Encyclopedia of Language Development

Frequency Effects in Language Development

Contributors: Holger Diessel
Editors: Patricia J. Brooks & Vera Kempe
Book Title: Encyclopedia of Language Development
Chapter Title: "Frequency Effects in Language Development"
Pub. Date: 2014
Access Date: September 21, 2014
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9781452258768
Online ISBN: 9781483346441

DOI: http://dx.doi.org/10.4135/9781483346441.n71
Print pages: 222-224

http://dx.doi.org/10.4135/9781483346441.n71*University of Jena*
Until recently, frequency received little attention in child language research. In the nativist approach, it is generally assumed that language acquisition involves a universal set of innate categories that are not immediately affected by frequency. However, there is now increasing evidence that language development is crucially influenced by the child's experience with concrete words and utterances and that frequency of occurrence plays an important role in the acquisition of grammatical categories.

Two general types of frequency are commonly distinguished to characterize the effects of frequency in language: (1) token frequency, that is, the repeated occurrence of a particular word or construction (in a given corpus or time frame); and (2) type frequency, that is, the number of distinct types that are subsumed under a particular category. For instance, this entry includes 22 tokens of the noun *frequency* and more than 300 different types of the category noun (*e.g., frequency, language*, and *development*).

Token frequency strengthens the representation of linguistic elements in memory, which in turn facilities their activation in future language use. For instance, the more often a speaker encounters a technical term such as *neologism*, the more easily he or she will remember it. Type frequency, by contrast, is assumed to correlate with linguistic productivity. In general, the more types that are subsumed under a category, the higher the probability that it will be extended to novel uses. For instance, the English past tense is commonly expressed by the suffix *-ed* (e.g., *walk/walked*), but there are also several irregular past-tense forms (e.g., *sing/sang, keep/kept*, and *hit/hit*). Because the regular past tense subsumes a much larger number of verb types than the irregular past tense forms, it is more likely to be applied to novel verbs than any of the irregular forms (e.g., *google/googled*).

One aspect of language development that has been analyzed in light of (token) frequency is the order of acquisition. There is evidence from a wide range of studies that the frequency of linguistic elements in the ambient language (i.e., the language children experience) correlates with their (first) appearance in child language. However, as frequency is not the only factor influencing the order of acquisition, frequent expressions do not generally appear prior to infrequent ones. In order to assess the impact of frequency on the course of language development, child-language researchers have investigated input frequency in conjunction with other aspects of the

ambient language that may affect the order in which linguistic elements are learned, for example, the complexity of words and constructions, their communicative functions and perceptual salience, and their semantic and structural relationships to other items and constructions. However, because some of these factors are not independent of frequency, it is often difficult to determine the relative effect of frequency on the order of acquisition.

In addition, frequency plays an important role in children's distributional analysis of the ambient language. Language is a sequential medium in which morphemes, words, and phrases are combined to strings of linguistic elements. Analyzing these strings is an important aspect of language development that is crucially influenced by (type and token) frequency. The more often two or more linguistic elements are combined to a string, the stronger the connection is between them, which can be expressed in terms of transitional probabilities or related statistical measures.

For instance, because the word *blue* is more frequently combined with *sky* than with *sun, blue sky* constitutes a tighter string than *blue sun*, as indicated by the fact that the transitional probability between *blue* and *sky* is higher than that between *blue* and *sun.* **[p. 222 ↓]** The same approach can be used to analyze the distributional properties of grammatical categories: As adjectives are usually placed before nouns, the string adjective-noun (e.g., *the best solution*) is more tightly organized than the string noun-adjective (e.g., *the best solution possible*), which has a very low transitional probability in English.

Because the linguistic strings of the ambient language include many inconsistencies (e.g., false starts, hesitation signals, and repairs), it was long assumed that distributional cues (or transitional probabilities) are only of minor importance for language learning. However, recent corpus studies revealed that the ambient language is much more regular in this regard than previously assumed. For instance, Martin Redington, Nick Chater, and Steven Finch showed that there are strong distributional cues for grammatical categorization in the language preschool children experience, which could help them to acquire syntactic categories. For instance, as nouns and verbs are accompanied by different function morphemes, children could use the distribution of frequent grammatical markers such as *a, the, is, has*, and *will* to learn the grammatical distinction between nouns and verbs. Subsequent studies demonstrated that the

distributional cues of the ambient language are even more powerful if they are analyzed together with phonological and semantic features.

Complementary to this line of research, child-language researchers have investigated children's ability to recognize distributional regularities in continuous speech. The results of this research suggest that young children are extremely sensitive to transitional probabilities between adjacent words and categories. For instance, in a seminal study, J. Saffran, R. Aslin, and E. Newport exposed 8-months-old infants to four nonce words that were combined to a continuous string of meaningless elements (e.g., *tupiro-bidakupadoti-bidaku*). After training, the infants were tested under two conditions. In one condition, they listened to a new string of the same four nonce words in random order (e.g., *bidaku-tupiro-golabu-padoti*), and in the other condition, they heard a random string of syllables that were formed from the syllables of the four nonce words (e.g., *bi-tu-ro-da-ti-ku-ti-go-pa-do*). Because there were fewer words than syllables, the transitional probability between words (condition 1) was higher than that between syllables (condition 2). Although the infants had heard the nonce words for only two minutes during training, they noticed the different transitional probabilities in the two conditions, suggesting that they had identified the four nonce words as conventionalized linguistic units.

Subsequent studies showed that young children are also able to extract syntactic categories from continuous speech based on distributional regularities. In one of these studies, Rebecca Gomez and LouAnn Gerken taught 12-month-old infants an artificial language consisting of a set of nonce words that appeared in five different sentence types distinguished by word order. The study showed that children are able to abstract away from the particular properties of individual words and to subsume them under general syntactic categories based on similar (or identical) distributional properties.

Together, these studies provide strong evidence for the importance of frequency in (first) language acquisition. Frequency is the driving force of general cognitive processes that are involved in inductive learning and that are not only relevant for language development but also for many other cognitive phenomena that are grounded in experience.

HolgerDiessel*University of Jena*

http://dx.doi.org/10.4135/9781483346441.n71
See Also:

- Distributional Knowledge and Language Learning
- Grammatical Categories
- Induction in Language Learning
- Item-Based/Exemplar-Based Learning
- Multiple Cues in Language Acquisition
- Statistical Learning

Further Readings

Bybee, J., ed. and P. Hopper, eds. Frequency and the Emergence of Linguistic Structure . Amsterdam: John Benjamins, 2001. http://dx.doi.org/10.1075/tsl.45

Diessel, H. “Frequency Effects in Language Acquisition, Language Use, and Diachronic Change.” New Ideas in Psychology , v. 25 (2007). http://dx.doi.org/10.1016/j.newideapsych.2007.02.002

Ellis, N. C. “Frequency Effects in Language Processing. A Review With Implications for Theories of Implicit and Explicit Language Acquisition.” Studies in Second Language Acquisition , v. 24 (1996).

Gomez, R. L. and L.Gerken. “Infant Artificial Language Learning and Language Acquisition.” Trends in Cognitive Science , v. 4 (2000). http://dx.doi.org/10.1016/S1364-6613%2800%2901467-4

Marcus, G. F., S.Vijayan, S. B.Rao, and P. M.Vishton. “Rule Learning by Seven-Months-Old Infants.” Science , v. 283 (1999). http://dx.doi.org/10.1126/science.283.5398.77

Monaghen, P., N.Chater, and M.Christiansen. “The Differential Role of Phonological and Distributional Cues in Grammatical Categorization.” Cognition , v. 96 (2005).

Redington, M., N.Chater, and S.Finch. “Distributional Information: A Powerful Cue for Acquiring Syntactic Categories.” Cognitive Science , v. 22 (1998). http://dx.doi.org/10.1207/s15516709cog2204_2

Saffran, J. R., R. N.Aslin, and E. L.Newport. “Statistical Learning in 8-Months-Old Infants.” Science , v. 274 (1996). http://dx.doi.org/10.1126/science.274.5294.1926