

## The role of frequency in language acquisition

*Katherine Demuth*

The role of frequency in language acquisition has largely been ignored, in part by early disclaimers by both psychologists and linguists. For example, Brown (1973), in his study of morphological development by Adam, Eve and Sarah, declared that the order of grammatical morpheme acquisition was not determined by frequency. Likewise, Chomsky (1965), in defining the goal of modern formal linguistics to focus on ‘competence’ rather than ‘performance’, suggested that the collection and analysis of actual speech samples be left to the field of ‘sociology’. As a consequence, the possible role of frequency effects on the formation of linguistic representations has been largely missing from both formal linguistic and more psychological approaches to language learning. This has been consistent with a more categorical approach to the nature of grammatical representations, where sentences are either grammatical nor not.

The field began to change in the late 1980s. With the increasing availability of computers, it was possible to collect and analyze large corpora of speech. This was especially relevant for the field of acquisition, since longitudinal spontaneous speech corpora provide much of the primary data on individual children’s course of language development. It is during this time that contributions to the CHILDES database (MacWhinney 2000) also increased. A rough count (excluding recent Talkbank contributions) indicates that the database included 4 corpora in 1984, 104 in 1995, and 210 in 2006 (MacWhinney p.c.). Although each corpus varies in size, it is clear that many more researchers have access to various language acquisition corpora today.

The increased availability of computerized corpora also coincided with developments in linguistic theory that began to change the notion of categorical ‘rules’ that were never violated. Optimality Theory (Prince & Smolensky 1993, 2004) provided the opportunity for thinking about surface output forms in terms of a constraint satisfaction problem. Under an optimality-theoretic view grammatical judgments assigned a ‘?’ could be interpreted as ‘acceptable’ under certain (discourse, morphological) conditions. This notion of interacting constraints was particularly appealing to researchers of language acquisition, where change was documented over time. Thus,

rather than an abrupt parameter-setting the gradual learning curve could be understood in terms of a continual process of constraint reranking, gradually approximating the adult form.

This constraint-reranking approach to acquisition turned out to be particularly fruitful for dealing with problems of higher-level phonological acquisition, where issues of syllable and word structure could be dealt with in terms of reranking constraints, with Markedness constraints initially ranked higher than Faithfulness constraints (e.g., Demuth 1995; Gnanadesikan 1995, 2004; Pater 1997). Some scholars have now taken this further, suggesting that the constraints should be 'weighted', with frequency information playing a role in determining weights and when/how they are reranked (e.g. Boersma & Levelt 1999). Appealing to frequency effects in syllable structure was particularly useful in understanding the course of acquisition of syllable structure in Dutch (Level, Schiller & Levelt 2000). It has also been proposed to help account for language-specific differences in the timing of when coda consonants are acquired (earlier in English and German than in Spanish) and in when initial weak syllables appear (earlier in Spanish than in English or German) (e.g., Roark & Demuth 2000). Thus, both language-internal and cross-language developmental paths have been shown to be sensitive to the frequency of language-specific structures – at least in the domain of prosodic phonology. It has also been shown to occur not only in production, but also in perception, at the level of segments (e.g., Anderson, Morgan & White 2003).

Thus, at least for phonology, there appears to be some effect of frequency, where the more frequent structures are earlier acquired. This raises the possibility that more frequent structures may be easier to be learned because they are typically also 'unmarked'. Stites, Demuth & Kirk (2004), however, show that markedness and frequency do not always coincide. For example, more sonorant consonants (e.g., /l, m, n/) are less marked than less sonorant consonants (e.g., /t, d, k, g, b, p/) in the coda of a syllable, but less sonorant /t, d/ are much higher in frequency as codas in English than are the sonorant consonants. This provides a context for teasing apart the markedness and frequency issues. The authors show that most children acquire the most frequent codas first, though a few acquire the more sonorant, unmarked codas first. This suggests that some individual variation may be accounted for by adhering to either frequency or markedness patterns. Zamuner, Gerken & Hammond (2004) take this further, showing that frequency effects within the syllable rhyme are better at predicting order of acquisition. This again raises the question of what units the learner (and linguist) are computing over.

But why might children acquire more frequent structures first? One possibility is that the more frequently a certain linguistic unit occurs, the harder it is for the learner to ignore it. That is, from an information-theoretic perspective, if it occurs often, it must be attended to, it is not 'noise' in the system. This is easy to capture in optimality theoretic terms. Imagine a young learner who produces no codas, such that *dog* is realized as /da/. If 60% of the syllables the child hears contain a coda consonant, this means that the majority of the syllables the child utters will be ill-formed, resulting in massive Faithfulness violations. Since this is a lot of ill-formedness, the English-speaking child will probably be faster to start producing coda consonants than a child learning Spanish, where only 25% of syllables have codas. Thus, we predict that, for most English learners, the acquisition of complex syllable structure will proceed more rapidly than it does in Spanish, and this appears to be the case (cf. Demuth, Culbertson & Alter 2006; Lleó 2003).

Thus, perhaps the high frequency of a particular form merely forces the child to attend to a particular grammatical structure earlier than they otherwise might. This is known in the adult psycholinguistics literature as a 'priming' effect (e.g., Bock 1986; Bock & Loebell 1990), which has been shown to effect adults' use of syntactic constructions. That is, if adults are previously exposed to a certain syntactic construction, they are more likely to use it in subsequent speech. This raises the possibility that frequency effects occur not only in children's developing phonologies, but also in their syntax. A number of results suggest that this is the case.

One of the first studies to suggest that frequency played a role in children's syntax came from Demuth's (1989) study of passive acquisition in the Bantu language Sesotho. This study showed that passives are much more frequent in Sesotho child-directed speech than they are in English. Demuth argued that this much higher frequency of Sesotho passives could help explain why passives are acquired in this language by the age of 2;8. Furthermore, this also predicted that even English-speaking children could show competence with such grammatical structures under appropriate priming conditions. Brooks & Tomasello (1999) have recently shown that this is the case, eliciting passives from 3-year-olds with novel verbs. These findings are not surprising for those who work on adult psycholinguistics, but they have been less routinely explored in the domain of language acquisition. The frequency of different grammatical structures is part of our linguistic competence, and influences not only how we process and produce language, but also which aspects of language will be learned first, and which aspects are most likely to change (e.g., Bybee & Hopper 2001).

We should therefore be able to make frequency-based predictions about when certain grammatical structures might appear in different languages – all else being equal. But all else is not always equal. For example, French also has codas in only 25% of syllables, but coda consonants appear to be acquired earlier than in Spanish. However, coda consonants are more likely to be preserved in stressed rather than unstressed syllables and in monosyllables rather than disyllables – at least by English speaking children (Demuth, Culbertson & Alter 2006; Kirk & Demuth 2006). Thus, frequency alone cannot explain this cross-linguistic difference. We suggest that, in this case, the longer average word length of Spanish, plus the stress differences between the two languages, account for the differences here.

This again raises the problem how to count. That is, are learners (which we DO assume are keeping track of all kinds of statistics) computing statistics over tokens (as often assumed in the acquisition literature) or types? And what are the linguistic units over which they are keeping these statistics? Recent modeling work indicates that in the domain of morphological segmentation, type information is more useful than token information (Goldwater 2006). Obviously, further modeling and empirical research is needed to determine what learners are counting, how they are counting, and how this triggers perceptual awareness and comprehension/production competence in different linguistic domains.

All these findings suggest the need for an intergraded model of language acquisition, where competing constraints (syntactic, semantic, discourse, phonological, prosodic, processing, frequency effects, etc.) all play a role in determining how and when which aspects of language are acquired. The frequency with which a particular grammatical construction occurs in a child's experience is therefore also part of that child's knowledge of language. Only when all these other constraints are controlled for can we really look at 'syntactic' competence in isolation.

Syntactic experiments are much better designed today than they were in the past, controlling for more of the possible discourse/pragmatic, semantic, and processing confounds. Researchers are also more aware of processing constraints, controlling for sentence length and complexity. Some also take lexical frequency into account, ensuring that children know the vocabulary items to be used. Less controlled are prosodic effects, which play a significant role in when and how certain grammatical function items (such as determiners) are acquired (e.g., Gerken 1996; Demuth 2007; Tremblay & Demuth 2007). Frequency is thus only one of the factors that can influence when and how learners demonstrate knowledge of grammar.

## References

- Anderson, Jennifer, James L. Morgan & Katherine S. White  
2003 A statistical basis for speech sound discrimination. *Language and Speech* 46: 155–182.
- Bock, Kathryn  
1986 Syntactic persistence in language production. *Cognitive Psychology* 18: 355–387.
- Bock, Kathryn & Helga Loebell  
1990 Framing Sentences. *Cognition* 35: 1–39.
- Boersma, Paul & Clara Levelt  
1999 Gradual Constraint-Ranking Learning algorithm predicts acquisition order. In *Proceedings of the 30<sup>th</sup> Child Language Research Forum*, Eve V. Clark (ed.), 229–237. Stanford: The University of Chicago Press (CSLI: Center for the Study of Language and Information).
- Brooks, Patricia & Michael Tomasello  
1999 Young children learn to produce passives with nonce verbs. *Developmental Psychology* 35: 29–44.
- Brown, Roger  
1973 *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Bybee, Joan & Paul Hopper (eds.)  
2001 *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins.
- Chomsky, Noam  
1965 *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Demuth, Katherine  
1989 Maturation and the acquisition of Sesotho passive. *Language* 65: 56–80.  
1995 Markedness and the development of prosodic structure. In *Proceedings of the North East Linguistic Society* 25, Jill Beckman (ed.), 13–25. Amherst, MA: GLSA, University of Massachusetts.  
2007 Acquisition at the prosody-morphology interface. In *Proceedings of the 2<sup>nd</sup> Conference on Generative Approaches to Language Acquisition North America (GALANA)*, Alyona Belikova, Luisa Meroni & Mari Umeda (eds.), 84–91. Somerville, MA: Cascadia Proceedings Project.
- Demuth, Katherine, Jennifer Culbertson & Jennifer Alter  
2006 Word-minimality, epenthesis, and coda licensing in the acquisition of English. *Language and Speech* 49: 137–174.
- Gerken, LouAnn  
1996 Prosodic structure in young children's language production. *Language* 72: 683–712.

- Gnanadesikan, Amalia  
2004 Markedness and faithfulness constraints in child phonology. In *Constraints in phonological acquisition*, René Kager, Joe Pater & Wim Zonneveld (eds), 73–109. Cambridge: Cambridge University Press.
- Goldwater, Sharon  
2006 *Nonparametric Bayesian Models of Lexical Acquisition*. Unpublished Ph.D. diss., Brown University.
- Kirk, Cecilia & Katherine Demuth  
2006 Accounting for variability in 2-year-olds' production of coda consonants. *Language Learning and Development* 2: 97–118.
- Levelt, Clara, C., Niels O. Schiller & Willem Levelt  
2000 The acquisition of syllable types. *Language Acquisition* 8: 237–264.
- Lleó, Conxita  
2003 Prosodic licensing of codas in the acquisition of Spanish. *Probus* 15: 257–281.
- MacWhinney, Brian  
2000 *The CHILDES Project: Tools for analyzing talk. 3<sup>rd</sup> Edition. Vol 2: The Database*. Mahwah, NJ: Erlbaum.
- Pater, Joe  
1997 Minimal violation and phonological development. *Language Acquisition* 6: 201–253.
- Prince, Alan & Paul Smolensky  
1993/ /2004 *Optimality Theory: Constraint Interaction in Generative Grammar*. Cambridge, MA: Blackwell.
- Roark, Brian & Katherine Demuth  
2000 Prosodic constraints and the learner's environment: A corpus study. In *Proceedings of the 24th Annual Boston University Conference on Language Development*, S. Catherine Howell, Sarah A. Fish & Thea Keith-Lucas (eds.), 597–608. Somerville, MA: Cascadilla Press
- Stites, Jessica, Katherine Demuth & Cecilia Kirk  
2004 Markedness versus frequency effects in coda acquisition. In *Proceedings of the 28th Annual Boston University Conference on Language Development*, Alejna Brugos, Linnea Micciulla & Christine E. Smith (eds.), 565–576. Somerville, MA: Cascadilla Press.
- Tremblay, Annie & Katherine Demuth  
2007 Prosodic licensing of determiners in children's early French. In *Proceedings of the 2<sup>nd</sup> Conference on Generative Approaches to Language Acquisition North America (GALANA)*, Alyona Belikova, Luisa Meroni & Mari Umeda (eds.), 426–436. Somerville, MA: Cascadilla Proceedings Project.
- Zamuner, Tania, LouAnn Gerken & Michael Hammond  
2004 Phonotactic probabilities in young children's speech production. *Journal of Child Language* 31: 515–536.