboldt, W. (1832) 1999. *On Language: The Diversity of Human Language-Structure and Its Influence on the Mental Development of Mankind*. Cambridge, UK: Cambridge University Press.

Vygotsky, L. S. 1978. "Interaction between Learning and Development." In *Mind in Society: The Development of Higher Psychological Processes*, edited by M. Cole, V. John-Steiner, S. Scribner, and E. Souberman, 79–91. Cambridge, MA: Harvard University Press.

Wagner, S. E., and S. A. Tagliamonte. 2016. "Vernacular Stability: Comparative Evidence from Two Lifespan Studies." Paper presented at New Ways of Analyzing Variation (NWAV) 45, Vancouver, 3–6 November 2016. Abstract 161.

Walker, A., and J. Hay. 2011. "Congruence between 'Word Age' and 'Voice Age' Facilitates Lexical Access." *Laboratory Phonology* 2 (1): 219–37.

Walther, D. B., E. Caddigan, L. Fei-Fei, and D. M. Beck. 2009. "Natural Scene Categories Revealed in Distributed Patterns of Activity in the Human Brain." Journal of *Neuroscience* 29 (34): 10573–81.

Wasow, T. 2002. *Postverbal Behavior.* Stanford, CA: Center for Study of Language and Information.

Wasow, T., T. F. Jaeger, and D. Orr. 2011. "Lexical Variation in Relativizer Frequency." In *Expecting the Unexpected: Exceptions in Grammar*, Trends in Linguistics: Studies and Monographs 216, edited by H. J. Simon and H. Wiese, 175–96. Berlin: Mouton de Gruyter.

Waxman, S. R., and D. B. Markow. 1995. "Words as Invitations to Form Categories: Evidence from 12-to 13-Month-Old Infants." *Cognitive Psychology* 29 (3): 257–302.

Webelhuth, G., and C. J. Dannenberg. 2006. "Southern American English Personal Datives: The Theoretical Significance of Dialectal Variation." *American Speech* 81 (1): 31–55.

Werker, J. F., J. H. Gilbert, K. Humphrey, and R. C. Tees. 1981. "Developmental Aspects of Cross-Language Speech Perception." *Child Development* 52 (1): 349–55.

Werker, J. F., and R. C. Tees. 1984a. "Cross-Language Speech Perception: Evidence for Perceptual Reorganization during the First Year of Life." *Infant Behavior and Development* 7 (1): 49–63.

———. 1984b. "Phonemic and Phonetic Factors in Adult Cross-Language Speech Perception." *Journal of the Acoustical Society of America* 75 (6): 1866–78.

Wible, D., and N. L. Tsao. 2017. "Constructions and the Problem of Discovery: A Case for the Paradigmatic." *Corpus Linguistics and Linguistic Theory*, 8 July 2017. doi:10.1515/cllt-2017-0008.

Williams, E. 1994. "Remarks on Lexical Knowledge." *Lingua* 92: 7–34.

———. 2007. "Dumping Lexicalism." In *The Oxford Handbook of Linguistic Interfaces*, edited by G. Ramchand and C. Reiss, 353–82. Oxford: Oxford University Press.

Willits, J. A., M. S. Amato, and M. C. MacDonald. (2015). "Language Knowledge and Event Knowledge in Language Use." *Cognitive Psychology* 78: 1–27.

Wilson, F. 2009. "Processing at the Syntax-Discourse Interface in Second Language Acquisition." PhD thesis, University of Edinburgh.

Wittgenstein, L. 1953. *Philosophical Investigations*. New York: Macmillan.

Wonnacott, E., J. K. Boyd, J. Thomson, and A. E. Goldberg. 2012. "Input Effects on the Acquisition of a Novel Phrasal Construction in 5 Year Olds." *Journal of Memory and Language* 66 (3): 458–78.

Wonnacott, E., H. Brown, and K. Nation. 2017. "Skewing the Evidence: The Effect of Input Structure on Child and Adult Learning of Lexically-Based Patterns in an Artificial Language." *Journal of Memory and Language* 95: 36–48.

Wonnacott, E., E. L. Newport, and M. K. Tanenhaus. 2008. "Acquiring and Processing Verb Argument Structure: Distributional Learning in a Miniature Language." *Cognitive Psychology* 56 (3): 165–209.

Woodard, K., L. R. Gleitman, and J. C. Trueswell. 2016. "Two- and Three-Year-Olds Track a Single Meaning during Word Learning: Evidence for Propose-but-Verify." *Language Learning and Development* 12 (3): 252–61.

Woodward, A. L., E. M. Markman, and C. M. Fitzsimmons. 1994. "Rapid Word Learning in 13- and 18-Month-Olds." *Developmental Psychology* 30: 553–66.

Wray, A. 2002. *Formulaic Language and the Lexicon*. Cambridge, UK: Cambridge University Press.

Wulff, S., A. Stefanowitsch, and S. T. Gries. 2007. "Brutal Brits and Persuasive Americans: Variety-Specific Meaning Construction in the *Into*-Causative." In *Aspects of Meaning Construction*, edited by G. Radden, K.-M. Köpcke, and T. Berg, 265–81. Amsterdam: John Benjamins.

Xu, F. 2002. "The Role of Language in Acquiring Object Kind Concepts in Infancy." *Cognition* 85 (3): 223–50.

Xu, F., and J. B. Tenenbaum. 2007. "Word Learning as Bayesian Inference." *Psychological Review* 114 (2): 245.

Yang, C. D. 2016. *The Price of Linguistic Productivity: How Children Learn to Break the Rules of Language*. Cambridge, MA: MIT Press.
Yoshida, H., and L. B. Smith. 2008. “What’s in View for Toddlers? Using a Head Camera to Study Visual Experience.” *Infancy* 13 (3): 229–48.
Yuan, S., C. Fisher, and J. Snedeker. 2012. “Counting the Nouns: Simple Structural Cues to Verb Meaning.” *Child Development* 83 (4): 1382–99.
Yurovsky, D., C. Yu, and L. B. Smith. 2013. “Competitive Processes in Cross-Situational Word Learning.” *Cognitive Science* 37 (5): 891–921.
Zanuttini, R., and J. B. Bernstein. 2014. “Transitive Expletives in Appalachian English.” In *Micro-syntactic Variation in North American English*, edited by R. Zanuttini and L. R. Horn, 143–77. Oxford: Oxford University Press.
Zeldes, A. 2013. *Productivity in Argument Selection: From Morphology to Syntax*. Trends in Linguistics, Studies and Monographs 260. Berlin: Mouton de Gruyter.
Zeschel, A. 2012. *Incipient Productivity: A Construction-Based Approach to Linguistic Creativity*. Cognitive Linguistics Research 49. Berlin: Mouton de Gruyter.
Zettersten, M., and G. Lupyan. 2018. “Finding Categories through Words: More Nameable Features Improve Category Learning.” *PsyArXiv*, 2 March 2018. doi:10.17605/OSF.IO/UZ2M9.
Zettersten, M., E. Wojcik, V. L. Benitez, and J. Saffran. 2018. “The Company Objects Keep: Linking Referents Together during Cross-Situational Word Learning.” *Journal of Memory and Language* 99: 62–73.
Zhang, X., and C. Mai. 2018. “Effects of Entrenchment and Preemption in Second Language Learners’ Acceptance of English Denominal Verbs.” *Applied Psycholinguistics* 39 (2): 413–36.
Ziegler, J., A. Kocab, and J. Snedeker. 2017. “The Effect of Population Size on Intergenerational Language Convergence: An Artificial Language Learning Paradigm.” Poster presented at the 42nd Boston University Conference on Language Development, Boston, MA, November 2017.
Zwicky, A. M. 1971. “In a Manner of Speaking.” *Linguistic Inquiry* 2 (2): 223–33.
Zwitserlood, P. 1989. “The Locus of the Effects of Sentential-Semantic Context in Spoken-Word Processing.” *Cognition* 32 (1): 25–64.

Index