SSLA, **24**, 143–188
DOI: 10.1017.S0272263102002024

FREQUENCY EFFECTS IN LANGUAGE PROCESSING

A Review with Implications for Theories of Implicit and Explicit Language Acquisition

Nick C. Ellis
University of Wales, Bangor

This article shows how language processing is intimately tuned to input frequency. Examples are given of frequency effects in the processing of phonology, phonotactics, reading, spelling, lexis, morphosyntax, formulaic language, language comprehension, grammaticality, sentence production, and syntax. The implications of these effects for the representations and developmental sequence of SLA are discussed. Usage-based theories hold that the acquisition of language is exemplar based. It is the piecemeal learning of many thousands of constructions and the frequency-biased abstraction of regularities within them. Determinants of pattern productivity include the power law of practice, cue competition and constraint satisfaction, connectionist learning, and effects of type and token frequency. The regularities of language emerge from experience as categories and prototypical patterns. The typical route of emergence of constructions is from formula, through low-scope pattern, to construction. Frequency plays a large part in explaining sociolinguistic variation and language change. Learners' sensitivity to frequency in all these domains has implications for theories of implicit and explicit learning and their interactions. The review concludes by considering the history of frequency as an explanatory concept in theoretical and applied linguistics, its 40 years of exile, and its necessary reinstatement as a bridging variable that binds the different schools of language acquisition research.

Address correspondence to: Nick C. Ellis, School of Psychology, University of Wales, Bangor, Gwynedd LL57 2AS, United Kingdom; e-mail: n.ellis@bangor.ac.uk.

© 2002 Cambridge University Press 0272-2631/02 $9.50

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

"Perception is of definite and probable things" (James, 1890, p. 82).

There is a lot more to the perception of language than meets the eye or ear. A percept is a complex state of consciousness in which antecedent sensation is supplemented by consequent ideas that are closely combined to it by association. The cerebral conditions of the perception of things are thus the paths of association irradiating from them. If a certain sensation is strongly associated with the attributes of a certain thing, that thing is almost sure to be perceived when we get that sensation. Where the sensation is associated with more than one reality, however, unconscious processes weigh the odds, and we perceive the most probable thing: "*all brain-processes are such as give rise to what we may call* FIGURED *consciousness*" (James, 1890, p. 82, emphasis in original). Accurate and fluent language perception, then, rests on the comprehender having acquired the appropriately weighted range of associations for each element of the language input.

Psycholinguistic and cognitive linguistic theories of language acquisition hold that all linguistic units are abstracted from language use. In these usage-based perspectives, the acquisition of grammar is the piecemeal learning of many thousands of constructions and the frequency-biased abstraction of regularities within them. Language learning is the associative learning of representations that reflect the probabilities of occurrence of form-function mappings. Frequency is thus a key determinant of acquisition because "rules" of language, at all levels of analysis (from phonology, through syntax, to discourse), are structural regularities that emerge from learners' lifetime analysis of the distributional characteristics of the language input. Learners have to *figure* language out.

This review illustrates (a) how frequency underpins regularity effects in the acquisition of orthographic, phonological, and morphological form, and (b) that learning accords to the power law of practice. It shows, for example, that there are effects of bigram frequency in visual word identification and of phonotactic knowledge in speech segmentation, effects of spelling-to-sound correspondences in reading, and cohort effects in spoken-word recognition. There are effects of neighbors and the proportion of friends (items that share surface-pattern cues and have the same interpretation) to enemies (items that share surface-pattern cues but have different interpretations) in reading and spelling, morphology, and spoken-word recognition. At higher levels, it can be shown that language comprehension is determined by the listeners' vast amount of statistical information about the behavior of lexical items in their language and that, at least for English, verbs provide some of the strongest constraints on the resolution of syntactic ambiguities. Comprehenders know the relative frequencies with which individual verbs appear in different tenses, in active versus passive structures and in intransitive versus transitive structures, the typical kinds of subjects and objects that a verb takes, and many other such facts. Such information is acquired through experience with input that exhibits these distributional properties. It is not some idiosyncratic fact

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

in the lexicon isolated from "core" grammatical information; rather, it is relevant at all stages of lexical, syntactic, and discourse comprehension. Comprehenders tend to perceive the most probable syntactic and semantic analyses of a new utterance on the basis of frequencies of previously perceived utterance analyses. Language users tend to produce the most probable utterance for a given meaning on the basis of frequencies of utterance representations. These accounts readily contribute to explanations of sociolinguistic variation.

The review discusses how these effects can be simulated using mathematical or computational models. The effects of frequency in input are modulated by the need to simultaneously satisfy the constraints of all other constructions that are represented in the learner's system. The interactions of input and existing representation can be described as Bayesian interactions in a rich network of interacting associations and connections, some competing, others mutually reinforcing as a result of the many redundancies of language and representation. Recent work shows that in syntax, as in phonology, the productivity of pattern depends on type frequency of the construction. The implications for theories of SLA are described: a developmental sequence—from formula, through low-scope pattern, to construction—is proposed as a useful starting point to investigate the emergence of constructions and the ways in which type and token frequency affect the productivity of patterns.

The review finishes with consideration of the consequences of exemplar views of language for theories of implicit and explicit learning. To the extent that language processing is based on frequency and probabilistic knowledge, language learning is implicit learning. This does NOT deny the importance of *noticing* (Schmidt, 1993) in the initial registration of a pattern-recognition unit. NOR does it deny a role for explicit instruction. Language acquisition can be speeded by explicit instruction. The last 20 years of empirical investigations into the effectiveness of L2 instruction demonstrate that focused L2 instruction results in large target-oriented gains, that explicit types of instruction are more effective than implicit types, and that the effectiveness of L2 instruction is durable. An outline is given of the mechanisms by which explicit knowledge affects implicit learning.

FREQUENCY LEARNING

Humans are sensitive to the frequencies of events in their experience. Ask them to make explicit judgments from memory about the relative frequency with which things happen and they are typically pretty good at it. College students, for example, can accurately estimate the frequency with which words occur in a list (Hasher & Chromiak, 1977) and can also estimate, with a high correlation to actual counts, the frequency of English words (Shapiro, 1969), of single letters (Attneave, 1953), and even of pairs of letters (Underwood, 1971). There are few individual differences in these abilities; children are just about as good as young adults, and this ability remains robust in the face of

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

various neurological insults or terminal decline (Hasher & Zacks, 1984). As Hasher and Chromiak state,

> That we can rank order events with as seemingly little meaning as bigrams suggests that the processing of frequency may fall into the domain of what Posner & Snyder (1975) have called "automatic processes." That is, of processes which the organism runs off both without any awareness of the operation, with no intention of doing so, and with little effort, in the sense that the tagging of frequency has little impact on one's ability to simultaneously attend to other aspects of a situation, such as the interpretation of an ongoing conversation. (p. 173)

It does not seem like we spend our time counting the units of language. We are instead conscious of communicating, and yet in the course of conversation we naturally acquire knowledge of the frequencies of the elements of language and their mappings. The mechanism underlying such unconscious counting is to be found in the plasticity of synaptic connections rather than abacuses or registers, but it constitutes counting nevertheless.

The experiments listed here all involve explicit judgments made on the basis of prior implicit learning (i.e., acquisition of knowledge about the underlying structure of a complex stimulus environment by a process that takes place naturally, simply, and without conscious operations [Ellis, 1994b]). It is fair comment that, however natural the original learning, the outcome task is as artificial a request as asking someone the number of days it rains in September or the weight of a bucket of water. It is more naturalistic simply to observe their behavioral responses to environmental contingencies—for example, to see how often they take their raincoats or how they prepare to heft. When one observes language knowledge like this, as the psycholinguistic demonstrations that I next review will attest, it is clear just how well listeners are tuned to the frequencies of their language input at all levels. In contrast to traditional psychology experiments or school tests that involve explicit memory (involving a conscious process of remembering a prior episodic experience or fact such as questions like *What did you have for breakfast?* or *What is the capital of Andorra?*), the outcome measures in these studies involve implicit memory (i.e., facilitation of the processing of a stimulus as a function of encounters with an identical or related stimulus but where the subject at no point has to consciously recall the prior event). Human excellence is often evidenced as implicit memory. Consider, as examples, the sublime Olympic gymnastic skills that result from lifetime practice and the ubiquitous but nonetheless equally impressive human abilities to categorize: We have never been instructed in the defining features of birds, English words, cups, mugs, or grammatical sentences, yet we are fast and accurate at the classification of these elements of our world.

Indeed, it is human categorization ability that provides the most persuasive testament to our incessant unconscious figuring. We know that natural categories are fuzzy rather than nomothetic. Wittgenstein's (1953) consideration of

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

the concept "game" showed that no set of features covers all the things that we call games, ranging from soccer, through chess, bridge, and poker, to solitaire. Instead, what organizes these exemplars into the game category is a set of family resemblances among these members; son may be like mother, and mother like sister, but in a very different way. We learn about these families, like our own, from experience. Exemplars are similar if they have many features in common and few distinctive attributes (features belonging to one but not the other); the more similar two objects are on these quantitative grounds, the faster people are at judging them to be similar (Tversky, 1977). Prototypes, as exemplars that are most typical of a category, are similar to many members of that category and not similar to members of other categories. Again, the operationalization of this criterion predicts the speed of human categorization performance; people more quickly classify as birds sparrows (or other average sized, average colored, average beaked, average featured specimens) than they do birds with less common features or feature combinations like geese or albatrosses (Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Prototypes are judged faster and more accurately, even if they themselves have never been seen before. For example, someone who has never seen a sparrow, yet who has experienced the rest of the run of the avian mill, will still be fast and accurate in judging it to be a bird (Posner & Keele, 1970). Of course, frequency and dimensional central tendency are not the only factors that determine activation of candidate schemata; there are moderating effects of recency of use and of context (Anderson, 1980, chap. 5; Harnad, 1987). Even so, it is very clear that although people do not go around consciously counting features, they nevertheless have very accurate knowledge of the underlying frequency distributions and their central tendencies. We are very good at this association learning—it is the raw basis of human cognition.

The understanding of human categorization ability is one of the major research enterprises within cognitive science. It is the subject of many different investigations using computational simulations that count features and feature combinations in different ways. These include connectionist, semantic network, prototype, schema, and exemplar models. In these models, it is usual that each learning event updates a statistical representation of a category independently of other learning events. Exemplar-based models have multiple instances in memory, one for each exemplar. Prototype models count features and derive statistical abstractions of the central tendencies. In connectionist models, each repetition increases the strength of the connections between the relevant feature units and the category unit. One long-standing question is whether this counting increments in units of events or individuals, of tokens or types. Traditional theories of perception and classification have it that token events constitute the basic unit of categorization and that as people categorize each token they store a memory of every such categorization event. When assessing whether a subsequent entity belongs to the same category, people retrieve these memories, assess their similarity to the new entity, and

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

admit the entity to the category if sufficiently similar. Such event-based approaches count tokens not types; they do not allow the cognitive system to track the same individual across multiple events nor assume that the cognitive system stores or uses information about individual types in classification decisions. However, there is growing interest in different models of categorization that are affected by type rather than by token frequency. In token-counting models of face recognition, a perceiver's prototype of the category of human faces would be as affected by the *n*-thousandth glimpse of a friend as it would the first of a stranger. In models that count types, however, renewed sighting of the familiar friend would have little or no effect, whereas experience of a new individual type would sway the balance of the central tendency. Recently, Barsalou, Huttenlocher, and Lamberts (1998) reported a series of experiments demonstrating that human subjects track individuals across events, establish representations of them, and use these representations in categorization. They argued that these findings are consistent with "representationalism," the view that an internal model of the world constitutes a physical level of representation in the brain and that the brain does not simply capture the statistical properties of events in an undifferentiated dynamic system. Thus, human categorization may be more sensitive to type frequency than to simple token frequency. This is a key issue for language acquisition, and a later section of this paper is accordingly devoted to it.

Type or token units, exemplar, prototype, or connectionist mechanism, these are importantly different variants of figuring, but it is all counting, one way or another, and it is all unconscious. What of the language census?

PHONOLOGY AND PHONOTACTICS

Speakers do not try to learn phonotactics. First-year linguistics undergraduates complain about the difficulty in explicitly analyzing the phonology of one's native language. Yet they are very good at judging whether nonwords are nativelike or not, and young children are sensitive to these regularities when trying to repeat nonwords (Treiman & Danis, 1988). Phonotactic competence simply emerges from using language, from the primary linguistic data of the lexical patterns that a speaker knows. Frisch, Large, Zawaydeh, and Pisoni (2001) asked native speakers to judge nonword stimuli, using a 7-point rating scale, for whether they were more or less like English words. The nonwords were created with relatively high or low probability licit phonotactic patterns as determined by the logarithm of the product of probabilities of the onset and rime constituents of the nonword. The mean wordlikeness judgments for these nonword stimuli had an extremely strong relationship with expected probability ($r = .87$). An emergentist account of phonotactic competence is thus that any new nonword is compared to the exemplars that are in memory; the closer it matches their characteristics, the more wordlike it is judged. Frisch et al. explained individual differences in the same way: Participants with relatively larger mental lexicons were more likely to judge low probabil-

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

ity nonwords as well formed, which suggests that well-formedness is determined by a lexicon-based probability cutoff such that the larger a speaker's lexicon, the more low-frequency words that have low-frequency phonotactic patterns, and the more likely it is that the speaker knows an exemplar like the nonword stimulus in some way. Well-formedness judgments for English show that the probability of a novel word, given the frequency distribution of phonological constituents in the listener's lexicon, provides a foundation for phonotactic well-formedness.

It is no wonder that adults are sensitive to the sequential dependencies of language: They have been tallying them since they were infants in order to segment and recognize connected speech. The linguistic knowledge acquired during the first year is largely the result of distributional analyses of the input, including the vowel space, consonant categories, phonotactic rules, phonological regularities, and frequent biphones of the native language (Jusczyk, 1997). Infants automatically analyze simple distributional statistics in the language they hear, to discover the word boundaries (Brent, 1999; Elman, 1990). Saffran, Aslin, and Newport (1996) demonstrated that 8-month-old infants exposed for only 2 minutes to unbroken strings of nonsense syllables (e.g., *bidakupado*) are able to detect the difference between three-syllable sequences that appeared as a unit and sequences that also appeared in their learning set but in random order. These infants achieved this learning on the basis of statistical analysis of phonotactic sequence data, right at the age when their caregivers start to notice systematic evidence of their recognizing words. In subsequent experiments, Saffran, Johnson, Aslin, and Newport (1999) exposed participants to continuous nonlinguistic auditory sequences whose elements were organized into so-called tone words. As in their previous studies, statistical information was the only word-boundary cue available to learners. Both adults and infants succeeded at segmenting the tone stream, with performance indistinguishable from that obtained with syllable streams. These results suggest that the same learning mechanism previously shown to be involved in word segmentation can also be used to segment sequences of nonlinguistic stimuli, putting these results firmly within the broad experimental area of sequence learning that is the research base of theories of implicit learning (Ellis, 1994a; Reber, 1993; Stadler & Frensch, 1998).

However, there are other additional cues to word boundaries beyond sequential probabilities, including pauses at the end of utterances and intonational cues. One important insight from connectionist explorations of cue utility is the demonstration that, although any one source of information may be insufficient to produce a solution, the combined effect of multiple cue sources can dramatically improve performance (Redington & Chater, 1998). Thus, for the example of segmenting words from speech, Christiansen, Allen, and Seidenberg (1998) showed that phonotactic information on its own is enough to produce 47% accuracy in word segmentation, whereas utterance-boundary information and relative stress information alone produce even less accurate results. However, when these three sources of cues are combined in

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

a connectionist simulation, performance levels exceed 70% accuracy. A count in one dimension alone may not solve a categorization problem, just as the height of a pot is insufficient to resolve whether it is a cup or a mug. Categorization research demonstrates that our unconscious tallying takes place on many multiple dimensions at once. In other words, our unconscious statistics greatly surpasses the simple univariate stuff—it runs to multidimensional scaling and beyond. Infants, similarly, can abstract and generalize beyond specific word order to show evidence of category-based abstraction and the combination of multiple sources of information in artificial-language learning experiments (see Gómez & Gerken, 2000, for review).

A final example of our implicit knowledge of phonotactics is found in cohort effects in spoken-word recognition. We seem to handle this by using the initial phoneme of a word to activate the set of all words in the lexicon that have this same phoneme. Then, as the speech signal unfolds over time and more information is received, we narrow the set down. In the cohort model of speech recognition (Marslen-Wilson, 1990), activation in the cohort varies so that items are not simply "in" or "out." Rather, higher frequency words get more activation from the same evidence than do low-frequency words. This assumption provides a means for accounting for lexical similarity effects whereby a whole neighborhood of words is activated but the higher frequency words get more activation. The result is that listeners are slower to recognize low-frequency words with high-frequency neighbors because the competitors are harder to eliminate (Lively, Pisoni, & Goldinger, 1994). These effects demonstrate that our language-processing system is sensitive both to the frequency of individual words and to the number of words that share the same beginnings (at any length of computation).

READING AND SPELLING

Another area in which language learners have been shown to be sensitive to the frequencies and consistencies of mappings is in relating symbols and their sounds while reading aloud. To the extent that readers are able to construct the correct pronunciations of novel words or nonwords, they have been said to be able to apply sublexical "rules" or mappings that relate graphemes to phonemes (Coltheart, Curtis, Atkins, & Haller, 1993; Patterson & Morton, 1985) or larger orthographic units to their corresponding rimes or syllables (Ehri, 1998; Glushko, 1979; Goswami, 1999; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995). It is likely to be the operation of this system that explains why regular or consistent words are read better than irregular or inconsistent words. For the case of adult fluency in English, words with regular spelling-sound correspondences (e.g., *mint*) are read with shorter naming latencies and lower error rates than words with exceptional correspondences (cf. *pint*) (Coltheart, 1978). In development, exception words (e.g., *blood*, *bouquet*) are acquired later than are regular words (e.g., *bed*, *brandy*) (Coltheart & Leahy, 1996). Similarly, in fluent performance, words that are consistent in

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

their pronunciation in that they are spoken like their neighbors, who share orthographic body and phonological rime (*best* is regular and consistent in that all *-est* bodies are pronounced in the same way), are named faster than inconsistent items (*mint* is regular in terms of its grapheme-phoneme conversion [GPC] rule, but inconsistent in that it has *pint* as a neighbor) (Glushko). The magnitude of the consistency effect for any word depends on the summed frequency of its friends (having similar spelling pattern and similar pronunciation) in relation to that of its enemies (having similar spelling pattern but dissimilar pronunciation) (Jared, McRae, & Seidenberg, 1990). Adult naming latency decreases monotonically with increasing consistency on this measure (Taraban & McClelland, 1987). In development, Laxon, Masterson, and Coltheart (1991) have shown that, for regular words, consistent (*pink*, all *-ink*) and consensus (*hint*, mostly as in *mint*, but cf. *pint*) items are acquired earlier than ambiguous ones (*cove* vs. *love*, *move*). Similarly, irregular words—those in deviant gangs in which the several items sharing that rime are all pronounced in the same irregular fashion (e.g., *look*, *book*, *cook*, etc., or *calm*, *balm*, *palm*)—are acquired earlier than ambiguous ones (*love*). Because of the power law of learning, these effects of regularity and consistency are more evident with low-frequency words than with high-frequency ones on which performance is closer to asymptote (Seidenberg, Waters, Barnes, & Tanenhaus, 1984). As with the learning of other quasiregular language domains, these effects of consistency or ambiguity of spelling-sound correspondence within language have been successfully simulated in connectionist (Harm & Seidenberg, 1999; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989; Zorzi, Houghton, & Butterworth, 1998) and exemplar-based (Ellis & Hicks, 2000) computational models.

These investigations compared the learnability and processing of words of different degrees of spelling-sound ambiguity within a language. What about the larger crosslinguistic issue: What are the effects of the overall ambiguity of a language's symbol-sound mappings on its speakers' rate of literacy acquisition? The orthographies of languages like Finnish, Italian, Welsh, Spanish, Dutch, Turkish, and German are on the whole much more transparent than those of opaque languages like English and French. In transparent orthographies, the mappings from letters to sounds are consistent. In opaque orthographies, the same grapheme may represent different phonemes in different words and, as illustrated for English in the preceding paragraph, there are many words that are irregular in terms of default grapheme-phoneme mapping. These language differences in overall orthographic transparency have a determining effect on rate of reading acquisition. In alphabetic languages it is commonly believed that there is a prolonged alphabetic stage of reading in which words are decoded on the basis of learned symbol-sound associations and that this provides the practice that allows the eventual development of skilled orthographic reading abilities (e.g., Ehri, 1979, 1998; Ellis & Cataldo, 1990; Frith, 1985; Marsh, Friedman, Welch, & Desberg, 1981). An orthographic-transparency hypothesis therefore predicts that children learning to read a

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

transparent orthography in which sound-symbol mappings are regular and consistent should learn to read and spell faster than those learning an opaque orthography in which the cues to pronunciation are more ambiguous. Empirical research supports this prediction. For example, children learning to read German are more able to read their transparent orthography instantiation of pairs of translation equivalents (e.g., *Pflug-plough*) than are matched learners of English (Landerl, Wimmer, & Frith, 1997). Spanish children are able to read more of a sample of eight monosyllabic and eight disyllabic words in their language than are matched French or English children (Goswami, Gombert, & de Barrera, 1998). Turkish children are able to read and spell with high degrees of accuracy by the end of their first grade (Öney & Durgunoglu, 1997). Finally, Welsh children are able to accurately read aloud significantly more of their language than are English children matched for reading instruction, background, locale, and mathematical ability (Ellis & Hooper, in press). These crosslinguistic effects of orthographic transparency demonstrate that where the statistics are simpler, the figuring is faster.

LEXIS

Just as we are sensitive to the combinatorial sequential probabilities of sublexical units, so, too, the recognition and production of words is a function of their frequency of occurrence in the language. For written language, high-frequency words are named more rapidly than low-frequency ones (Balota & Chumbly, 1984; Forster & Chambers, 1973), they are more rapidly judged to be words in lexical decision tasks (Forster, 1976), and they are spelled more accurately (Barry & Seymour, 1988). Auditory word recognition is better for high-frequency than low-frequency words (Luce, 1986; Savin, 1963). Kirsner (1994) has shown that there are strong effects of word frequency on the speed and accuracy of lexical recognition processes (speech perception, reading, object naming, and sign perception) and lexical production processes (speaking, typing, writing, and signing), in children and adults as well as in L1 and L2. He proposed a life-span practice model to explain these frequency effects whereby processing speed can be explained simply by reference to the power law of learning (Anderson, 1982; Ellis & Schmidt, 1997; Newell, 1990), which is generally used to describe the relationships between practice and performance in the acquisition of a wide range of cognitive skills. That is, the effects of practice are greatest at early stages of learning, but they eventually reach asymptote. We may not be counting the words as we listen or speak, but each time we process one there is a reduction in processing time that marks this practice increment, and thus the perceptual and motor systems become tuned by the experience of a particular language.

Jusczyk (1993, 1997) argued that exemplar-based models of memory are as relevant to the origins of representations in infants as they are to adult language processing (e.g., Goldinger, 1996, 1998; Hintzman, 1986, 1988; Jacoby & Brooks, 1984). In these models, memories are formed for every speech input

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

experience and these traces retain detail (e.g., speaker's voice characteristics and context). Computer modeling, for example with Minerva 2 (Hintzman), shows how abstraction is an automatic consequence of aggregate activation of high-frequency exemplars, with regression toward central tendencies as numbers of highly similar exemplars increase. Thus, there is a single-voice advantage (in that words repeated in the same voice are better recognized than those in a different voice), and this advantage is greater for low-frequency words; "old" words that have been frequently experienced in various places by a variety of speakers inspire "abstract" echoes, obscuring context and voice elements of the study trace (Goldinger, 1998, p. 255). Vihman (2000) proposed a frequency-sensitive episodic processing and exemplar learning account of the transition from analogic echoic lexical representation to segmental phonology.

MORPHOSYNTAX

The conclusion reached regarding orthography was that systems that are noisy and inconsistent are harder to figure out than those that are reliable and categorical. There is now a large body of research demonstrating that greater ambiguity in the mappings between the forms and functions of language causes less successful learning because of a larger degree of competition among the cues in the learning set (Bates & MacWhinney, 1987; MacWhinney, 1987). Some cues are more reliable than others, and the language learner's task is to work out the most valid predictors. The Competition Model shows how Bayesian cue use can result in the activation of a single interpretative hypothesis in resolution of interacting cue—some conflicting, some agreeing—that vary in their frequency, reliability, and validity (MacWhinney, 1997c, 1999).

These effects have been well evidenced in the domain of morphosyntax. Consider, for example, how the acquisition of grammatical gender is determined by the degree of transparency of its morphophonological marking. Brooks, Braine, Catalano, Brody, and Sudhalter (1993) demonstrated that children and adults showed better learning of the noun subclasses of artificial languages when ambiguity was reduced through the presence of a subset of nouns of each subclass sharing a phonological feature than they did in a condition in which the phonological features were less reliable cues in distinguishing the subclasses. Taraban and Roark (1996) manipulated the ambiguity in the mapping of noun forms onto genders in two sets of French nouns. They showed that learning the same set of feminine nouns took longer if the nouns in the masculine class were as a set more ambiguous in the mappings of their noun endings onto gender. This demonstration is important because it illustrates how the presence of nontransparent marking not only affects the speed at which the nontransparent items themselves are acquired but also slows down the learning of the whole system. Recent studies have simulated language-learning data using simple connectionist models that relate cues and

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

their functional interpretations. For example, the simulations of Kempe and MacWhinney (1998) showed why the Russian case inflection system is acquired more rapidly than is that of German: Even though case marking in Russian is more complex than in German, the Russian inflections are more reliable cues to sentence interpretation.

It has been claimed that there are only effects of frequency of past-tense forms when fluent speakers generate irregular morphological forms and that these are not to be seen with regular inflections. The early relevant experiments were conducted by Prasada, Pinker, and Snyder (1990) and Seidenberg and Bruck (1990), who showed that when fluent native English speakers saw verb stems on a screen and were required to produce the past-tense form as quickly as possible, they took significantly less time (16–29 ms in three experiments) for irregular verbs with high past-tense frequencies (e.g., *went*) than for irregular verbs with low past-tense frequencies (e.g., *slung*), even when stem frequencies were equated. However, in these experiments there was no effect on latency of the past-tense frequency of regular verbs whose past tense was generated by adding *-ed*. Ellis and Schmidt (1997) argued that this apparent interaction is simply a result of the power law of learning—that is, that frequency effects are not easily observed for regular forms because performance on these is close to asymptote. The same argument has been put forward by Plaut et al. (1996) for the case of reading, and the general conclusion follows that frequency by regularity interactions are a natural and necessary result of the power law of practice. To verify this for morphology, Ellis and Schmidt measured production of regular and irregular forms as learners practiced an artificial second language where regularity and frequency were factorially combined. Accuracy and latency data demonstrated frequency effects for both regular and irregular forms early in the acquisition process. However, as learning progressed, the frequency effect on regular items diminished whereas it remained the same for irregular items. Other studies have confirmed this effect of frequency on the latency of fluent adult production of regular morphological forms (Stemberger & MacWhinney, 1986) and on the accuracy of this process in children with specific language impairment (Oetting & Rice, 1993).

Morphological processing, like reading and listening, also shows effects of neighbors and false friends where regular, inconsistent items (e.g., *bake-baked* is similar in rime to neighbors *make-made* and *take-took*, which have inconsistent past tenses) are produced more slowly than entirely regular ones such as *hate-hated*, *bate-bated*, *date-dated* (Daugherty & Seidenberg, 1994; Seidenberg & Bruck, 1990). I realize I am risking bathos in driving home the point about consciousness, but if I ask you to quickly render the sentence *Sam mixes the dough, bakes the bread, and calls the family for breakfast* in the past tense, I will warrant that the last thing that comes to your mind while dealing with the second verb is: "Hmm, *bake* rhymes with several verbs that have irregular preterit forms, I'd better be careful in its production."

The acquisition of morphology has been the major proving ground for connectionist models of learning in quasiregular domains. The pioneers were

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

Rumelhart and McClelland (1986), who showed that a simple learning model reproduced the characteristics of young children learning the morphology of the past tense in English to a remarkable degree. The model generated the so-called U-shaped learning curve for irregular forms; it exhibited a tendency to overgeneralize, and, in the model as in children, different past-tense forms for the same word could coexist at the same time. Yet there was no "rule," as the authors note: "It is possible to imagine that the system simply stores a set of rote-associations between base and past tense forms with novel responses generated by 'on-line' generalizations from the stored exemplars" (p. 267). This original past-tense model was very influential: It laid the foundations for the connectionist approach to language research; it generated a large number of criticisms (Pinker & Prince, 1988; Lachter & Bever, 1988), some of which are undeniably valid; and, in turn, it spawned a number of revised and improved connectionist models of different aspects of the acquisition of the English past tense. These recent models successfully capture the regularities that are present in associating phonological form of lemma with phonological form of inflected form (Daugherty & Seidenberg, 1994; MacWhinney & Leinbach, 1991; Marchman, 1993; Plunkett & Marchman, 1993), and between referents (+past tense or +plural) and associated inflected perfect or plural forms (Cottrell & Plunkett, 1994; Ellis & Schmidt, 1997), closely simulating the error patterns, profiles of acquisition, differential difficulties, false-friends effects, reaction times for production, and interactions of regularity and frequency that are found in human learners (both L1 and L2), and, by abstracting the regularities from experienced exemplars, they successfully apply the default case to "wug" items in generalization tests. They therefore strongly support the notion that acquisition of morphology is also a result of simple associative learning principles operating in a massively distributed system abstracting the statistical regularities of association using optimal inference.

FORMULAIC LANGUAGE PRODUCTION

Just as we learn the common sequences of sublexical components of our language, the tens of thousands of phoneme and letter sequences large and small, so also we learn the common sequences of words. Formulas are lexical chunks that result from binding frequent collocations. Large stretches of language are adequately described by finite state grammars as collocational streams where patterns flow into each other. Sinclair (1991), then director of the Cobuild project, the largest lexicographic analysis of the English language to date, summarized this in the Principle of Idiom:

> A language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments. To some extent this may reflect the recurrence of similar situations in human affairs; it may illustrate a natural tendency to economy of effort; or it may be motivated in part by the exigencies of real-time conversation. (p. 110)

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

Instead of a rather minor feature, compared with grammar, Sinclair suggests that for normal texts, the first mode of analysis to be applied is the idiom principle, as most text is interpretable by this principle. Most of the material that Sinclair was analyzing in the Bank of English was written text. Comparisons of written and spoken corpora demonstrate that collocations are even more frequent in spoken language (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Brazil, 1995; Leech, 2000). Parole is flat and Markovian because it is constructed "off the top of one's head" and there is no time to work it over. Utterances are constructed as intonation units, and substantive units are fairly strongly constrained to have a modal length of four words in English, a fact indicative of the cognitive limits on how much information can be fully active in the mind at any one time (Chafe, 1994). The majority of substantive intonational units have the grammatical form of single clauses, though many others are parts of clauses, and they are often highly predictable in terms of their lexical concordance (Hopper, 1998). Language reception and production are mediated by learners' representations of chunks of language:

> Suppose that, instead of shaping discourse according to rules, one really pulls old language from memory (particularly old language, with all its words in and everything), and then reshapes it to the current context: "context shaping," as Bateson puts it, "is just another term for grammar." (Becker, 1983, p. 218)

Frequency of collocation determines the growth of lexical phrases. These play out as frequency effects within the realm of idioms, too. High-frequency idioms are easier to comprehend than less familiar ones because their figurative meanings are known and these, by dint of practice, have become more salient than their literal ones (Giora, 1997; Nippold & Rudzinski, 1993). The greater the working-memory demands of the processing task, the greater the need to rely on formulas: "It is easier for us to look something up than to compute it" (Bresnan, 1999). There is no more enjoyable demonstration of this principle than the analyses by Kuiper (1996) of so-called smooth talkers—that is, sports commentators and auctioneers who are in communicative contexts that place an inordinate pressure to observe what is transpiring around them, to analyze these happenings in short-term memory, and to formulate speech reports describing what is observed in real time without getting left behind. Smooth talkers use many formulas in their speech, such as recurrent sequences of verbal behavior, whether conventional or idiosyncratic, which are sequentially and hierarchically organized. The faster the action, the more difficult it is for the commentator to provide an instantaneous commentary. By contrasting slow-action commentators (cricket, real estate auctions) with faster-action commentators (horse races, antique and livestock auctions), Kuiper showed that the faster-action commentators made greater use of formulas than did the slower-action ones. Horses for courses: The faster commentaries of race-casters are highly formulaic, easily recognized, exciting, and communicatively instantaneous in their effect. Compare this with reporting

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

that is as rich as the accompanying fruit cake when cricket commentators slowly conjure their descriptions of the pigeons on the pitch during another comforting rain interval.

The importance of such lexical units or idiomatic phrases is acknowledged in SLA research under discussion of holophrases (Corder, 1973), prefabricated routines and patterns (Hakuta, 1974), formulaic speech (Wong-Fillmore, 1976), memorized sentences and lexicalized stems (Pawley & Syder, 1983), formulas (R. Ellis, 1994), sequences in SLA (N. Ellis, 1996b, 2001), discourse management (Dörnyei & Kormos, 1998; Tannen, 1987), register (Biber & Finegan, 1994), style (Brewster, 1999), and lexical patterns and collocational knowledge (Carter, 1998; Hoey, 1991; Lewis, 1993; Schmitt, 2000). As Nattinger (1980) pointed out, "for a great deal of the time anyway, language production consists of piecing together the ready-made units appropriate for a particular situation and . . . comprehension relies on knowing which of these patterns to predict in these situations" (p. 341). So too did Pawley and Syder:

> In the store of familiar collocations there are expressions for a wide range of familiar concepts and speech acts, and the speaker is able to retrieve these as wholes or as automatic chains from the long-term memory; by doing this he minimizes the amount of clause-internal encoding work to be done and frees himself to attend to other tasks in talk-exchange, including the planning of larger units of discourse. (p. 192)

The "store of familiar collocations" of the native language speaker is very large indeed. The sheer number of words and their patterns variously explains why language learning takes so long, why it requires exposure to authentic sources, and why there is so much current interest in corpus linguistics in SLA (Biber, Conrad, & Reppen, 1998; Collins Cobuild, 1996; Hunston & Francis, 1996; McEnery & Wilson, 1996). Nativelike competence, fluency, and idiomaticity require an awful lot of figuring out which words go together.

LANGUAGE COMPREHENSION

Corpus linguistics, using powerful computers for statistical and distributional analysis of large language corpora, has demonstrated the importance of the idiom principle and has catalogued the majority of the multitude of underlying chunks of language and the ways in which lexical items, with their particular valences and subcategorization requirements, operate within these patterns. Language learners use their heads to figure out the same things, and their understanding of language is driven by this lexical knowledge. The last 20 years of linguistics demonstrates a trend whereby theories of grammar have increasingly put more syntax into the lexicon and correspondingly less into rules. The last 20 years of psycholinguistics have encouraged even greater emphasis on the lexical cues that guide sentence interpretation (McKoon & Ratcliff, 1998). The Competition Model (Bates & MacWhinney, 1987; MacWhinney, 1987, 1997c) emphasizes lexical functionalism, in which syntactic patterns are

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

controlled by lexical items. Lexical items provide cues to functional interpretations for sentence comprehension or production. Some cues are more reliable than others. The language learner's task is to work out which are the most valid predictors. The Competition Model is the paradigmatic example of constraint-satisfaction accounts of language comprehension.

Consider the particular cues that relate subject-marking forms to subject-related functions in the English sentence *The learner counts the words*. They are preverbal positioning (*learner* before *counts*), verb agreement morphology (*counts* agrees in number with *learner* rather than *words*), sentence-initial positioning, and use of the article *the*. Case-marking languages, unlike English, would additionally include nominative and accusative cues in such sentences. The corresponding functional interpretations include actor, topicality, perspective, givenness, and definiteness. Competition Model studies analyze a corpus of exemplar sentences that relate such cue combinations with their various functional interpretations to determine the regularities of the ways in which a particular language expresses, for example, agency. They then demonstrate how well these probabilities determine (a) cue use when learners process that language and (b) cue acquisition. The ease of learning an inflection is determined by its cue validity, a function of how often an inflection occurs as a cue for a certain underlying function (cue availability) and how reliably it marks this function (cue reliability) (MacWhinney, 1997c). This present section will concentrate on cue use in language comprehension.

Consider the sentences in (1) beginning with *The plane left for the*. . . . Does the second word refer to a geometric element, an airplane, or a tool? Does the third imply a direction, or is it the past tense of the verb *leave* in active or in passive voice?

(1) a. *The plane left for the East Coast.*
    b. *The plane left for the reporter was missing.*

What of the likelihood of the past-tense passive interpretation of *left* in sentence (2)? Is it greater or less than that for sentence (1b) (Seidenberg, 1997)?

(2) *The note left for the reporter was missing.*

Psycholinguistic experiments show that fluent adults resolve such ambiguities by rapidly exploiting a variety of probabilistic constraints derived from previous experience. There is the first-order frequency information to the effect that *plane* is much more frequent in its vehicle than its other possible meanings and *left* is used more frequently in active rather than passive voice. Thus, the ambiguity is strongly constrained by the frequency with which the ambiguous verb occurs in transitive and passive structures, of which reduced relative clauses are a special type (MacDonald, 1994; MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell, 1996). Additionally, there are combinatorial constraints to the effect that *plane* is an implausible modifier of the noun

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

left, so *plane left* is not a high-probability noun phrase, and sentence (2) is therefore more easy to comprehend as a reduced relative clause than sentence (1b) because it is much more plausible for a note to be left than to leave. In this fashion, interpretation is also constrained by combinatorial lexical information (MacDonald; Tabossi, Spivey-Knowlton, McRae, & Tanenhaus, 1994; Trueswell, Tanenhaus, & Garnsey, 1994). The math is complicated because the combination of constraints is nonlinear in that manipulations of noun agency or discourse context can successfully promote the rarer, reduced relative interpretation only when properties of the ambiguous verb make this interpretation a viable one (MacDonald et al.).

Consider the interpretation of prepositional phrases as a second general class of example of parsing as constraint satisfaction.

(3) *The spy saw the cop with the binoculars.*

In (3), the PP *with the binoculars* can be interpreted as modifying either the verb, such that it is an instrument of seeing, or the direct object NP, such that the spy saw the cop who had the binoculars rather than some other cop (Seidenberg & MacDonald, 1999). Corpus data suggest that about 60% of such PPs modify the direct object NP, and the remaining 40% modify the verb (Collins & Brooks, 1995; Hindle & Rooth, 1993). Thus any simple structural metric that resolved the ambiguity on the basis of overall frequency bias (e.g., Mitchell, Cuetos, Corley, & Brysbaert, 1995) will make errors in the initial interpretation of this ambiguity approximately 40% of the time. The Minimal Attachment metric used in the best-known structural account, the Garden Path model (Frazier, 1987), initially chooses the verb attachment and therefore makes about 60% errors. Yet human comprehenders figure out this ambiguity quite easily. What information are they using to achieve this? There is the first-order frequency information: Many prepositions do not modify nouns and verbs with equal frequency (e.g., *of* almost exclusively modifies NPs, as in *bottle of beer*, *tally of the likelihoods*). Choosing the verb versus noun attachment based solely on the frequency bias of the preposition will resolve the ambiguity correctly about 72% of the time (Collins & Brooks, 1995). Other key lexical information comes from the verb. For example, perception verbs (*see*, *hear*, *feel*, etc.) are not usually modified with instruments or manners, and they thus bias toward the noun-modification interpretation of the PP. Action verbs, which usually collocate with a PP expressing the instrument or manner of the action (e.g., *write with a pen*, *eat in a hurry*), promote the interpretation in which the PP modifies the verb (Spivey-Knowlton & Sedivy, 1995; Taraban & McClelland, 1988). There are also additional, weaker cues provided by the direct object NP: Definite NPs promote the verb-modification interpretation of the PP whereas indefinite NPs promote the noun-modification interpretation (Spivey-Knowlton & Sedivy). Again, such constraints combine in a nonlinear manner: The weaker noun definiteness constraint has little effect when a verb strongly promotes the verb-modification interpretation, but its effects can be

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

clearly seen when the verb is one of perception. The surprise that results from breaking the usual frequency-driven expectancies in language processing underpins both the processing regressions necessitated by garden-path sentences and the humor of puns like those that open and close this paper in which idiomatic meaning overtakes literal interpretation.

Just as frequent analyses are preferred and are easy to process, so it is clear that less frequent constructions are more difficult because they are less predictable and less well practiced. In the aptly named article "Learning and losing syntax: Practice makes perfect and frequency builds fortitude," St. John and Gernsbacher (1998) analyzed why passive and cleft-object constructions are more difficult to learn (de Villiers & de Villiers, 1985), are harder to comprehend, and break down more easily under processing stress (e.g., the very rapid presentation conditions of Miyake, Carpenter, & Just, 1994) or brain damage (Bates, Frederici, & Wolfeck, 1987) than active and cleft-subject constructions. St. John and Gernsbacher proposed that one reason is simply that they are less frequent and concomitantly less practiced. They modeled this frequency effect by using a simple recurrent connectionist model (St. John & McClelland, 1990) and trained it on the four sentence constructions (simple active, simple passive, cleft-subject, and cleft-object) employed by Miyake et al., with, in different simulations, each construction trained more frequently than the others. Generalization to new sentences was high, which demonstrates mastery of the syntax rather than memorization of the training instances, and the high-frequency construction was mastered first and proved more robust under simulated brain damage.

These psycholinguistic studies of sentence processing show that fluent adults have a vast statistical knowledge about the behavior of the lexical items of their language. They know the strong cues provided by verbs, in English at least, in the interpretation of syntactic ambiguities. Fluent comprehenders know the relative frequencies with which particular verbs appear in different tenses, in active versus passive and in intransitive versus transitive structures, the typical kinds of subjects and objects that a verb takes, and many other such facts. This knowledge has been acquired through experience with input that exhibits these distributional properties and through knowledge of its semantics. This information is not just an aspect of the lexicon, isolated from "core" syntax; rather, it is relevant at all stages of lexical, syntactic, and discourse comprehension (Seidenberg & MacDonald, 1999). Frequent analyses are preferred to less frequent ones.

THE VARIABLE NATURE OF "GRAMMATICALITY"

The gold standard empirical test in linguistics traditionally involves judgments of grammatical acceptability: "One way to test the adequacy of a grammar proposed for L is to determine whether or not the sequences that it generates are actually grammatical, i.e., acceptable to a native speaker, etc." (Chomsky, 1957, p. 13). Chomsky went on to write:

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

> Any grammar of a language will *project* the finite and somewhat accidental corpus of observed utterances to a set (presumably infinite) of grammatical utterances. In this respect, a grammar mirrors the behavior of the speaker who, on the basis of a finite and accidental experience with language, can produce or understand an indefinite number of new sentences. (p. 15, emphasis in original)

Grammaticality judgments rest on the introspection of language users who have abstracted information about the language from their particular finite and accidental experience of it. One can make some nomothetic conclusions from the application of this measure across speakers (e.g., *Furiously sleep ideas green colorless* might well warrant a firm, 100% solid * from any speaker of English), yet it is clear that there is marked variation across individuals in other respects, which is why this measure is used to investigate sociolinguistic variation (Labov, 1975) and the interlanguage of second language learners (e.g., Davies & Kaplan, 1998; Mandell, 1999). As Labov (1969) put it:

> Today, a great many linguists study English through their own intuitions; they operate "out of their heads" in the sense that they believe that they can ask and answer all the relevant questions themselves. But even if a teacher comes from the same background as his students, he will find his grammar has changed: that he no longer has sound intuitions about whether he can say *Nobody don't know nothing about it* instead of *Nobody knows nothing about it.* (p. 8)

It seems that the marked individual differences concerning judgments of grammatical acceptability directly reflect the patterns that have been abstracted from the particular frequencies evidenced in each learner's finite and particular accidental experience. I will return to this issue in the later section on sociolinguistics.

Nor are individuals constant in their opinions. The subsequently very large literature on the nature and utility of grammaticality judgments and other linguistic intuitions demonstrates the instability and unreliability of such discriminations (Schütze, 1996) even in fluent L1 speakers. Sometimes this is because, as in (4), the combined cue strengths do not fall clearly one way or another.

(4) a. *Tom is one of those clumsy people who cuts himself shaving.*
    b. *Tom is one of those clumsy people who cuts themself shaving.*
    c. *Tom is one of those clumsy people who cut themselves shaving.*
    d. *other*

Instability, however, also comes from the fluid, ever-changing nature of each individual's system, wherein frequency patterns change as experience accumulates and where the strength of a particular pattern can be temporarily boosted as a result of a recent encounter; recall that the three key factors that determine the activation of candidate schema are frequency, recency, and

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

context. Grammatical categorization is no different from other classifications in these respects (Lakoff, 1987, p. 58; Langacker, 1987; Taylor, 1998). The knowledge of a speaker-hearer cannot be understood as a fixed grammar but rather as a statistical ensemble of language experiences that changes slightly every time a new utterance is processed. R. Ellis (1999) used these ideas, as expressed in N. Ellis (1996b), to explain free variation in L2 learner language.

There are several compelling demonstrations of standard implicit memory effects in the priming of grammaticality judgments and of speech production. The first evidence of syntactic priming came from Bock (1986), who showed that using a particular syntactic pattern in one's speech primes use of that pattern in subsequent speech. On each priming trial, participants were cued to produce a priming sentence in one of several syntactic forms. Several intervening trials later they then viewed an unrelated event in a picture and were asked to describe it in a sentence. The probability of a particular syntactic form being used in the description increased when that form had occurred in the prime, and this happened under presentation conditions that minimized the participants' attention to their speech, to the syntactic features of the priming sentences, and to connections between the priming sentences and the subsequent pictures. These effects of priming are specific to features of sentence form and independent of sentence content, the priming of particular lexical items, thematic roles, or word sequences (Bock & Loebell, 1988). Subsequent replications and extensions of this research demonstrated that these effects are long lasting and can persist over at least 10 intervening sentences (Bock & Griffin, 2000) and intervals of at least 20 minutes (Boyland & Anderson, 1998), thus showing that syntactic priming operates on the same time scale as implicit learning (Pickering, Branigan, Cleland, & Stewart, 2000). Fluent language users tend to produce the most probable utterance for a given meaning on the basis of the frequencies and recencies of utterance representations.

Evidence of syntactic priming of grammaticality judgments comes from Luka and Barsalou (1998) and Luka (1999), who showed that recently read sentences are judged to be more grammatically acceptable. In their first experiment, sentences that had been read for content during an earlier part of the experiment were judged as more grammatically acceptable during a subsequent test phase, a finding that demonstrates an implicit memory or the "mere exposure" effect (Bornstein, 1989) for identically repeated natural language stimuli. A subsequent experiment investigated whether this priming effect generalized across priming and test sentences that were syntactically similar as defined by a phrase-structural description but that did not have the same content words. The priming effect was found for related syntactic structures containing novel content words, and thus the priming effect observed in the first experiment was not due to simple lexical repetition.

These standard, implicit learning and memory effects for grammatical constructions show that competence as assessed using grammaticality judgments seems hardly more constant than grammatical performance. Both are affected by frequency and recency of use of construction. Grammatical analyses that a

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

language user has frequently or recently experienced are preferred to analyses that must be newly constructed.

SYNTAX

The sensitivity of morphosyntax, language comprehension, production, and grammaticality to patterns of frequency of usage has important implications for the structure of the grammatical system—namely, that representations must have variable strengths reflective of their frequency, and connections must similarly be variable in weight.

In defining the nature of grammaticality and in introducing the two now-renowned exemplar test items in (5) and (6), Chomsky (1957) argued:

(5) *Colorless green ideas sleep furiously.*

(6) *Furiously sleep ideas green colorless.*

> The notion "grammatical in English" cannot be identified in any way with the notion "high order of statistical approximation to English." It is fair to assume that neither sentence (5) nor (6) (nor indeed any part of these sentences) has ever occurred in an English discourse. Hence, in any statistical model for grammaticalness, these sentences will be ruled out on identical grounds as being equally 'remote' from English. Yet (5), though nonsensical, is grammatical, while (6) is not. . . . Evidently, one's ability to produce and recognize grammatical utterances is not based on notions of statistical approximation and the like. (pp. 15–16)

This passage is the *locus classicus* for why generative grammars have traditionally paid no heed to the statistics of language. The conclusion is later summarized: "The notion of grammaticalness cannot be identified with meaningfulness (nor does it have any special relation, even approximate, to the notion of statistical order of approximation)" (p. 106). Yet in the footnotes to this defining text, things are left rather more open:

> Given the grammar of a language, one can study the use of the language statistically in various ways; and the development of probabilistic models for the use of language (as distinct from the syntactic structure of language) can be quite rewarding. Cf. B. Mandelbrot, "Structure formelle des textes et communication: deux études," *Word* 10.1–27 (1954); H. A. Simon, "On a class of skew distribution functions," *Biometrika* 42.425–400 (1955).
>
> One might seek to develop a more elaborate relation between statistical and syntactic structure than the simple order of approximation model we have rejected. I would certainly not care to argue that any such relation is unthinkable, but I know of no suggestion to this effect that does not have obvious flaws. (p. 17, fn. 4)

It comes as no surprise that as original a thinker as Chomsky, who was shortly to be working with Miller, the father of psycholinguistics, was pointing in

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

these quotations to the work of Mandelbrot, one of the forerunners of Chaos/Complexity theory, and Simon, the progenitor of the information processing model of human cognition, artificial intelligence (AI), and bounded rationality. There may not have been a suitable model of the relation between statistical and syntactic structure in 1957, but the very enterprises Chomsky identified here have since developed this understanding. Psycholinguistics provided the evidence of the effects of frequency that have been reviewed in this paper. Cognitive psychology developed ways of rationalizing intelligence in terms of models of optimal inference in the presence of uncertainty. AI, and more recently connectionism, provided the computational framework for testing input-based theories as simulations. Finally, complexity theory served as the foundation for recent characterizations of language as an emergent system (e.g., MacWhinney, 1997b, 1999, for L1; Ellis, 1998, and Larsen-Freeman, 1997, for L2).

Computational implementations of generative grammars that are large enough to cover a nontrivial subset of natural language assign to many sentences an extremely large number of alternative syntactic analyses, yet fluent humans perceive only one or two of these when faced with the same input. Such models may be judged successful if the defining criterion is that it describes the space of possible analyses that sentences may get. However, the combinatorial explosion of syntactic analyses and corresponding semantic interpretations are very problematic if the criterion is rather to predict which analyses human comprehenders actually assign to natural language utterances (Bod, 1998; Church & Patil, 1982; Martin, Church, & Patil, 1981). The Natural Language Processing community has moved to the use of stochastic grammars to overcome these problems. Examples include stochastic context-free grammar (Sampson, 1986), stochastic unification-based grammar (Briscoe, 1994), stochastic head-driven phrase-structure grammar (Brew, 1995), stochastic lexical-functional grammar (Kaplan, 1999), and data-oriented parsing (DOP) (Bod). Bod described experienced-based DOP models of language, which learn how to provide appropriate linguistic representations from an unlimited set of utterances by generalizing from examples of representations of previously occurring utterances. These probabilistic models operate by decomposing the given representations into fragments and recomposing those pieces to analyze new utterances. Bod showed that any systematic restriction of the fragments seems to jeopardize the statistical dependencies that are needed for predicting the appropriate structure of a sentence. This implies that the productive units of natural language cannot be defined in terms of a minimal set of rules, constraints, or principles, but rather they need to be defined in terms of a large redundant set of previously experienced structures with virtually no restriction on size or complexity. The behavior of the society of syntax is determined by the interactions and associations of all of its members; if communities are excised, or if new individuals join, then the whole ecology changes.

For these reasons it is becoming increasingly the case that models of syntax, parsing, and natural language processing contain mechanisms for gathering information about the frequencies of language representation along with

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

probabilistic algorithms that guide the access and disambiguation of linguistic knowledge (Jurafsky, 1996; Jurafsky & Martin, 2000). Connectionist models of sentence processing are still in their infancy. Nevertheless, Allen and Seidenberg (1999) reported a simulation of the emergence of grammaticality. The model was sensitive to sequential word probabilities and to sequential semantic probabilities (e.g., the representation of the word *mother* included the features female-parent, parent, female, relative, human, living-thing, organism, animate, entity). After being exposed to a training corpus of 100 utterances, the model gave high grammaticality scores to utterances of high word and high semantic transition probabilities such as *which girl did you invite*. It also gave values that were close to these levels for sentences in which the transitional probabilities between words was low but the transitional probabilities between semantic types was high. Allen and Seidenberg claimed that this type of sentence corresponds to *colorless green ideas sleep furiously* in which the semantic subsequences [property property entity act manner] are consistent with semantic sequences that appeared in the training set, but the bigram frequencies of the words (e.g., colorless-green, green-ideas) were low or zero in the training corpus. The model gave very low grammaticality scores for sentences such as *on invited cake street time the* that had low word transition probability and low semantic transition probability—that is, sentences that Allen and Seidenberg claim to parallel examples of the type *furiously sleep ideas green colorless*.

Other recurrent connectionist models of language include those of Elman (1990) and Redington and Chater (1998), which demonstrate how learning the grammatical categories and requirements of words and word groups reduces to the analysis of the sequence in which words work in chunks. These part-of-speech categories emerge from the language input without any semantics or real-world grounding, and their category structure is hierarchical, soft, and implicit. Landauer and Dumais (1997) presented a latent semantic analysis model that simulated language learners' acquisition of lexical semantics from text. The same types of network architecture that discover sublexical regularities of language also discover important grammatical and semantic information; sequential analysis provides useful language representation at all levels of grain. The networks move from processing frequency relations between mere surface regularities to representing something more abstract, without this being built in as a prespecified syntactic or other linguistic constraint. Relatively general architectural constraints give rise to language-specific representational constraints as a product of processing the input strings. These linguistically relevant representations are an emergent property. They emerge from the analyses of co-occurrence patterns.

TYPE VERSUS TOKEN FREQUENCY

How exactly does the frequency of patterns in the input affect acquisition? We have already discussed how representational strength is determined by the

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

power law of practice rather than raw frequency, as well as how the effects of frequency in input are modulated by the need to simultaneously satisfy the constraints of all other constructions that are represented in the learners' system. (See Ellis & Schmidt, 1998, pp. 329–331 for a detailed explanation of these two factors in interaction in the example case of morphosyntax.) This section considers the separable effects of token frequency and type frequency using examples at the level of lexis and constructions.

Token frequency is how often particular words or specific phrases appear in the input. Type frequency, on the other hand, is how many different lexical items can be applied to a certain pattern, paradigm, or construction. Type frequency refers to the number of distinct lexical items that can be substituted in a given slot in a construction, whether it is a word-level construction for inflection or a syntactic construction specifying the relation among words. For example, the "regular" English past tense *-ed* has a very high type frequency because it applies to thousands of different types of verbs, whereas the vowel change exemplified in *swam* and *rang* has much lower type frequency. Bybee (1995), Bybee and Thompson (2000), and the researchers gathered in Bybee and Hopper (2001) showed how the productivity of phonological, morphological, and syntactic patterns is a function of type rather than token frequency. In contrast, high token frequency promotes the entrenchment or conservation of irregular forms and idioms; the irregular forms only survive because they are high frequency. These findings support language's place at the center of cognitive research into human categorization, reviewed above, which also emphasizes the importance of type frequency in classification.

Type frequency determines productivity because: (a) the more lexical items that are heard in a certain position in a construction, the less likely it is that the construction is associated with a particular lexical item and the more likely it is that a general category is formed over the items that occur in that position; (b) the more items the category must cover, the more general are its criterial features and the more likely it is to extend to new items; and (c) high type frequency ensures that a construction is used frequently, thus strengthening its representational schema and making it more accessible for further use with new items (Bybee & Thompson, 2000).

IMPLICATIONS FOR THEORIES OF L2 ACQUISITION

These psycholinguistic demonstrations that frequency sensitivity pervades all aspects of language processing have profound implications for theories of language acquisition—namely, that language learning is exemplar based. The evidence reviewed here suggests that the knowledge underlying fluent use of language is not grammar in the sense of abstract rules or structure but a huge collection of memories of previously experienced utterances. These exemplars are linked, with like kinds being related in such a way that they resonate as abstract linguistic categories, schema, and prototypes. Linguistic regulari-

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

ties emerge as central tendencies in the conspiracy of the database of memories of utterances.

For language learners to be accurate and fluent in their generalizations they need to have processed sufficient exemplars that their accidental and finite experience is truly representative of the total population of language of the speech community in terms of its overall content, the relative frequencies of that content, and the mappings of form to functional interpretation. The enormity of the lexical pool, the range of frequencies from 60,000 per million down to 1 per million and below, and the wide range of different linguistic constructions, when considered from the point of view of sampling theory, makes it clear that the necessary representative experience for fluency must be vast indeed.

Traditional descriptive and pedagogical grammars relate well to these theories of acquisition, both in their induction and in their descriptive grain, by focusing on constructions as recurrent patterns of linguistic elements that serve some well-defined linguistic function. These may be at the sentence level (e.g., imperative, ditransitive, or yes-no question) or below (NP, PP, etc.). Whereas generative grammar denied constructions, viewing them as epiphenomena resulting from the interaction of higher level principles and parameters and lower level lexicon, cognitive linguistics (Barlow & Kemmer, 2000; Croft & Cruse, 1999; Langacker, 1987; Ungerer & Schmid, 1996), and Construction Grammar in particular (Goldberg, 1995), has brought constructions back to the fore, suspecting instead that it is the higher-level systematicities that emerge from the interactions of constructions large and small.

A construction is a conventional linguistic unit—that is, part of the linguistic system, accepted as a convention in the speech community, and entrenched as grammatical knowledge in the speaker's mind. Constructions may be complex, as in [Det Noun], or simple, as in [Noun] (traditionally viewed as "lexicon"); they may represent complex structure above the word level, as in [Adj Noun], or below the word level, as in [NounStem-PL] (traditionally viewed as "morphology"); and they may be schematic, as in [Det Noun], or specific, as in [the United Kingdom] (traditionally viewed as "lexicon"). Hence, "morphology," "syntax," and "lexicon" are uniformly represented in a construction grammar, unlike both traditional grammar and generative grammar, and chunks of language much larger than the analytic units of morphemes or even words are the usual units of storage and processing. Constructions are symbolic. That is, in addition to specifying the properties of an utterance's defining morphological, syntactic, and lexical form, a construction also specifies the semantic, pragmatic, and discourse functions that are associated with it. Constructions form a structured inventory of a speaker's knowledge of the conventions of his or her language (Langacker, 1987, pp. 63–66), in which schematic constructions can be abstracted over the less schematic ones that are inferred inductively by the speaker in acquisition. A construction may provide a partial specification of the structure of an utterance. Hence, an utterance's structure is specified by a number of distinct constructions. Constructions are indepen-

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

dently represented units in a speaker's mind. Any construction with unique, idiosyncratic formal or functional properties must be represented independently to capture a speaker's knowledge of his or her language. However, absence of any unique property of a construction does not entail that it is not represented independently and simply derived from other, more general or schematic constructions. Frequency of occurrence may lead to independent representation of even so-called regular constructional patterns. In this usage-based perspective, the acquisition of grammar is the piecemeal learning of many thousands of constructions and the frequency-biased abstraction of regularities within them (Croft & Cruse, 1999).

Many constructions are based on particular lexical items, ranging from simple (*Howzat!* in cricket) to complex (*Beauty is in the eye of the beholder*). But other constructions are more abstract. Goldberg (1995) focused on complex argument-structure constructions such as the ditransitive (*Pat faxed Bill the letter*), the caused motion (*Pat pushed the napkin off the table*), and the conative (*Sam kicked at Bill*). Goldberg held that these abstract and complex constructions themselves carry meaning, independently of the particular words in the sentence. For example, even though the verb *kick* does not typically imply transfer of possession, it works in the ditransitive *Pat kicked Bill the football*, and even though one is hard-pressed to interpret anything but an intransitive *sneeze*, the caused motion *Pat sneezed the napkin off the table* is equally good. These abstract argument-structure constructions thus create an important top-down component to the process of linguistic communication. They would predict, for example, a reasonable analogic understanding of the sentence *Eloquence is in the ear of the hearkener*. Such influences are powerful mechanisms for the creativity of language, possibly even as manifest in derivational phenomena such as denominal verbs (e.g., *They tabled the motion*) and deverbal nouns (e.g., *Drinking killed him*) (Tomasello, 1998b). As reviewed earlier for human classification more generally, constructions show prototype effects. For example, ditransitive constructions have a central sense of "agent-successfully-causes-recipient-to-receive-patient" (*Bill gave* [*handed*, *passed*, *threw*, *took*] *her a book*) and various more peripheral meanings such as future-transfer (*Bill bequeathed* [*allocated*, *granted*, *reserved*] *her a book*) and enabling-transfer (*Bill allowed* or *permitted her one book*).

Currently, we must look to child language research for the development of construction-based theories of acquisition (Tomasello, 1998a, 2000). These accounts emphasize the piecemeal learning of concrete exemplars. Longitudinal child acquisition data suggest that, to begin with, each word is treated as a semantic isolate in the sense that the ability to combine it with other words is not accompanied by a parallel ability with semantically related words. Bowerman (1976) noted that her daughter acquired the *more* + X construction long before other semantically similar relational words like *again* and *all-gone* came to be used in the similar pivot position in two-word utterances. Pine and Lieven (1993, 1997), Lieven, Pine, and Dresner Barnes (1992), and Pine, Lieven,

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

and Rowland (1998) have since demonstrated widespread lexical specificity in L1 grammar development. Children's language between the ages of 2 and 3 years old is much more "low-scope" than theories of generative grammar have argued. A high proportion of children's early multiword speech is produced from a developing set of slot-and-frame patterns. These patterns are often based around chunks of one or two words or phrases and have slots into which the child can place a variety of words, for instance subgroups of nouns or verbs (e.g., *I can't* + verb; *where's* + noun + *gone?*). Children are very productive with these patterns and both the number of patterns and their structure develop over time. They are, however, lexically specific: Pine, Lieven, and colleagues analyzed recordings of 2- to 3-year-old children speaking with their mothers, and they measured the overlap between the words used in different slots in different utterances. For example, if a child had two patterns, *I can't* + X and *I don't* + X, they determined whether the verbs used in the X slots came from the same group and whether any other *can*- or *do*- auxiliaries could be used. There was typically very little or no overlap, an observation that supports the conclusions that (a) the patterns are not related through an underlying grammar (i.e., the child does not "know" that *can't* and *don't* are both auxiliaries or that the words that appear in the patterns all belong to a category of Verb), (b) there is no evidence for abstract grammatical patterns in the 2- to 3-year-old child's speech, and (c) in contrast, the children are picking up frequent patterns from what they hear around them and only slowly making more abstract generalizations as the database of related utterances grows.

Tomasello (1992) proposed the Verb Island hypothesis in which it is the early verbs and relational terms that are the individual islands of organization in young children's otherwise unorganized grammatical system. In the early stages, the child learns about arguments and syntactic markings on a verb-by-verb basis, and ordering patterns and morphological markers learned for one verb do not immediately generalize to other verbs. Positional analysis of each verb island requires long-term representations of that verb's collocations, and thus this account of grammar acquisition implies vast amounts of long-term knowledge of word sequences. Only later are syntagmatic categories formed by abstracting regularities from this large dataset in conjunction with morphological marker cues (at least in case-marking languages). Goldberg (1995) argued that certain patterns are more likely to be made more salient in the input because they relate to certain fundamental perceptual primitives, and thus that the child's construction of grammar involves both the distributional analysis of the language stream and the analysis of contingent perceptual activity:

> Constructions which correspond to basic sentence types encode as their central senses event types that are basic to human experience . . . that of someone causing something, something moving, something being in a state, someone possessing something, something causing a change of state or location, something undergoing a change of state or location, and something having an effect on someone. (p. 39)

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

Goldberg and Sethuraman (1999) showed how individual, semantically prototypic verbs are "pathbreakers" in that they seed the growth of verb-centered argument-structure patterns. Generalizations of the verb-centered instances emerge gradually as the verb-centered categories themselves are analyzed into more abstract argument-structure constructions. The verb is a better predictor of sentence meaning than any other word in the sentence. Nevertheless, children ultimately generalize to the level of constructions because these are much better predictors of overall meaning. Although verbs thus predominate in seeding low-scope patterns and eventually more abstract generalizations, Pine et al. (1998) have shown that such islands are not exclusive to verbs and that the theory should be extended to include limited patterns based on other lexical types such as bound morphemes, auxiliary verbs, and case-marking pronouns. Nevertheless, the focus on verbs echoes their central role in determining the syntactic structure of a sentence and, as reviewed previously, their providing the strongest constraints on the resolution of syntactic ambiguities.

In sum, theories of the acquisition of first-language grammatical constructions maintain that there is a developmental sequence from formula, through low-scope pattern, to construction. Second and foreign language acquisition is different from L1 acquisition in numerous respects. First, in conceptual development: In child language acquisition, knowledge of the world and knowledge of language are developing simultaneously, whereas adult SLA builds on preexisting conceptual knowledge. Moreover, adult learners have sophisticated formal operational means of thinking and can treat language as an object of explicit learning—that is, of conscious problem-solving and deduction—to a much greater degree than can children (Ellis, 1994a). Second, in language input: The typical L1 pattern of acquisition results from naturalistic exposure in situations where caregivers naturally scaffold development (Tomasello & Brooks, 1999), whereas classroom environments for second or foreign language teaching can distort the patterns of exposure, function, medium, and social interaction (Ellis & Laporte, 1997). Third, in transfer from L1: Adult SLA builds on preexisting L1 knowledge (Kellerman, 1995; MacWhinney, 1992; Odlin, 1989). Nevertheless, the L1 acquisition sequence—from formulas, through low-scope patterns, to constructions—could serve well as a reasonable default in guiding the investigation of the ways in which exemplars and their type and token frequencies determine the second language acquisition of structure.

SOCIOLINGUISTICS

Sociolinguistics is concerned with variation: regional and social variations of language, diachronic change, and intra-individual variation tuned to differing communicative situations. Differences between individuals result from their differing histories of input: "No two speakers have the same language, because no two speakers have the same experience of language" (Hudson, 1996, p. 11); "the more speakers talk to each other, the more they will talk alike, and so

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

linguistic variation will pattern along lines of social contact and interaction" (Kemmer & Israel, 1994, p. 167). If linguistic regularities emerge as central tendencies in a conspiracy of the database of memories for utterances, it is easy to see how input-frequency based accounts might lead to this variation and change. Psychological processing models of language are well-suited collaborators alongside theories of social interaction and accommodation, as well as Vygotskian and social psychology, in the sociolinguistic understanding of mechanisms of variation (Tarone, 1988, 1997).

The Systematicity of Variability

Sociolinguistic variation provides the strongest empirical evidence against the self-containedness of grammatical knowledge of the individual (Croft, 1995). Variation occurs in the speech of single individuals and hence must be a part of a speaker's knowledge of the language. Fluent speakers have many variants as part of their grammatical competence, and they employ them differently depending on their social position in the speech community and the circumstances of the conversational interaction. To the degree that the systematicity of an individual's variable grammar is tied to the individual's interactions in social settings, the grammar is not self-contained. Labov (1972) importantly demonstrated the systematicity of variability and proposed that variable rules may stipulate the linguistic and social contexts in which a rule is more or less likely to apply (e.g., omission of the contracted verb, as in *We're here*, in Black English occurs more often before consonants than before vowels and more often in informal situations than formal). Such explanation in terms of rules may be too top-down for present tastes and "rulelike" would be a preferred conceptualization (Ellis, 1996a; Ellis & Schmidt, 1998), but the statistical sensitivity that is displayed is another powerful example of learner's frequency matching:

> [Labov's work] . . . implies that it is wrong to think of an individual as mastering a single idiolect, and understanding others' speech only insofar as it resembles his own. Rather, it seems that each speaker learns a structured range of alternative speech patterns, together with the correlations between variation in his social environment and variation in that dialectal continuum. There is nothing particularly surprising in the finding that speakers are familiar with a variety of speech-styles, of course, but many of us had supposed that such knowledge was patchy and largely inaccurate—as speakers' conscious beliefs about such facts certainly are. What is staggering about Labov's work is the subtlety, consistency, and mathematical regularity it reveals in speakers' use of statistical linguistic variables and hearers' reactions to them. (Sampson, 1980, pp. 128–129)

As with other aspects of language, there are attempts to formalize this knowledge in computational models. For example, one of the controversial features of Word Grammar (WG) (Hudson, 1990) is that it allows the grammar to include sociolinguistic information about speakers and addressees because

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

the relationships "speaker" and "addressee" are available (and are used elsewhere, notably in treating deixis). Thus, it is possible to include both *pavement* and *sidewalk* in the same grammar but to show that they are used by different kinds of people—that is (roughly speaking), that they belong to different "dialects" in that the speaker of *pavement* is British and the speaker of *sidewalk* is American. Hudson also discusses ways to give similar kinds of information about registers (e.g., formality and other matters of style). The WG model is based on a semantic network that incorporates spreading activation across paths of variable weight. At present, the strengths of these weights are determined and set by the modeler. Emergentists would want better to incorporate these ideas into connectionist types of model in which such relations are learned rather than hand-tuned.

Language Change

Individual learner grammars incorporate variation. This variation changes through use in ways that can lead to the propagation of a change in the speech community that will be established as such in the mental representations of speakers' (variable) grammars, thus resulting in diachronic language change. Boyland (1996) pointed out that the changes in form that occur in the grammaticization process closely resemble changes that occur as nonlinguistic skills are practiced and become automatized. Building on such ideas, Bybee (1995, 2000) and Bybee and Hopper (2001) have developed a model of grammaticization as the process of automatization of frequently occurring sequences of linguistic elements. With repetition, sequences of units that were previously independent come to be processed as a single unit or chunk (Ellis, 1996b). This repackaging has two consequences: The identity of the component units is gradually lost, and the whole chunk begins to reduce in form. As described previously, these basic principles of automatization apply to all kinds of motor activities—for example, playing a musical instrument, cooking, or playing an Olympic sport. They also apply to grammaticization. A phrase such as *(I'm) going to* + Verb, which has been frequently used over the last couple of centuries, has been repackaged as a single processing unit. The identity of the component parts is lost (children are often surprised to see that *gonna* is actually spelled *going to*), and the form is substantially reduced.

Thus in Bybee's model, frequency is the driving force of language change in five ways. First, frequency of use leads to weakening of semantic force by habituation. Second, phonological changes of reduction and fusion of grammaticizing constructions are conditioned by their high frequency. Third, increased frequency conditions a greater autonomy for a construction, which means that the individual components of the construction (such as *go*, *to*, or *-ing* in the example of *be going to*) weaken or lose their association with other instances of the same item (as the phrase reduces to *gonna*). Fourth, the loss of semantic transparency accompanying the rift between the components of the grammaticizing construction and their lexical congeners allows the use of

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

the phrase in new contexts with new pragmatic associations, leading to semantic change. Fifth and finally, autonomy of a frequent phrase makes it more entrenched in the language and often conditions the preservation of otherwise obsolete morphosyntactic characteristics (Bybee, 2000). Frequency is the determinant of lexical and construction strength that are key components of Bybee's schema-network models. It is similarly the driving force of other usage-based accounts of language change (Boyland, 1996; Croft, 2000; Kemmer & Israel, 1994).

In sum, frequency is a mechanism underlying much of sociolinguistic variation, and the measurement of frequency of use is consequently an essential descriptive foundation (Biber, 1996; Biber & Finegan, 1994; Biber et al., 1998). To understand variation we must do what learners themselves do in order to vary appropriately: In the words of a recent evolution in American English, we must "Go figure."

IMPLICATIONS FOR THEORIES OF IMPLICIT AND EXPLICIT LEARNING

This review has shown that much of language learning is the gradual strengthening of associations between co-occurring elements of the language and that fluent language performance is the exploitation of this probabilistic knowledge. Because the conscious experiences of language learning do not revolve around counting, to the extent that language processing is based on frequency and probabilistic knowledge, language learning is implicit learning.

Two provisos must be added to put language learning in the proper balance. The first concerns initial registration of representations, the second explicit instruction. First, the frequency effects detailed here describe the tuning of the language system through use. The initial registration of a language representation may well require attention and conscious identification. Schmidt (1990) has properly emphasized the importance of apperception for language acquisition in his "noticing" hypothesis. Schmidt proposed that the subjective experience of noticing is the necessary and sufficient condition for the conversion of input to intake in SLA. An extension of the noticing hypothesis is that what must be attended to and noticed is not just the input in a global sense but whatever features of the input are relevant for the target system (Schmidt, 1993). That is, to acquire phonology one must attend to phonology, and to acquire pragmatics, one must notice both linguistic forms and the relevant contextual features, and so forth:

> Noticing is used here as a technical term to refer only to registration of the occurrence of a stimulus event in conscious awareness and subsequent storage in long term memory, not the detection of form/meaning relationships or inductive formation of hypotheses or other processes that may lead to the organization of stored knowledge into a linguistic system. (Schmidt, 1994, p. 179)

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

The strong form of the noticing hypothesis is that attention must be paid to some aspect of the stimulus environment, and that aspect must be noticed before a mental representation of that stimulus can first be formed. However, the strong form of an implicit tallying hypothesis to accompany it is that once a stimulus representation is firmly in existence, that stimulus, be it for example the letter *b* or a rise in pitch at the end of questions, need never be noticed again; yet as long as it is attended to for use in the processing of future input for meaning, its strength will be incremented and its associations will be tallied and implicitly cataloged. This latter is the same assumption as that first proposed by Logan (1988) in his instance theory of automatization: "Encoding into memory is an obligatory, unavoidable consequence of attention" (p. 493). What remains in dispute falls between these two hypotheses and primarily centers on the size of grain of representation: Are all types of association equally learnable by implicit means, or are there some relations that require explicit learning? The material reviewed here, as well as the research marshaled in psycholinguistics texts (e.g., Altman, 1997; Gernsbacher, 1994; Harley, 1995), shows us to be very well tuned to our language input at all levels. Once low-level mental representations are formed, they can be grouped together hierarchically—that is, "chunked" by unconscious processes of association to form larger units that are henceforth used in pattern recognition. As Ellis (1996b) argued:

> Although learners need not be aware of the processes of such pattern extraction, they will later be aware of the product of these processes since the next time they experience that pattern it is the patterned chunk that they will be aware of in working memory, not the individual components (for illustration, while young children learn about clocks they devote considerable attention to the position of hands on an analog face in relation to the pattern of numerals; when experienced adults consult their watch they are aware of the time, and have no immediate access to lower-level perceptual information concerning the design of the hands or numerals [Morton, 1967]). (p. 109)

Nevertheless, there do seem to be more complex associations that require more conscious explicit learning and hypothesis testing to acquire. The experiments of Ellis, Lee, and Reber (1999) provided evidence suggesting that this is the case for some long-distance discontinuous dependencies in language acquisition. Similarly, Cohen, Ivry, and Keele (1990) and Curran and Keele (1993) showed that, whereas unique sequences can be acquired implicitly in artificial grammar learning experiments, ambiguous sequences require more attentional forms of learning.

The second proviso is that these findings do not deny a role for explicit instruction, the efficacy of which is now well established. Language acquisition can be speeded by explicit instruction. Reviews of the experimental and quasiexperimental investigations into the effectiveness of L2 instruction (e.g., Ellis & Laporte, 1997; Hulstijn & DeKeyser, 1997; Lightbown, Spada, & White,

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

1993; Long, 1983; Spada, 1997), particularly the Norris and Ortega (2000) comprehensive meta-analysis of the last 20 years' empirical work, demonstrated that focused L2 instruction results in large target-oriented gains, that explicit types of instruction are more effective than implicit types, and that the effectiveness of L2 instruction is durable. This is not to say that just providing learners with pedagogical rules will make them into fluent language users. Far from it (Krashen, 1985; Krashen & Terrell, 1983), because then the learner neither gets the exemplars nor the tuning. Pedagogical rules are only properly effective when demonstrated in operation with a number of illustrative exemplars of their application (Ellis, 1993).

The real stuff of language acquisition is the slow acquisition of form-function mappings and the regularities therein. This skill, like others, takes tens of thousands of hours of practice, practice that cannot be substituted for by provision of a few declarative rules. Communicative approaches give input, time-on-task, and opportunity for relating form and function. All of this is necessary for developing the associations required for language learning. Naturalistic environments provide motivation and plenty of opportunity for output practice as well. These are situations that guarantee sufficient quantity of language and proper involvement in the communicative functions of language. However, without any focus on form or consciousness raising (Sharwood Smith, 1981), formal accuracy is an unlikely result; relations that are not salient or essential for understanding the meaning of an utterance are otherwise only picked up very slowly, if at all (Schmidt, 1990; Terrell, 1991). Focus on forms alone can teach some declarative rules of grammar but at its worst can be accompanied by too little time on the task of language use itself. Focus-on-form instruction, which is rich in communicative opportunities and which at the same time makes salient the association between communicative function and structure, can facilitate language acquisition (Doughty & Williams, 1998; Long, 1991). As MacWhinney (1997a) put it:

> Students who receive explicit instruction, as well as implicit exposure to forms, would seem to have the best of both worlds. They can use explicit instruction to allocate attention to specific types of input . . . , narrow their hypothesis space . . . , tune the weights in their neural networks . . . , or consolidate their memory traces. From the viewpoint of psycholinguistic theory, providing learners with explicit instruction along with standard implicit exposure would seem to be a no-lose proposition. (p. 278)

The communicative functions of language motivate the learner to the task. Noticing lays out the problem. Consciousness-raising can speed its solution. Figuring provides the final tally of native levels of fluency and idiomaticity.

FREQUENCY IN APPLIED LINGUISTICS: LADO AND LATER

The acknowledgment of frequency as a causal variable in language learning, though not prominent in the last 40 years of applied linguistics or second lan-

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

guage studies, is certainly not new. It was a key concept in the theories and practices of American structural linguists such as Harris (1955, 1968). Harris's *Mathematical structures of language* described a recurrent dependence process for identifying sentence boundaries at the points where the sequential likelihood between word neighbors shows a discontinuity. He also described how morphemes could be segmented out of a string of phonemes using the same process. This work set firm foundations for more recent recurrent models of language sequence analysis (Brent, 1999; Christiansen et al., 1998; Elman, 1990). There is no more thorough investigation of the effects of different frequencies of behavior and contingencies of reinforcement than that established in behaviorist psychology by Skinner (1957, 1969). Lado's theory of second language learning, built on the behaviorist principles of learning (including the fundamental principle of contiguity, the law of exercise, the law of intensity, the law of assimilation, and the law of effect), was that grammatical structure is a system of habits (Lado, 1957, p. 57; 1964, pp. 37–45). For Lado, acquisition is the learning of patterns of expression, content, and their association, a concept closely akin to the constructions of the previous section. Language learning occurs through experience, and because all experiences leave a trace in the memory store, all previous experiences are a factor, either facilitating or inhibiting the learning of a new language, hence his later development of contrastive analysis. His theory emphasized the shift from attentional to unconscious processing through automatization:

> Knowing a language is defined as the power to use its complex mechanism through bundles of habits while only the thread of the argument and some matters of selection and agreement are under attention. This power is achieved gradually by strengthening the facilities for partial use through repeated experiences so that attention is freed from the mechanics of use. In other words, before learning, attention must control every minor element of speaking or listening expression, content, or association. Since attention is limited in the number of items that can be held simultaneously under it, learning occurs when attention can be shifted to larger units and patterns and eventually to the total utterance by increased facility with the smaller units and patterns. (1964, p. 41)

I have only recently discovered these interests of his in automaticity and attention. However, when I consider them alongside his other investigations of short-term memory constraints as determinants of individual differences in language learning (Lado, 1965), I admit to some weary feelings of familiarity in observing the inadvertent similarity of our research paths. As Lado's ideas on language learning appear in their final form as lexicosemantic theory, they are surprisingly reminiscent of cognitive linguistic approaches: "This view is lexical and semantic rather than syntactic in nature. It claims that humans acquire and learn words, names, titles, expressions, sayings, and formulas as undifferentiated lexical items (lexemes) first" (Lado, 1990, p. 96). In this view, as the number of lexemes increases, so systems develop for storing and retrieving them as needed and a grammatical system emerges from these:

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

> The combinations are amorphous at first and evolve into patterns and eventually rules that permit the creation of new words and sentences. Grammatical meaning attaches to word order, inflections, derivations, etc. from meaningful samples that are experienced by learners in interaction with speakers, and then the patterns and rules are applied to create and understand new phrases and sentences. (1990, p. 99)

So why is this work of so little current influence? There are two special factors besides the usual explanation that most papers have a limited shelf-life; applied linguistics is no less influenced by fashion and topicality than any other discipline, and publication citation half-life is typically humblingly brief. The first is that progress in behaviorist theories of language learning floundered following Chomsky's (1959) masterly and highly influential critique of Skinner's (1957) *Verbal Behavior*. The second is that behavioral approaches fell from favor in applied linguistics in eventual reaction against Lado's implementation of his theory in the Audio-Lingual Method (ALM) (Lado, 1948, 1964). However laudable his belief in an applied psycholinguistics as a scientific approach to language learning and teaching, Lado's operationalization championed behaviorist principles of learning at the expense of language and the learner. Despite his premise of language learning as the learning of patterns of expression, content, and their association, the ALM involved "mimicry-memorization" in pattern drills in which the role of understanding was minimized as much as possible. The major emphasis was on the mechanical production of the utterance as a language form and on the development of automatic responses of the desired nature, that is, good habits. At its worst, the method involved mindless repetition and meaningless drills, with unmotivated learners interacting with teaching machines using far from authentic materials. Further, unlike grammar-translation methods, there was no capitalizing on learner's metalinguistic abilities or explicit instruction. The ALM was a blinkered and limited implementation of theories of frequency and habit in language learning. Language learning involves much more than mindless repetition. *Mindful* repetition in an engaging communicative context by motivated learners is somewhat closer to the mark.

So it was that in the 1970s the ALM became unfashionable, and few applied linguists considered frequency very much thereafter. Nonetheless, Larsen-Freeman (1976) determined it was the only thing that could adequately explain the common acquisition order of English morphemes to which ESL learners, despite their different ages and language backgrounds, all seem to adhere; Goldschneider and DeKeyser's (2001) recent meta-analysis of the morpheme acquisition studies confirms that this so-called natural order is a function of the frequency of occurrence of these morphemes in adult native-speaker speech. Frequency remained as a tacit factor, of course, in all input-based theories of SLA (see Gass, 1997, for review). Only in the last decade, however, has it been explicitly reintroduced to studies of SLA after its theoretical development in psycholinguistics, connectionism, and child language acquisition. A notable example of this is the Competition Model as a sophisticated theory of

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

frequency-based interactions and the associated tests of its predictions in both L1 acquisition and SLA (Bates & MacWhinney, 1987; Kempe & MacWhinney, 1998; MacWhinney, 1987, 1997c, 1999).

CONCLUSIONS AND CONNECTIONS

This review has demonstrated a range of influences of frequency in a wide variety of theoretical approaches to language across the whole gamut of language-relevant representations. Frequency is a necessary component of theories of language acquisition and processing. In some guises it is a very rudimentary causal variable. Learners analyze the language input that they are exposed to; practice makes perfect. In other guises it is incredibly complex. The multiplicity of interacting elements in any system that nontrivially represents language makes the prediction of the patterns that will eventually emerge as difficult as forecasting the weather, the evolution of an ecological system, or the outcome of any other complex system (Ellis, 1998; Larsen-Freeman, 1997; MacWhinney, 1997b, 1999). This is why complexity requires investigation using computational simulations.

Of course, frequency is not a sufficient explanation; otherwise we would never get beyond the definite article in our speech. There are many other determinants of acquisition. Semantic basicness, salience, communicative intent, and relevance are major determining factors in the acquisition process (Slobin, 1997). Whereas frequency explains some of the variance of morpheme acquisition order, it does so in interaction with other determinants such as perceptual salience, semantic complexity, morphophonological regularity, and syntactic category (Goldschneider & DeKeyser, 2001). A hardliner might want to argue that all of these aspects are emergent from patterns of experience as well, but such generality is too abstract to buy very much explanation at all, other than to say that we are adaptive organisms. Our language acquisition is limited by cognitive constraints, and our limited working memories (Ellis, 1996b) determine bounds to our rationality:

> The capacity of the human mind for formulating and solving complex problems is very small compared with the size of the problems whose solution is required for objectively rational behaviour in the real world—or even for a reasonable approximation to such objective rationality. (Simon, 1957, p. 198)

As cognitive linguistics emphasizes, our language builds on our experience and our embodiment that represents the world in a very particular way. The different degrees of salience or prominence of elements involved in situations that we wish to describe affect the selection of subject, object, adverbials, and other clause arrangement. Figure and ground segregation and perspective taking are mirrored in language and have systematic relations with syntactic structure. All of the systems of working memory, all perceptual representa-

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

tional systems, and our attentional resources and supervisory systems are involved in collating the regularities of our experience and of cross-modal associations relating them to language use. Our common physical, social, and cognitive inheritance affects us all and commonly moderates the processing of our language experience; there are universals of language acquisition that derive from them. The Aspect Hypothesis—whereby language learners are initially influenced by the inherent semantic aspect of verbs or predicates in the acquisition of tense and aspect markers associated with or affixed to these verbs and may be less sensitive to tense or grammatical aspect (Andersen, 1991; Bardovi-Harlig, 2000; Shirai & Andersen, 1995)—is a good example of such influences in both first and second language acquisition.

Nevertheless, the ubiquitous influence of frequency gives researchers of language acquisition from different disciplines some common currency. Corpus linguists count and analyze the patterns in the input, the different registers for sociolinguists, and the different stages of proficiency and interlanguage for child and L2 acquisition researchers. Psycholinguists investigate the influence of these patterns in language processing. Psychologists, linguists, and connectionists alike are concerned with the role of input frequency in determining acquisition, the differential effects on implicit and explicit learning, the ways that attention can modulate frequency effects, and the representations and regularities that emerge from these processes. There are important collaborations to be made between child language researchers, second language researchers, and cognitive linguists. Psychophysicists investigate the relations between input frequency, sensation, and memory consolidation, and neuroscientists research how synaptic plasticity substantiates these changes. Educators try to ensure that learners meet a representative sample of authentic language experience so that the learner representations that emerge approximate those of native speakers. . . . Enough drawing near and binding tight.

The role of frequency has largely been ignored in theoretical and applied linguistics for the last 40 years or so. As this review has shown, there is now ample justification for its reinstatement as an all-pervasive causal factor. In the final analysis of successful language acquisition and language processing, it is the language learner who counts.

REFERENCES

Allen, J., & Seidenberg, M. S. (1999). The emergence of grammaticality in connectionist networks. In B. MacWhinney (Ed.), *The emergence of language* (pp. 115–152). Mahwah, NJ: Erlbaum.

Altman, G. T. M. (1997). *The ascent of Babel*. Oxford: Oxford University Press.

Andersen, R. W. (1991). Developmental sequences: The emergence of aspect marking in second language acquisition. In T. Huebner & C. A. Ferguson (Eds.), *Crosscurrents in second language acquisition and linguistic theories* (pp. 305–324). Amsterdam: Benjamins.

Anderson, J. R. (1980). *Cognitive psychology and its implications*. San Francisco, CA: W. H. Freeman.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*, 369–406.

Attneave, F. (1953). Psychological probability as a function of experienced frequency. *Journal of Experimental Psychology*, *46*, 81–86.

Balota, D. A., & Chumbley, J. L. (1984). Are lexical decisions a good measure of lexical access? The

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 340–357.

Bardovi-Harlig, K. (2000). *Tense and aspect in second language acquisition: Form, meaning, and use*. Oxford: Blackwell.

Barlow, M., & Kemmer, S. (Eds.). (2000). *Usage based models of language*. Stanford, CA: CSLI Publications.

Barry, C., & Seymour, P. H. K. (1988). Lexical priming and sound-to-spelling contingency effects in nonword spelling. *Quarterly Journal of Experimental Psychology*, *40*(A), 5–40.

Barsalou, L. W., Huttenlocher, J., & Lamberts, K. (1998). Basing categorization on individuals and events. *Cognitive Psychology*, *36*, 203–272.

Bates, E., Frederici, A., & Wulfeck, B. (1987). Sentence comprehension in aphasia: A crosslinguistic study. *Brain and Language*, *32*, 19–67.

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

Becker, A. L. (1983). Toward a post-structuralist view of language learning: A short essay. *Language Learning*, *33*, 217–220.

Biber, D. (1996). Investigating language use through corpus-based analyses of association patterns. *International Journal of Corpus Linguistics*, *1*, 171–197.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. New York: Cambridge University Press.

Biber, D., & Finegan, E. (Eds.). (1994). *Sociolinguistic perspectives on register*. Oxford: Oxford University Press.

Biber, B., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, UK: Pearson Education.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*, 355–387.

Bock, J. K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, *129*, 177–192.

Bock, J. K., & Loebell, H. (1988). Framing sentences. *Cognition*, *35*, 1–39.

Bod, R. (1998). *Beyond grammar: An experience-based theory of language*. Stanford, CA: CSLI Publications.

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, *106*, 265–289.

Bowerman, M. (1976). Semantic factors in the acquisition of rules for word use and sentence construction. In D. Morehead & A. Morehead (Eds.), *Normal and deficient child language* (pp. 99–180). Baltimore, MD: University Park Press.

Boyland, J. T. (1996). *Morphosyntactic change in progress: A psycholinguistic approach*. Unpublished doctoral dissertation, University of California, Berkeley.

Boyland, J. T., & Anderson, J. R. (1998). Evidence that syntactic priming is long-lasting. In M. A. Gernsbacher (Ed.), *Proceedings of the 20th annual meeting of the Cognitive Science Society* (p. 1205). Mahwah, NJ: Erlbaum.

Brazil, D. (1995). *A grammar of speech*. Oxford: Oxford University Press.

Brent, M. R. (1999). Speech segmentation and word discovery: A computational perspective. *Trends in Cognitive Sciences*, *3*, 294–301.

Bresnan, J. (1999, August). *Linguistic theory at the turn of the century*. Plenary presentation at the 12th World Congress of Applied Linguistics, Tokyo.

Brew, C. (1995). Stochastic HPSG. *Proceedings of the seventh conference of the European Chapter of the Association for Computational Linguistics*, 83–99. Dublin: University College.

Brewster, J. (1999). Exploring gendered talk: Some effects of interactional style. In A. Tosi & C. Leung (Eds.), *Rethinking language education* (pp. 196–212). London: CILT.

Briscoe, T. (1994). Prospects for practical parsing of unrestricted text: Robust statistical parsing techniques. In N. Oostdijk & P. de Haan (Eds.), *Corpus-based research into language* (pp. 67–95). Amsterdam: Rodopi.

Brooks, P. J., Braine, M. D. S., Catalano, L., Brody, R., & Sudhalter, V. (1993). Acquisition of gender-like subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language*, *32*, 76–95.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*, 425–455.

Bybee, J. L. (2000). *Mechanisms of change in grammaticalization: The role of frequency*. Unpublished manuscript, University of New Mexico.

Bybee, J., & Hopper, P. (Eds.). (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins.

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

Bybee, J., & Thompson, S. (2000). Three frequency effects in syntax. *Berkeley Linguistic Society*, *23*, 65–85.
Carter, R. (1998). *Vocabulary: Applied linguistic perspectives* (2nd ed.). London: Routledge.
Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton de Gruyter.
Chomsky, N. (1959). Review of the book *Verbal behavior*. *Language*, *35*, 26–58.
Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, *13*, 221–268.
Church, K., & Patil, R. (1982). Coping with syntactic ambiguity or how to put the block in the box on the table. *Computational Linguistics*, *8*, 139–149.
Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 17–30.
Collins Cobuild. (1996). *Grammar patterns 1: Verbs*. London: Harper Collins.
Collins, M., & Brooks, J. (1995). Prepositional phrase attachment through a backed-off model. *Proceedings of the third Workshop on Very Large Corpora*, 27–38. Cambridge, MA: MIT Press.
Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151–216). San Diego, CA: Academic Press.
Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*, 589–608.
Coltheart, M., & Leahy, J. (1996). Assessment of lexical and nonlexical reading abilities in children: Some normative data. *Australian Journal of Psychology*, *48*, 136–140.
Corder, S. P. (1973). *Introducing applied linguistics*. New York: Penguin.
Cottrell, G., & Plunkett, K. (1994). Acquiring the mapping from meaning to sounds. *Connection Science*, *6*, 379–412.
Croft, W. (1995). Autonomy and functionalist linguistics. *Language*, *71*, 490–531.
Croft, W. (2000). *Explaining language change: An evolutionary approach*. London: Longman.
Croft, W., & Cruse, A. (1999). *Cognitive linguistics*. Unpublished manuscript, University of Manchester.
Curran, T., & Keele, S. W. (1993). Attention and non-attentional forms of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 189–202.
Daugherty, K. G., & Seidenberg, M. S. (1994). Beyond rules and exceptions: A connectionist approach to inflectional morphology. In S. D. Lima, R. L. Corrigan, & G. K. Iverson (Eds.), *The reality of linguistic rules* (pp. 353–388). Amsterdam: Benjamins.
Davies, W. D., & Kaplan, T. I. (1998). Native speaker vs. L2 learner grammaticality judgements. *Applied Linguistics*, *19*, 183–203.
de Villiers, J. G., & de Villiers, P. A. (1985). The acquisition of English. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition: Vol. 1* (pp. 27–139). Mahwah, NJ: Erlbaum.
Dörnyei, Z., & Kormos, J. (1998). Problem-solving mechanisms in L2 communication: A psycholinguistic perspective. *Studies in Second Language Acquisition*, *20*, 349–385.
Doughty, C., & Williams, J. (Eds.). (1998). *Focus on form in classroom second language acquisition*. New York: Cambridge University Press.
Ehri, L. C. (1979). Linguistic insight: Threshold of reading acquisition. In T. G. Waller & G. E. MacKinnon (Eds.), *Reading research: Advances in theory and practice: Vol. 1* (pp. 63–114). San Diego, CA: Academic Press.
Ehri, L. C. (1998). Word reading by sight and by analogy in beginning readers. In C. Hulme & R. M. Joshi (Eds.), *Reading and spelling: Development and disorders* (pp. 87–111). Mahwah, NJ: Erlbaum.
Ellis, N. C. (1993). Rules and instances in foreign language learning: Interactions of explicit and implicit knowledge. *European Journal of Cognitive Psychology*, *5*, 289–318.
Ellis, N. C. (Ed.). (1994a). *Implicit and explicit learning of languages*. San Diego, CA: Academic Press.
Ellis, N. C. (1994b). Implicit and explicit processes in language acquisition: An introduction. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 1–32). San Diego, CA: Academic Press.
Ellis, N. C. (1996a). Analyzing language sequence in the sequence of language acquisition: Some comments on Major and Ioup. *Studies in Second Language Acquisition*, *18*, 361–368.
Ellis, N. C. (1996b). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, *18*, 91–126.
Ellis, N. C. (1998). Emergentism, connectionism, and language learning. *Language Learning*, *48*, 631–664.
Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33–68). New York: Cambridge University Press.

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

Ellis, N. C., & Cataldo, S. (1990). The role of spelling in learning to read. *Language and Education*, *4*, 1–28.

Ellis, N. C., & Hicks, W. (2000, July). *Chunker 1.0—A two layer exemplar-model of reading and spelling acquisition: Cross-linguistic simulations of learning to read in English, German, Welsh, and Dutch*. Paper presented at the seventh annual conference of the Society for the Scientific Study of Reading, Stockholm, Sweden.

Ellis, N. C., & Hooper, A. M. (in press). It is easier to learn to read in Welsh than in English: Effects of orthographic transparency demonstrated using frequency-matched cross-linguistic reading tests. *Applied Psycholinguistics*, *22*.

Ellis, N. C., & Laporte, N. (1997). Contexts of acquisition: Effects of formal instruction and naturalistic exposure on second language acquisition. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 53–83). Mahwah, NJ: Erlbaum.

Ellis, N. C., Lee, M. W., & Reber, A. R. (1999). *Phonological working memory in artificial language acquisition*. Unpublished manuscript, University of Wales, Bangor.

Ellis, N. C., & Schmidt, R. (1997). Morphology and longer-distance dependencies: Laboratory research illuminating the A in SLA. *Studies in Second Language Acquisition*, *19*, 145–171.

Ellis, N. C., & Schmidt, R. (1998). Rules or associations in the acquisition of morphology? The frequency by regularity interaction in human and PDP learning of morphosyntax. *Language and Cognitive Processes*, *13*, 307–336.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, R. (1999). Item versus system learning: Explaining free variation. *Applied Linguistics*, *20*, 460–480.

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

Forster, K. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257–287). Amsterdam: North Holland.

Forster, K., & Chambers, S. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*, 627–635.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance: Vol. 12. The psychology of reading* (pp. 559–586). Mahwah, NJ: Erlbaum.

Frisch, S. F., Large, N. R., Zawaydeh, B., & Pisoni, D. B. (2001). Emergent phonotactic generalizations in English and Arabic. In J. Bybee & P. Hopper (Eds.), *Proceedings of the Symposium on Frequency Effects and Emergent Grammar* (pp. 159–180). Amsterdam: Benjamins.

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. Patterson, M. Coltheart, & J. Marshall (Eds.), *Surface dyslexia* (pp. 301–330). Mahwah, NJ: Erlbaum.

Gass, S. (1997). *Input, interaction, and the development of second languages*. Mahwah, NJ: Erlbaum.

Gernsbacher, M. A. (1994). *A handbook of psycholinguistics*. San Diego, CA: Academic Press.

Giora, R. (1997). Understanding figurative and literal language. *Cognitive Linguistics*, *8*, 183–206.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 674–691.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, A., & Sethuraman, N. (1999). *Learning argument structure generalizations*. Unpublished manuscript, University of Illinois, Urbana–Champaign.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning*, *51*, 1–50.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Science*, *4*, 178–186.

Goswami, U. (1999). Causal connections in beginning reading: The importance of rhyme. *Journal of Research in Reading*, *22*, 217–240.

Goswami, U., Gomert, J. E., & de Barrera, L. F. (1998). Children's orthographic representations and linguistic transparency: Nonsense word reading in English, French, and Spanish. *Applied Psycholinguistics*, *19*, 19–52.

Hakuta, K. (1974). Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning*, *24*, 287–298.

Harley, T. A. (1995). *The psychology of language: From data to theory*. Hove, UK: Erlbaum and Taylor & Francis.

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*, 491–528.

Harnad, S. (Ed.). (1987). *Categorical perception: The groundwork of cognition*. New York: Cambridge University Press.

Harris, Z. (1955). From phoneme to morpheme. *Language*, *31*, 190–222.

Harris, Z. (1968). *Mathematical structures of language*. New York: Wiley and Sons.

Hasher, L., & Chromiak, W. (1977). The processing of frequency information: An automatic mechanism? *Journal of Verbal Learning and Verbal Behavior*, *16*, 173–184.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, *39*, 1372–1388.

Hindle, D., & Rooth, M. (1993). Structural ambiguity and lexical relations. *Computational Linguistics*, *19*, 103–120.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*, 411–428.

Hintzman, D. L. (1988). Judgements of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*, 528–551.

Hoey, M. P. (1991). *Patterns of lexis in text*. Oxford: Oxford University Press.

Hopper, P. J. (1998). Emergent grammar. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 155–176). Mahwah, NJ: Erlbaum.

Hudson, R. A. (1990). *English word grammar*. Oxford: Blackwell.

Hudson, R. A. (1996). *Sociolinguistics* (2nd ed.). New York: Cambridge University Press.

Hulstijn, J., & DeKeyser, R. (Eds.). (1997). *Testing SLA theory in the research laboratory* [Special issue]. *Studies in Second Language Acquisition*, *19*.

Hunston, S., & Francis, G. (1996). *Pattern grammar: A corpus driven approach to the lexical grammar of English*. Amsterdam: Benjamins.

Jacoby, L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G. Bower (Ed.), *The psychology of learning and motivation: Vol. 18* (pp. 1–47). San Diego, CA: Academic Press.

James, W. (1890). *The principles of psychology: Vol. 2*. New York: Holt.

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language*, *29*, 687–715.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *20*, 137–194.

Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing: An introduction to natural language processing, speech recognition, and computational linguistics*. Englewood Cliffs, NJ: Prentice-Hall.

Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics*, *21*, 3–28.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Kaplan, R. M. (1999, August). *Computational and theoretical linguistics*. Plenary presentation at the 12th World Congress of Applied Linguistics, Tokyo.

Kellerman, E. (1995). Crosslinguistic influence: Transfer to nowhere? *Annual Review of Applied Linguistics*, *15*, 125–150.

Kemmer, S., & Israel, M. (1994). Variation and the usage-based model. In K. Beals, J. Denton, R. Knippen, L. Melnar, H. Suzuki, & E. Zeinfeld (Eds.), *Papers from the thirtieth regional meeting of the Chicago Linguistics Society: Vol. 2. Parasession on variation and linguistic theory* (pp. 165–179). Chicago: Chicago Linguistics Society.

Kempe, V., & MacWhinney, B. (1998). The acquisition of case-marking by adult learners of Russian and German. *Studies in Second Language Acquisition*, *20*, 543–587.

Kirsner, K. (1994). Implicit processes in second language learning. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 283–312). San Diego, CA: Academic Press.

Krashen, S. D. (1985) *The input hypothesis: Issues and implications*. London: Longman.

Krashen, S., & Terrell, T. (1983). *The natural approach: Language acquisition in the classroom*. Oxford: Pergamon Press.

Kuiper, K. (1996). *Smooth talkers: The linguistic performance of auctioneers and sportscasters*. Mahwah, NJ: Erlbaum.

Labov, W. (1969). *A study of non-standard English*. Champaign, IL: National Council of Teachers of English.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (1975). Empirical foundations of linguistic theory. In R. Austerlitz (Ed.), *The scope of American linguistics* (pp. 77–133). Lisse, The Netherlands: Peter de Ridder Press.

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

Lachter, J., & Bever, T. (1988). The relation between linguistic structure and associative theories of language learning: A constructive critique of some connectionist learning models. *Cognition*, *28*, 195–247.

Lado, R. (1948). Pattern practice: Completely oral? *Language Learning*, *1*, 24–28.

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. Ann Arbor: University of Michigan Press.

Lado, R. (1964). *Language teaching: A scientific approach*. New York: McGraw-Hill.

Lado, R. (1965). Memory span as a factor in second language learning. *IRAL*, *3*, 123–129.

Lado, R. (1990). Toward a lexico-semantic theory of language and language learning. *The Georgetown Journal of Languages and Linguistics*, *1*, 96–100.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

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

Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German-English comparison. *Cognition*, *63*, 315–334.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Vol. 1. Theoretical prerequisites*. Stanford, CA: Stanford University Press.

Larsen-Freeman, D. (1976). An explanation for the morpheme acquisition order of second language learners. *Language Learning*, *26*, 125–134.

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics*, *18*, 141–165.

Laxon, V., Masterson, J., & Coltheart, V. (1991). Some bodies are easier to read: The effect of consistency and regularity on children's reading. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *43*(A), 793–824.

Leech, L. (2000). Grammars of spoken English: New outcomes of corpus-oriented research. *Language Learning*, *50*, 675–724.

Lewis, M. (1993). *The lexical approach: The state of ELT and the way forward*. Hove, UK: Language Teaching Publications.

Lieven, E. V. M., Pine, J. M., & Dresner Barnes, H. (1992). Individual differences in early vocabulary development: Redefining the referential-expressive dimension. *Journal of Child Language*, *19*, 287–310.

Lightbown, P. M., Spada, N., & White, L. (Eds.). (1993). *The role of instruction in second language acquisition* [Special issue]. *Studies in Second Language Acquisition*, *15*.

Lively, S. E., Pisoni, D. B., & Goldinger, S. D. (1994). Spoken word recognition. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 265–318). San Diego, CA: Academic Press.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527.

Long, M. H. (1983). Does second language instruction make a difference? A review of research. *TESOL Quarterly*, *17*, 359–382.

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39–52). Amsterdam: Benjamins.

Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception and Psychophysics*, *39*, 155–158.

Luka, B. J. (1999). Familiarity breeds consent: Is structural facilitation evidence for the implicit knowledge of syntactic structures? (Doctoral dissertation, University of Chicago, 1999). *Dissertation Abstracts International*, *60*(6-B), 2973.

Luka, B., & Barsalou, L. (1998). The role of sentence priming on the implicit memory of syntactic structures. In M. A. Gernsbacher (Ed.), *Proceedings of the 20th annual meeting of the Cognitive Science Society* (p. 1240). Mahwah, NJ: Erlbaum.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, *9*, 157–201.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.

MacWhinney, B. (Ed.). (1987). *Mechanisms of language acquisition*. Mahwah, NJ: Erlbaum.

MacWhinney, B. (1992). Transfer and competition in second language learning. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 371–390). Amsterdam: North Holland.

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

MacWhinney, B. (1997a). Implicit and explicit processes: Commentary. *Studies in Second Language Acquisition*, *19*, 277–282.
MacWhinney, B. (1997b). Models of the emergence of language. *Annual Review of Psychology*, *49*, 199–227.
MacWhinney, B. (1997c). Second language acquisition and the competition model. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 113–144). Mahwah, NJ: Erlbaum.
MacWhinney, B. (Ed.). (1999). *The emergence of language*. Mahwah, NJ: Erlbaum.
MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. *Cognition*, *40*, 121–157.
Mandell, P. B. (1999). On the reliability of grammaticality judgement tests in second language acquisition research. *Second Language Research*, *15*, 73–99.
Marchman, V. A. (1993). Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience*, *5*, 215–234.
Marsh, G., Friedman, M. P., Welch, V., & Desberg, P. (1981). A cognitive-developmental theory of reading acquisition. In T. G. Waller & G. E. MacKinnon (Eds.), *Reading research: Advances in theory and practice: Vol. 3* (pp. 199–221). San Diego, CA: Academic Press.
Marslen-Wilson, W. D. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 148–172). Cambridge, MA: MIT Press.
Martin, W. K., Church, K., & Patil, R. (1981). *Preliminary analysis of a breadth-first parsing algorithm: Theoretical and experimental results* (Tech. Rep. No. MIT/LCS/TR-261). Cambridge, MA: Massachusetts Institute of Technology, Laboratory for Computer Science.
McEnery, T., & Wilson, A. (1996). *Corpus linguistics*. Edinburgh, UK: Edinburgh University Press.
McKoon, G., & Ratcliff, R. (1998). Memory-based language processing: Psycholinguistic research in the 1990s. *Annual Review of Psychology*, *49*, 25–42.
Mitchell, D. C., Cuetos, F., Corley, M. M. B., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research*, *24*, 469–488.
Miyake, A., Carpenter, P. A., & Just, M. A. (1994). A capacity approach to syntactic comprehension disorders: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology*, *11*, 671–717.
Morton, J. (1967). A singular lack of incidental learning. *Nature*, *215*, 203–204.
Nattinger, J. R. (1980). A lexical phrase grammar for ESL. *TESOL Quarterly*, *14*, 337–344.
Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
Nippold, M. A., & Rudzinski, M. (1993). Familiarity and transparency in idiom explanation: A developmental study of children and adolescents. *Journal of Speech and Hearing Research*, *36*, 728–737.
Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528.
Odlin, T. (1989). *Language transfer*. New York: Cambridge University Press.
Oetting, J. B., & Rice, M. L. (1993). Plural acquisition in children with specific language impairment. *Journal of Speech and Hearing Research*, *36*, 1236–1248.
Öney, B., & Durgunoglu, A. Y. (1997). Beginning to read in Turkish: A phonologically transparent orthography. *Applied Psycholinguistics*, *18*, 1–15.
Patterson, K. E., & Morton, J. C. (1985). From orthography to phonology: An attempt at an old interpretation. In K. E. Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 335–359). Mahwah, NJ: Erlbaum.
Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–225). London: Longman.
Pickering, M. J., Branigan, H. P., Cleland, A. A., & Stewart, A. J. (2000). Activation of syntactic information during language production. *Journal of Psycholinguistic Research*, *29*, 205–216.
Pine, J. M., & Lieven, E. V. M. (1993). Reanalyzing rote-learned phrases: Individual differences in the transition to multi-word speech. *Journal of Child Language*, *20*, 551–571.
Pine, J. M., & Lieven, E. V. M. (1997). Slot and frame patterns in the development of the determiner category. *Applied Psycholinguistics*, *18*, 123–138.
Pine, J. M., Lieven, E. V. M., & Rowland, C. F. (1998). Comparing different models of the development of the English verb category. *Linguistics*, *36*, 807–830.
Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, *29*, 195–247.
Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *94*, 523–568.
Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, *48*, 21–69.
Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, *83*, 304–308.
Posner, M. I., & Snyder, C. R. (1975). Attention and cognitive control. In G. H. Bower & J. T. Spence (Eds.), *Psychology of learning and motivation: Vol. 3* (pp. 55–85). San Diego, CA: Academic Press.
Prasada, S., Pinker, S., & Snyder, W. (1990, November). *Some evidence that irregular forms are retrieved from memory but regular forms are rule-governed*. Paper presented at the 31st meeting of the Psychonomic Society, New Orleans, LA.
Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford: Oxford University Press.
Redington, M., & Chater, N. (1998). Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes*, *13*, 129–192.
Rosch, E., & Mervis, C. B. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*, 192–233.
Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.
Rumelhart, D., & McClelland, J. (1986). On learning the past tense of English verbs. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models* (pp. 272–326). Cambridge, MA: MIT Press.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.
Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*, 27–52.
Sampson, G. (1980). *Schools of linguistics*. Stanford, CA: Stanford University Press.
Sampson, G. (1986). A stochastic approach to parsing. *Proceedings of COLING'86*, 151–155.
Savin, H. B. (1963). Word-frequency effects and errors in the perception of speech. *Journal of the Acoustic Society of America*, *35*, 200–206.
Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158.
Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics*, *13*, 206–226.
Schmidt, R. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165–210). San Diego, CA: Academic Press.
Schmitt, N. (2000). *Vocabulary in language teaching*. New York: Cambridge University Press.
Schütze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.
Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science*, *275*, 1599–1603.
Seidenberg, M. S., & Bruck, M. (1990, November). *Consistency effects in the generation of past tense morphology*. Paper presented at the 31st meeting of the Psychonomic Society, New Orleans, LA.
Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, *23*, 569–588.
Seidenberg, M. S., & McClelland, J. (1989). A distributed, developmental model of word acquisition. *Psychological Review*, *94*, 523–568.
Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior*, *23*, 383–404.
Shapiro, B. J. (1969). The subjective estimate of relative word frequency. *Journal of Verbal Learning and Verbal Behavior*, *8*, 248–251.
Sharwood Smith, M. (1981). Consciousness raising and the second-language learner. *Applied Linguistics*, *2*, 159–168.
Shirai, Y., & Andersen, R. W. (1995). The acquisition of tense-aspect morphology. *Language*, *71*, 743–762.
Simon, H. A. (1957). *Models of man: Social and rational*. New York: Wiley and Sons.
Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.
Skinner, B. F. (1969). *Contingencies of reinforcement: A theoretical analysis*. Englewood Cliffs, NJ: Prentice-Hall.
Slobin, D. I. (1997). The origins of grammaticizable notions: Beyond the individual mind. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition: Vol. 5* (pp. 265–323). Mahwah, NJ: Erlbaum.
Spada, N. (1997). Form-focused instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, *30*, 73–87.
Spivey-Knowlton, M., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55*, 227–267.
St. John, M. F., & Gernsbacher, M. A. (1998). Learning and losing syntax: Practice makes perfect and frequency builds fortitude. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Foreign language learning: Psycholinguistic studies of training and retention* (pp. 231–255). Mahwah, NJ: Erlbaum.
St. John, M. F., & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence*, *46*, 217–257.
Stadler, M.A., & Frensch, P. A. (Eds.). (1998). *Implicit learning handbook*. Thousand Oaks, CA: Sage Publications.
Stemberger, J. P., & MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory and Cognition*, *14*, 17–26.
Tabossi, P., Spivey-Knowlton, M. J., McRae, K., & Tanenhaus, M. K. (1994). Semantic effects on syntactic ambiguity resolution: Evidence for a constraint-based resolution process. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance: Vol. 15. Conscious and nonconscious information processing* (pp. 589–615). Cambridge, MA: MIT Press.
Tannen, D. (1987). Repetition and variation as spoken formulaicity in conversation. *Language*, *63*, 574–605.
Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language*, *26*, 608–631.
Taraban, R., & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language*, *27*, 597–632.
Taraban, R., & Roark, B. (1996). Competition in language-based categories. *Applied Psycholinguistics*, *17*, 125–148.
Tarone, E. (1988). *Variation in interlanguage*. London: Edward Arnold.
Tarone, E. (1997). Analyzing IL in natural settings: A sociolinguistic perspective of second-language acquisition. *Communication and Cognition*, *30*, 137–150.
Taylor, J. R. (1998). Syntactic constructions as prototype categories. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 177–202). Mahwah, NJ: Erlbaum.
Terrell, T. (1991). The role of grammar instruction in a communicative approach. *The Modern Language Journal*, *75*, 52–63.
Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. New York: Cambridge University Press.
Tomasello, M. (Ed.). (1998a). *The new psychology of language: Cognitive and functional approaches to language structure*. Mahwah, NJ: Erlbaum.
Tomasello, M. (1998b). The return of constructions [Review of the book *Constructions: A Construction Grammar approach to argument structure*]. *Journal of Child Language*, *75*, 431–447.
Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Science*, *4*, 156–160.
Tomasello, M., & Brooks, P. (1999). Early syntactic development: A Construction Grammar approach. In M. Barrett (Ed.), *The development of language* (pp. 116–190). London: University College London Press.
Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 145–152.
Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, *124*, 107–136.
Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, *35*, 566–585.

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285–318.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.

Underwood, B. J. (1971). Recognition memory. In H. H. Kendler & J. T. Spence (Eds.), *Essays in neobehaviorism* (pp. 313–335). New York: Appleton-Century-Crofts.

Ungerer, F., & Schmid, H. J. (1996). *An introduction to cognitive linguistics*. London: Longman.

Vihman, M. (2000). *From prosodic response to segmental patterning: The origins of representations for speech*. Unpublished manuscript, University of Wales, Bangor.

Wittgenstein, L. (1953). *Philosophical investigations*. New York: Macmillan.

Wong-Fillmore, L. (1976). *The second time around*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud. A connectionist dual process model. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1131–1161.

https://doi.org/10.1017/S0272263102002024 Published online by Cambridge University Press