ELSEVIER

New Ideas in Psychology 25 (2007) 108–127

NEW IDEAS IN PSYCHOLOGY

www.elsevier.com/locate/newideapsych

Frequency effects in language acquisition, language use, and diachronic change

Holger Diessel*

Friedrich-Schiller-Universitat Jena, Institut fur Anglistik/Amerikanistik, Ernst-Abbe-Platz 8, 07743 Jena, Germany

Available online 18 April 2007

Abstract

Recent work in psychology and linguistics has shown that frequency of occurrence is an important determinant of language acquisition, language use, and diachronic change. This paper surveys the effects of frequency on the use and development of language and considers the psychological mechanisms that underlie the various frequency effects. The paper shows that frequency has an impact on the emergence of linguistic structure and that some well-known cross-linguistic tendencies arise from frequency effects.

0. Introduction

One of the most basic principles of modern linguistics is the rigid division between grammar and language use, which is reflected in Chomsky's dichotomy of competence and performance. Based on this division, it is commonly assumed that grammatical structures are independent of their use; that is, in this approach grammar is seen as a closed and stable system that is not affected by pragmatic and psycholinguistic principles involved in language use.

This view of grammar has been challenged in recent years, both in psychology and linguistics. Inspired by research with connectionist models, psychologists and cognitive scientists have argued that grammatical structures emerge from processing linguistic data. In this approach, grammar is a dynamical system that is constantly changing by virtue of psychological processes involved in language use (cf. Elman et al., 1996; Elman &

*Tel.: +49 3641 944530.
E-mail address: holger.diessel@uni-jena.de.

0732-118X/$ - see front matter
doi:10.1016/j.newideapsych.2007.02.002

McClelland, 1984; MacDonald, Pearlmutter, & Seidenberg, 1994; Tabor, Juliano, & Tanehnaus, 1997).

A similar model of grammar has been proposed in cognitive and functional approaches to the study of language. Based on data from language acquisition, corpus linguistics, grammaticalization, and linguistic typology, functionally oriented linguists have argued that linguistic structure is shaped by pragmatic and cognitive pressures that influence the use of language (cf. Barlow & Kemmer, 2000; Bod, Hay, & Jannedy, 2003; Bybee, 1985, 1995, 2001; Bybee & Hopper, 2001; Diessel, 2004; Hawkins, 2004; Hopper, 1987).

One aspect that all dynamical models of grammar emphasize is that frequency of occurrence is an important determinant of linguistic structure and language use. There is a wealth of recent results suggesting that frequency has an impact on the comprehension, production, and emergence of linguistic categories and rules (cf. Bod et al., 2003; Bybee and Hopper, 2001). This paper discusses the influence of frequency on the use and structure of language and considers the psychological mechanisms that underlie the various frequency effects.

The paper is divided into five sections. Section 1 is concerned with statistical approaches to language acquisition, Sections 2 and 3 deal with frequency effects in sentence comprehension and production, Section 4 considers the effect of frequency in diachronic change, and Section 5 discusses some cross-linguistic tendencies that arise from frequency effects.

1. Language acquisition

In generative linguistics, it is commonly assumed that grammar cannot be learned from experience alone. Specifically, it has been claimed that the ambient language does not provide sufficient information to extract grammatical categories and constraints from the input. This has become known as the argument from the poverty of the stimulus, which is perhaps the best-known argument for linguistic innateness (cf. Crain & Pietroski, 2001). According Chomsky (1965, p. 78), there is an enormous gap between the intricacies of human grammar and the data that children encounter in the ambient language, which can only be closed because children are endowed with an innate universal grammar.

This view of grammatical development has been challenged in recent corpus-based studies of the ambient language (cf. Mintz, Newport, & Bever, 2002; Monaghan, Chater, & Christiansen, 2005; Redington, Chater, & Finch, 1998). What all of these studies have shown is that there is much more information in the input than commonly assumed. For instance, Redington et al. (1998) have shown that the ambient language provides a wealth of distributional information that could help the child to learn the basic word classes. Using data from the English corpus of the CHILDES database (cf. MacWhinney, 2000), Redington et al. examined the distributional properties of the 1000 most frequent words in the input data. For each target word, they collected Bigram statistics to determine the context; that is, they considered the two words preceding a target word and determined how often a particular target word occurred after the 150 most frequent context words, i.e. the two words preceding the target word. If we represent this information in a contingency table and combine the frequencies in the rows into 'context vectors', we obtain numerical representations of the distributional properties of the target words (cf. Table 1).

Based on these representations, Redington et al. conducted a cluster analysis that grouped the context vectors into lexical classes based on their (numerical) similarities.

Table 1
Contingency table of Bigram statistics [illustrative example] (Redington et al., 1998)

	Context 1 (in the __)	Context 2 (a big __)	Context 3 (I am __)	Context 4 (has been __)	Context vectors
Target word 1	210	321	2	0	210-321-2-0
Target word 2	376	917	1	5	376-917-1-5
Target word 3	0	1	1078	1298	0-1-1078-1298
Target word 4	1	4	987	1398	1-4-987-1398
etc.					

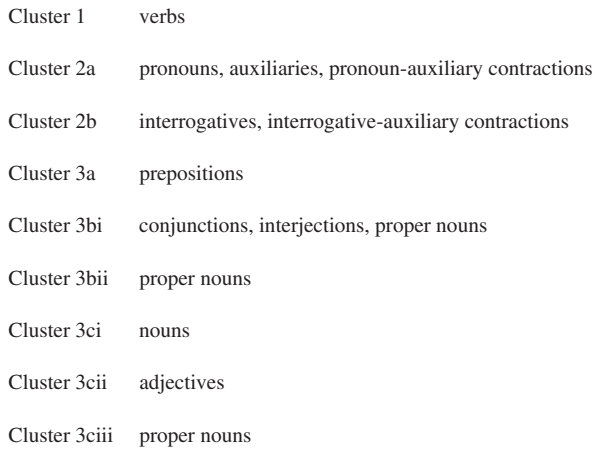

Fig. 1. Result of cluster analysis (Redington et al., 1998, p. 443).

As you can see in Fig. 1, the words that are combined in particular clusters correspond quite closely to the traditional word classes of English grammar, suggesting that distributional frequencies may play a crucial role in the acquisition of grammatical categories. Even if we disregard semantic and pragmatic information, children would find enough (distributional) information in the ambient language to learn the traditional parts-of-speech.

Using a similar methodological approach, Mintz et al. (2002) showed that information about phrasal boundaries can improve the categorization of particular word classes. In contrast to Redington et al., they defined the beginning of the context of a target word by the last function word preceding it. If we assume that children are able to draw on this information, i.e. if we assume that children are able to recognize phrasal boundaries, the distributional analysis is even more powerful to solve the categorization task.[1]

Elaborating this line of research, Monaghan et al. (2005) conducted an analysis in which they investigated the differential role of phonological and distributional cues to linguistic categorization. Using adult data from the CHILDES database, which they transformed into phonological transcriptions, they showed that there are strong correlations between

[1] As Jusczyk (1997) and colleagues have shown, even 1-year old infants are able to differentiate function words from content words based on phonological features. Since function words tend to occur at the beginning of a phrase, children could use this information to determine phrasal boundaries.

Table 2
Materials and conditions (Saffran et al., 1996)

a.	Nonce words:	(i) tupiro (ii) golabu (iii) bidaku (iv) padoti
b.	Training string:	tupiro-bidaku-padoti-bidaku-golabu …
c.	Transitional probabilities:	[tu–pi–ro]—[bi–da–ku]—[pa–do–ti]—[bi–da–ku]—… 1.0 1.0 0.25 1.0 1.0 0.25 1.0 1.0 0.25 1.0 1.0 0.25
d.	Test strings:	Condition 1: [bi–da–ku]— [tu–pi–ro]— [pa–do–ti] …
		Condition 2: ro–pa–ku–pi–do–da–tu–ku–go–ro–bu …

certain phonological features and particular word classes in the ambient language. Specifically, they demonstrated that the categorical distinctions between nouns and verbs and open class items and closed class items correlate with several phonological features: the number of syllables, the occurrence of reduced vowels, the presence of stress, the complexity of the onset, the occurrence of final voicing, and the occurrence of certain speech sounds (cf. Kelly, 1992). Interestingly, the phonological features do not just reinforce the information that children may extract from distributional regularities, but seem to be especially powerful in lexical domains in which distributional information is not so easily available. Distributional cues are especially useful for the categorization of high-frequency items, which children encounter many times in the same context; but they are less useful for the categorization of low-frequency words, which are not frequent enough to be associated with a particular syntactic context. However, since low-frequency items tend to be longer and phonologically more complex than high-frequency items, they provide more phonetic information, which can help the child to categorize linguistic expressions that are less frequent.

What all of these studies have demonstrated is that the ambient language provides a rich source of information that the child could use to learn grammatical categories. But do children really draw on this information? Given the complexity of a cluster analysis, one might doubt that young children are able to perform the necessary computations. However, recent experimental studies with infants suggest that children are extremely talented in detecting distributional patterns (cf. Jusczyk, 1997; Saffran, 2001; Saffran, Aslin, & Newport, 1996; Santelmann & Jusczyk, 1998). For instance, Saffran et al. (1996) showed that infants as young as 8 months are able to determine transitional probabilities in continuous speech. Transitional probabilities indicate the likelihood that an element X will occur after Y. Using four nonce words, each consisting of three syllables, they exposed their infants to a string of speech composed of the four nonce words in random order (cf. Tables 2a and b). After only 2 min of training, the infants were tested under two separate conditions (cf. Table 2c). In condition 1, they listened to a new string of the four nonce words combined in random order; and in condition 2, they listened to a string of randomly combined syllables that were obtained by decomposing the four nonce words. Since there are more syllables than words, the transitional probabilities are higher in condition 1 than in condition 2 (cf. Table 2d).[2] Using the listening-preference procedure

[2]Inside of a word, the transitional probability is 1.0, because after a given syllable there is only one particular syllable that can follow, but at the end of a word the transitional probability is 0.25 because the next string can continue with each of the four words.

(Jusczyk, 1997), Saffran et al. found that infants listened significantly longer to the string of syllables (condition 2) than to the string of nonce words (condition 1), suggesting that they noticed the higher transitional probability in condition 1. Put differently, the children's responses can be interpreted as evidence for the hypothesis that they had learned the four nonce words as linguistic units.

If this is correct, it seems reasonable to assume that statistical information plays a key role in grammatical development. Moreover, one might hypothesize that what children learn is a probabilistic grammar grounded in the child's language experience. In such an experienced-based grammar, linguistic categories and linguistic structures are associated with activation (or probability) values that are determined by their relative frequencies in language use (cf. Bod, 2003; Bod et al., 2003; Elman et al., 1996; Jurafsky, 1996; Diessel & Tomasello, 2005).

2. Sentence comprehension

In accordance with this view of grammar and grammatical development, recent work on sentence comprehension has argued that the interpretation of linguistic structures is crucially influenced by the speaker's past language experience (cf. Jurafsky, 1996; Jurafsky & Martin, 2000; MacDonald et al., 1994; Tabor et al., 1997). More precisely, these studies claim that frequency information is essential to deal with the occurrence of syntactic ambiguities. For instance, one type of ambiguity that has been subject to numerous investigations involves the attachment of a prepositional phrase (PP), which can be ambiguous between two interpretations: If a PP follows a noun phrase (NP), it can be either attached to the preceding noun (phrase) (cf. [V [NP [PP]]) or to the verb phrase [V [NP] [PP]]. Interestingly, several studies have found that the attachment site of an ambiguous PP varies with the occurrence of particular lexical items (cf. Ford, Bresan, & Kaplan, 1982; Spivey-Knowlton & Sedivy, 1995; Taraban & McClelland, 1988). Consider for instance the sentences in (1a-b), which Ford et al. (1982) used in an off-line experiment in which subjects had to perform a forced choice between two meanings.

(1) a. The woman discussed [the dogs [on the beach]].
b. The woman kept [the dogs] [on the beach].

Although both sentences have the same surface structure, 90% of the participants interpreted the PP in (1a) as an attribute of the preceding NP, i.e. they attached the PP to NP, whereas 95% interpreted the PP in (1b) as an immediate constituent of the verb phrase (VP), i.e. they attached the PP to VP. Since the test items are identical except for the verb, it must have been the verb that caused the different responses. But what distinguishes the two verbs *discuss* and *keep*?

One factor that seems to be important is their meaning. Note that the meaning of *keep* varies with the syntactic structure, but the meaning of *discuss* is basically the same in the two interpretations. However, Ford et al. argue that while semantic factors are important there is another, related factor that affected their subjects' responses. If we look at the frequency of the two subcategorization frames we find that while *discuss* predominately occurs with a single NP complement, *keep* tends to occur with two immediate constituents. Using data from the Penn Treebank, Jurafsky (1996) reports that three out of four instances of *discuss* occur with a single NP complement, whereas four out of five instances

of *keep* occur with a PP attached to the verb. In other words, *discuss* and *keep* tend to occur in different subcategorization frames, which is eventually motivated by their meanings, but may affect the interpretation of ambiguous PPs as an independent component. Specifically, one might hypothesize that other things being equal people tend to activate the syntactic structure they have encountered most frequently.

A similar analysis has been proposed for the resolution of ambiguous complements (cf. Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Hare, McRae, & Elman, 2004; Jurafsky, 2003; Roland, Elman, & Ferreira, 2006; Roland & Jurafsky, 2002; Trueswell, Tanenhaus, & Kello, 1993; Wiechmann, 2006). As can be seen in (2a-d), many transitive verbs can take both an NP-complement and an S-complement.

(2) a. John remembered [the proposal]$_{NP}$.
b. John remembered [the proposal was rejected by Mary]$_{S}$.
c. The police officer suspected [his motifs] $_{NP}$.
d. The police officer suspected [his motives would turn out to be true]$_{S}$.

However, individual verbs are often strongly biased towards one of the two types of complements. For instance, *remember* occurs primarily with nominal complements, while *suspect* occurs more frequently with complement clauses (cf. Jurafsky, 1996). Several experimental studies have argued and presented evidence that an NP-bias verb such as *remember* causes processing difficulties if it occurs with a sentential complement; whereas an S-bias verb such as *suspect* causes additional processing costs if it occurs with a simple NP-complement. Again, the processing effect can be explained by the combined effect of the verb's meaning and the speaker's experience with the verb. If a verb is commonly used with a particular subcategorization frame, which is ultimately motivated by semantic considerations, this frame is automatically activated in on-line processing.[3]

That frequency plays an important role in on-line processing has also been suggested in a study by Juliano and Tanenhaus (1993), who investigated ambiguous structures including the word *that* (see also Tabor et al., 1997). As can be seen in (3a-d), *that* serves a variety of syntactic functions in different contexts. It can be a demonstrative pronoun (3a), a demonstrative determiner (3b), a complementizer (3c), or a relative marker (3d).

(3) a. *That* is your seat.
b. *That* book is boring.
c. Peter told me *that* Sally can't come.
d. Joe showed me the cat *that* was chasing the dog.

Overall, *that* is most frequently used as a demonstrative pronoun; but interestingly the frequencies of the four categories vary with the syntactic context. For instance, Juliano and Tanenhaus (1993) report that in the Brown Corpus 89% of all tokens of *that* are demonstratives if *that* occurs at the beginning of a sentence and only 11% are complementizers; but after verbs 93% of all tokens of *that* are complementizers and only

[3] Hare et al. (2004) have shown that semantically ambiguous verbs such as *find* (cf. *We found a solution—We find that she is a genius*) may be biased towards different subcategorization frames, suggesting that it is the lexical meaning rather than the lexical form that influences the automatic activation of a particular subcategorization frame (cf. Wiechmann, 2006).

Table 3
Mean reading times (msec) at specified word positions (Juliano & Tanenhaus, 1993, p. 594)

Condition	that	cheap	hotels	was/were
(5a) DEM	414	453	484	**492**
(5b) COMP	409	470	479	433
(5c) DEM	399	456	454	438
(5d) COMP	381	461	479	563

7% are demonstratives. Thus, one might hypothesize that the context-based frequencies may influence the interpretation of *that* and the associated syntactic structure. In order to test this hypothesis, Juliano and Tanenhaus conducted a self-paced reading study using sentences in which *that* appeared (i) after a transitive verb (4a-b) or at the beginning of a sentence (4c-d). In both conditions, *that* occurred with two different syntactic functions: In sentences (4a) and (4c), it functions as a demonstrative determiner, and in sentences (4b) and (4d) it functions as a complementizer.

(4) a. The lawyer insisted [$\varnothing$ [*that* cheap hotel]$_{NP}$ was clean and comfortable]$_S$. DEM
b. The lawyer insisted [*that* [cheap hotels]$_{NP}$ were clean and comfortable]$_S$. COMP
c. [*That* cheap hotel]$_{NP}$ was clean and comfortable to our surprise. DEM
d. [*That* [cheap hotels]$_{NP}$ were clean and comfortable]$_S$ surprised us. COMP

As can be seen in Table 3, after verbs (cf. 4a-b) the average reading times of the copula, i.e. the element following the ambiguous NP, were particularly long when *that* functioned as a demonstrative (cf. 4a); but at the beginning of the sentence (cf. 4c-d) reading times were especially long when *that* functioned as a complementizer (cf. 4d). Juliano and Tanenhaus argue that the longer reading times are due to the relative frequencies of the demonstrative and the complementizer in the two contexts. If *that* occurs at the beginning of a sentence there is a bias to interpret it as a demonstrative, but if it occurs after a verb there is a bias to interpret it as a complementizer. If the rest of the sentence is incompatible with the initial parse (i.e. the language user's expectation), as in examples (4a) and (4d), the comprehender has to revise the initial interpretation, resulting in prolonged reading times.

In sum, there is strong evidence that frequency plays an important role in sentence comprehension. It seems that language users store an enormous amount of statistical information, determining their linguistic expectations, which in turn play an important role in syntactic ambiguity resolution.

3. Language production

Like sentence comprehension, language production is crucially influenced by the speaker's past language experience. The most obvious frequency effect in language production is that frequently used expressions are often phonetically reduced. There is a well-known correlation between the token frequency of linguistic expressions and their length or phonetic substance. Across languages, frequently used expressions tend to be short because they have been phonetically reduced (cf. Bybee, 2001; Hooper, 1976; Zipf, 1935).

How do we account for this correlation? Why are frequently used expressions often phonetically reduced? Jurafsky, Bell, Gregory, and Raymond (2001) provide a psycholinguistic explanation: speakers tend to produce linguistic expressions with minimal effort, but have to ensure that the hearer is able to understand what they mean. Since frequently used expressions are more expected (or more predictable) they may be identified even if they are phonetically reduced (see also Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky, 1999).

In a classical paper, Pollack and Pickett (1964) have shown that if we splice words out of continuous speech, only 50% of them are identifiable in isolation; that is, about 50% of all words are phonetically so impoverished that they are not recognizable. But people are still able to understand each other because they also rely on top–down information. These top–down processes involve semantic and pragmatic aspects, but are also determined by frequency. Other things being equal, frequently used expressions are more easily predictable, and thus more easily recognizable, than infrequent expressions, which explains why frequent expressions are more often reduced than infrequent ones. If the speaker reduces an infrequent expression, the hearer may not be able to interpret the reduced stimulus; but if he reduces a frequent expression, the hearer may be able to identify the underspecified term because high-frequency expressions are expected to occur in particular contexts. In addition, frequently used expressions may undergo phonetic reduction because speakers have more practice in producing them.

The correlation between frequency and phonetic reduction is especially striking in linguistic sequences. For instance, Krug (1998) has shown that the contraction of auxiliary verbs (e.g. *I've, he's, we'll*) varies with the string frequency of the subject and the auxiliary. String frequency is defined as the joint frequency of two words, X and Y (which can be normalized by dividing the frequency of the dyad, i.e. XY, by the total number of words in the corpus). Using data from the London Lund Corpus and the Bank of English Corpus, Krug showed that the occurrence of contracted auxiliaries varies with the string frequency of a particular subject and auxiliary.[4] Given that pronominal subjects are much more frequent than lexical subjects (cf. DuBois, 1987), it does not come as a surprise that auxiliary contraction is largely restricted to pronominal subjects and that the most frequent pronouns are the most common hosts of a clitic. In order to partial out the influence of phonetic factors on contraction, Krug (2003) conducted a study in which he concentrated on the auxiliary *have* after a set of pronominal subjects that all end in a vowel (e.g. *I've, you've, we've, they've, who've*). In accordance with the above hypothesis, he found a close correlation between the frequency of the subjects preceding *have* and the percentage of contractions, supporting his claim that string frequency influences phonetic reduction.

A related phenomenon has been investigated by Bybee and Scheibman (1999). Using transcripts of naturally occurring conversations, Bybee and Scheibman examined the various pronunciations of *don't*, which they grouped into four pronunciation classes: (1) tokens with an initial d-consonant and a full vowel [dõt, dõn], (2) tokens with an initial flap and a full vowel [æõt, æõ], (3) tokens with a flap and a reduced vowel [æə̃], and (4) tokens with just a reduced vowel [ə̃]. Table 4 shows that the reduction of *don't* varies with the words in the surrounding context.

[4]Interestingly, string frequency is a better measure in this case than conditional probability, suggesting that different measures are needed to account for different linguistic phenomena (cf. Krug, 2003).

Table 4
Don't variants by type of preceding and following item (Bybee & Scheibman, 1999, pp. 581–582)

	[dõt, dõ]	[rõt, rõ]	[rə̃]	[ə̃]	Total
Preceding					
I	16	22	38	12	88
you	7	7			14
we	2	6			8
they	1	3			4
Lexical NP	5				5
…					
Following					
know	2	8	24	5	39
think	7	6	6	1	20
have	1	7	1		9
have to	1	2	1		4
want	1	1	3		5
see	3	1			4
like		1		1	2
get	1	2			3
…					

The reduction process is most advanced with frequent context words. In particular, *don't* tends to be reduced after highly frequent pronominal subjects, notably after *I*, and before high frequency verbs, notably before *know* and *think*. Since there are no phonological features that could condition the reduction, Bybee and Scheibman conclude that it must be the frequent combination of *don't* with a particular subject and a particular verb that conditions the reduction.

Moreover, they claim that high-frequency strings such as *I don't know* and *I don't think* have turned into processing units that are stored independently of their components. Although these expressions seem to abide by general grammatical rules, they have assumed a life of their own. As Bybee and Scheibman have pointed out, *I don't know* and *I don't think* are not only special in that they tend to occur with a reduced form of *don't*, they also have special meanings. While they are formally negated, they do not express negation; rather, *I don't know* serves either to express the speaker's uncertainty or to indicate polite disagreement, and *I don't think* expresses an epistemic stance towards the associated proposition.

Bybee and Scheibman argue that the status of *I don't know* and *I don't think* as storage and processing units is eventually motivated by their frequent occurrence in language use. More specifically, they claim that the frequency of these expressions leads to a process of automatization, which is a general psychological mechanism that does not only affect the use of language but also other skills such as music and sports (cf. Logan, 1988; Schneider and Shiffrin, 1977; Shiffrin and Schneider, 1977). Automatization describes the process whereby a sequence of elements (e.g. a sequence of words) is gradually transformed through repeated use to a single chunk or holistic unit. In the course of this process, the elements of the string may lose their independence, boundaries are blurred, and the whole chunk is compressed and reduced.

4. Language change

Small biases in language production can lead to diachronic change. At first, these biases are on-line variants of more elaborate forms, but if the variants occur over an extended period of time they are often conventionalized and separated from their (historical) source. This is a well-known process of diachronic change that has been studied extensively in recent work on grammaticalization.

Grammaticalization is commonly defined as the process whereby lexical expressions (i.e. nouns and verbs) and demonstratives develop into grammatical markers, which may continue to develop into more strongly grammaticalized expressions (cf. Diessel, 2006; Hopper & Traugott, 1993; Lehmann, 1995). Interestingly, grammaticalization processes tend to follow universal pathways that originate from the same source. Here are some well-known examples.

- Across languages future-tense auxiliaries develop from two common sources: motion verbs such as *come* and *go* (e.g. *gonna*) and verbs of intention (e.g. *will*). (cf. Bybee, Perkins, & Pagliuca, 1994).
- Modal verbs such as *can* and *must* are commonly derived from verbs expressing permission and obligation (cf. Krug, 2000).
- Adpositions are often based on body part terms such as *stomach* or *head* that are metaphorically used as spatial expressions (e.g. *back*) (cf. Heine, Claudi, & Hünnemeyer, 1991; Heine & Reh, 1984).
- Complementizers are frequently derived from demonstrative pronouns (e.g. *that*) or from verbs of saying (cf. Diessel, 1999; Heine & Reh, 1984).
- Definite articles are almost always derived from demonstratives (e.g. þæ*s* > *the*) (cf. Diessel, 2006; Greenberg 1978).
- Tense and aspect affixes emerge from independent auxiliaries (Bybee et al. 1994).
- And case affixes are commonly derived from adpositions (Lehmann 1995).

Grammaticalization is crucially motivated by semantic (or conceptual) factors. For instance, motion verbs are often reanalyzed as future tense auxiliaries because time can be conceptualized as a path (e.g. *go* > *gonna*). But in addition to the semantic factors, frequency plays an important role in the process of grammaticalization. As has been amply demonstrated in the literature, linguistic expressions that undergo grammaticalization tend to lose some of their phonetic substance and/or their pragmatic and semantic force (cf. Bybee, 2003; Traugott, 1989). Both processes are due to frequent language use: Frequently used expressions tend to be phonetically reduced because they are highly predictable in a particular context (see above), and they are often semantically/pragmatically reduced (or weakened) because repetition reduces the psychological effect of the stimulus.

Since the reduction effect of frequency is based on a general psychological mechanism, grammatical processes tend to proceed in only one direction, i.e. grammaticalization is unidirectional, leading from content words to function words that develop into affixes, via clitics, before they disappear (cf. 5) (cf. Hopper and Traugott, 1993).

(5) content word > function word > clitic > affix > zero

Specific instantiations of this cline characterize the development of particular types of grammatical markers. For instance, third person pronouns commonly derive from

anaphoric demonstrative pronouns and may evolve into pronominal clitics, which turn into agreement markers before they disappear (cf. 6) (cf. Diessel, 1999, pp. 119–120).

(6) anaphoric DEM > 3.person pronoun > pronominal clitic > agreement marker > zero

Since the development is driven by high frequency in language use, it cannot be converted. Moreover, we can predict from this development that grammatical markers tend to be short. There is a well-known correlation between the length and the categorical status of linguistic expressions, which is eventually motivated by frequency of use. As can be seen in Table 5, the 20 most frequent words of the British National Corpus include only function words consisting of maximally three phonemes. The correlation between frequency, length, and categorical status has become known as Zipf's (1935) law, which is perhaps the most powerful generalization about the relationship between language structure and language use.

Interestingly, frequency is not just the driving force of phonetic reduction and grammaticalization, it can also be a conservative force. As Bybee and Thompson (1997) have shown, frequently used expressions are often resistant to analogical change. For instance, in English there has been continuous pressure to regularize irregular verb forms. Since the time of Old English, nearly 200 verbs have lost the stem vowel alternation and have adopted the regular past tense form (cf. Table 6).

If we look at the verbs that are still irregular in Present Day English, we find that most of them are very frequent. The frequent use has strengthened their representation in memory, which is why they have resisted the pressure from analogical change.

Interestingly, the conserving effect of token frequency is not restricted to isolated words but can also affect syntactic patterns. For instance, Bybee and Thompson (1997) have argued that auxiliaries and modals are exempt from *do*-support because these verbs occurred so frequently in Late Middle English and Early Modern English that they were not affected by the new word order patterns that emerged at that time (cf. Krug, 2000, 2003). In earlier periods of English, all verbs were fronted in yes–no questions and negated without *do*; but at the end of the Middle English period speakers began to insert the verb *do* into questions and negative sentences. However, the change was gradual and affected individual verbs at separate times. At the beginning of the Early Modern English period most verbs had switched to the new word order patterns except for a few frequent verbs,

Table 5
Top frequent words in the BNC (frequency per million words in spoken speech) (Leech et al., 2001)

1.	the	DET	39605	11.	n't	PART	12212
2.	I	PRO	29448	12.	in	P	11609
3.	you	PRO	25957	13.	we	PRO	10448
4.	and	CONJ	25210	14.	is	AUX	10164
5.	it	PRO	24508	15.	do	AUX	9594
6.	a	DET	18637	16.	they	PRO	9333
7.	's	AUX	17677	17.	was	AUX	8097
8.	to	INF	14912	18.	yeah	INT	7890
9.	of	P	14550	19.	have	AUX	7488
10.	that	PRO	14252	20.	what	PRO	7313

Table 6
Examples of regularized English verb forms

	Old form	New form
climb	clomb	climbed
creep	crope	crept
laugh	low	laughed
yield	yold	yielded
step	stope	stepped

which according to Bybee and Thompson resisted the pressure to change because they were strongly represented in memory.

Krug (2000) provided empirical evidence for this hypothesis. Investigating data from the Helsinki Corpus (i.e. from the time between 1420 and 1500), he found that both word order patterns were commonly used in Late Middle English. However, if we look at the verbs that were still used without *do* at that time, we only find frequent verbs, including modals and auxiliaries but also other high-frequency verbs such as *know, come, speak,* and *dare*. Apart from the modals and auxiliaries, all of these verbs were eventually attracted by the new word order patterns, but note that *know,* which is one of the most frequent verbs, can still occur without *do* in Present Day English (e.g. *They know not what they do*).

In sum, token frequency has two seemingly contradictory effects on diachronic change. On one hand, token frequency leads to phonetic reduction and the development of new linguistic forms; but on the other hand, token frequency can be a conservative force protecting high-frequency structures from analogical leveling. The final section shows that the two-frequency effects of diachronic change have given rise to some striking cross-linguistic tendencies that linguistic typologists subsume under the notion of markedness.

5. Typological markedness

The term markedness has a wide range of uses in linguistics (cf. Haspelmath, 2006). In what follows I concentrate on the typological markedness patterns that were first discussed by Greenberg (1966), see also Croft (2003). Typological markedness refers to cross-linguistic asymmetries in the encoding of grammatical phenomena. Croft (2003) distinguishes two different types of typological markedness: the structural encoding and the behavioral potential. Both types of markedness can be seen as emergent grammatical phenomena that have been shaped over time by frequency of use.

The notion of structural markedness refers to asymmetries in the morphological encoding of grammatical features. The classical example of structural markedness is the encoding of nominal number (cf. Croft, 2003, pp. 88–89). Across languages, the plural of nouns is often expressed by an affix while the singular is structurally unmarked. English is a good example: Most English count nouns form the plural by adding the suffix *–s*. The same strategy is found in many languages across the world, but there are also other patterns (cf. Dryer, 2005). For instance, some languages do not have any morphological plural marking (e.g. Mandarin Chinese), and some languages employ an extra morpheme for both singular and plural nouns (e.g. Latvian). What does not seem to occur, or is at least extremely rare, is the occurrence of a singular morpheme in languages in which the plural is morphologically unmarked. Thus, we may describe the cross-linguistic

markedness patterns in terms of an implicational universal: if singular nouns occur with an overt number marker, plural nouns are also marked for number. More generally, Croft (2003, p. 92) defines structural markedness as follows:

> The marked value of a grammatical category will be expressed by at least as many morphemes as is the unmarked value of that category.

Greenberg (1966) and Croft (2003) discuss a wide range of structural markedness patterns. For instance, if a language uses case affixes to mark grammatical relations, the subject is often structurally unmarked, i.e. the subject does not carry a case affix while all other grammatical relations are marked. Example (7) shows a transitive clause in Luiseno (Uto-Aztecan, North America) in which the object occurs with a case suffix while the subject does not carry a case marker.

(7) ʔáşwut kasila-**y** toow-q
eagle lizard-OBJ see-SG
'The eagle sees the lizard.'

There are also languages in which neither the subject nor the object occur with a case marker (e.g. English) and languages in which both subject and object are structurally marked (e.g. Latin); but languages with a marked subject and an unmarked object are extremely rare, (though they exist, e.g. Mojave). Table 7 shows a cross-linguistically typical paradigm of nominal inflection, exemplified by a Turkish noun, in which the singular subject does not take any morphological marker, while all other singular nouns carry a case suffix, and plural nouns take both a case suffix and a plural marker except for the plural subject which only occurs with a number marker.

Like nominal inflection, verbal inflection is often asymmetrical such that verbs in the present tense are marked by fewer morphemes than verbs of other tense and aspects categories. This is illustrated in Table 8, which presents the various tense and aspect forms of the Latin verb *laudāre* 'to praise'. As can be seen in this table, only in the present tense is the verb unmarked.

How do we explain the cross-linguistic asymmetries in the encoding of inflectional categories? Greenberg (1966, pp. 65–69), argued that structural markedness patterns reflect different frequencies in language use. Specifically, he claimed that the most frequent inflectional categories tend to be unmarked. There are two reasons why one would expect that structural markedness patterns correlate with frequency. First, since frequently used expressions tend to undergo reduction, one might hypothesize that frequent inflectional

Table 7
Case and number marking in Turkish

	Singular	Plural
Nominative	adam	adam-lar
Accusative	adam-K	adam-lar-K
Genitive	adam-K*n*	adam-lar-K*n*
Dative	adam-a	adam-lar-a
Locative	adam-da	adam-lar-da
Ablative	adam-dan	adam-lar-dan

Table 8
Tense and aspect inflection in Latin

Present	laudō	'I praise'
Past	laudā-bam	'I praised'
Future I	laudā-bō	'I will praise'
Perfect	laudāv-Ī	'I have praised'
Past Perfect	laudāv-eram	'I had praised'
Future II	laudāv-erō	'I will have praised'

categories are often unmarked because they have been phonetically reduced (see above). Second, since frequently used categories have a high (prior) probability they often act as the default, which does not need explicit coding. In fact, encoding the default, i.e. the expected value, would be redundant, i.e. it would violate a basic maxim of conversation (cf. Croft, 2003).

Additional support for the frequency account of structural markedness comes from a phenomenon called "local markedness" (Tiersma, 1982). Local markedness refers to the reversal of markedness patterns in individual words. For instance, collective nouns, denoting entities that typically occur in pairs or groups are often structurally marked in the singular. This can be a very common pattern, as for instance in Turkana (Nilotic, Africa), in which a wide range of nouns for animals, people, small things, food, and certain body parts carry an extra morpheme in the singular while the plural is unmarked (cf. Croft, 2003; Dimmendaal, 1983). For instance, the noun for 'shoe' occurs with the singular suffix *–àt* while the plural 'shoes' does not take a number marker (cf. 8).

(8) Îa-mukl 'shoes'
a-muk-**àt** 'shoe'

English does not have a singular marker, but note that English has a few count nouns for animals that commonly refer to groups and do not carry a plural marker: *fish, deer, sheep, moose, salmon* (cf. Croft, 2003; Tiersma, 1982).

Tiersma (1982) has demonstrated that local markedness patterns can also be found in analogical change. He presents data from Frisian in which certain nouns with alternating stem vowels in the singular and plural have undergone analogical leveling. In most nouns, the vowel of the plural has changed to the vowel of the singular; but in some cases, nouns referring to entities that are commonly experienced in pairs or groups have changed in the opposite direction, i.e. they have adopted the vowel of the plural in the singular (cf. Table 9).

Tiersma argues that the existence of local markedness patterns and local analogical changes are motivated by semantic factors and driven by frequency of use. Markedness reversals occur when the marked value of a morphological pattern is more frequent than the unmarked value. For instance, while the singular is overall more frequent than the plural, in collective nouns it is often the other way around (for some statistical support for this hypothesis see Greenberg, 1966 and Tiersma, 1982).

Like structural markedness, behavioral markedness reflects the effect of frequency in language use. Croft distinguishes between two subtypes of behavioral markedness: the

Table 9
Local analogical changes (Tiersma, 1982)

	Old forms Singular	Plural	Levelled forms Singular	Plural
Dominant pattern				
'coal'	koal	kwallen	koal	koalen
'whore'	hoer	hworren	hoer	hoeren
'meal'	miel	mjillen	miel	mielen
Local pattern				
'arm'	earm	jermen	jerm	jermen
'goose'	goes	gwozzen	gwos	gwozzen
'tooth'	kies	kjizzen	kjizze	kjizzen

Table 10
Inflection of *be*

	Present	Past
1st SG	am	was
2nd SG	are	were
3rd SG	is	was
1st PL	are	were
2nd PL	are	were
3rd PL	are	were

inflectional and distributional potential. Here, I concentrate on the inflectional potential, which Croft (2003, p. 97) defines as follows:

> If the marked value has a certain number of formal distinctions in an inflectional paradigm, then the unmarked value will have at least as many formal distinctions in the same paradigm.

Consider for instance the inflection of the English auxiliary *be* in Table 10.

There are two asymmetries in this table: first, *be* has more distinct forms in the present tense than in the past tense: there are three separate present tense forms, *am, are* and *is*, and only two past tense forms, *was* and *were*. Second, *be* has more distinct forms in the singular than in the plural; in present tense, there are three separate singular forms (*am, are, is*) and only one form in the plural (*are*), and in past tense, there are two separate singular forms (*was, were*) and only one form in the plural (*were*). Thus, we can conclude that the present tense and the singular are the unmarked values of the categories tense and verbal number. This is not just a particular property of the English verb *be* but a general typological pattern. Across languages, verbs tend to differentiate more forms in the present tense than in the past tense and more forms in the singular than in the plural (cf. Greenberg, 1966).

Similar asymmetries can be found in the inflectional paradigms of nominal expressions. Consider for instance the data in Table 11.

As can be seen in this table, there is a striking difference in inflection between pronouns and nouns in English. Pronouns are inflected for three categories—number, gender, and

Table 11
Inflection of nominal expressions

		Pronouns		Nouns	
		Singular	Plural	Singular	Plural
Subject	Masculine	he	they	tree	trees
	Feminine	she	they		
	Neuter	it	they		
Object	Masculine	him	them		
	Feminine	her	them		
	Neuter	it	them		

case—whereas nouns are only inflected for number. Moreover, while the singular pronouns distinguishes three different genders—masculine, feminine, and neuter—the plural pronouns are not inflected for gender. Again, these are not idiosyncratic properties of nominal inflection in English. Across languages, third-person pronouns tend to have more distinctions in the inflectional paradigm than lexical nouns, and singular (pro)nouns tend to express more inflectional categories than the corresponding plural forms (cf. Greenberg, 1966).

Like structural markedness, behavioral markedness is motivated by frequency of use; but in this case it is the conserving effect that accounts for the markedness pattern. Since frequent forms are more strongly represented in memory than infrequent forms, they are more easily memorized. This is why irregular forms tend to be frequent. Infrequent irregular forms are usually regularized because they are difficult to memorize. This also explains why suppletion, i.e. the most extreme form of formal irregularity (e.g. *go–went*), only occurs with very frequent expressions (cf. Bybee, 1985).

In sum, the cross-linguistic markedness patterns described by Greenberg and others arise from frequency effects in diachronic change. More precisely, structural markedness is shaped by the reduction effect of frequency (together with the tendency to leave the default unmarked; see above), and inflectional markedness is due to the conserving effect of frequency that protects frequently used expressions from leveling.

6. Conclusion

This paper has shown that frequency of occurrence affects the processes of language acquisition, sentence comprehension and production, and diachronic change. The various frequency effects are based on psychological mechanisms that may be divided into three basic types:

- First, the strengthening of linguistic representations. Frequency of use reinforces the representation of linguistic expressions in memory, which in turn influences their activation and interpretation in language use.
- Second, the strengthening of linguistic expectations. Since linguistic expressions are arranged in recurrent orders, the language user develops expectations as to which linguistic expressions may occur after a particular word or a particular category, which

influences the comprehension and production of linguistic units and can give rise to diachronic change.

- Third, the development of automatized chunks. Linguistic expressions that are frequently combined may become automatized, i.e. they may develop into a processing unit in which the boundaries between linguistic elements are blurred and the whole chunk is compressed and reduced.

The influence of frequency on linguistic structure challenges the rigid division between grammar and language use; it suggests a dynamical model of grammar in which linguistic structure is grounded in language use. Frequency of occurrence plays an important role in this model, but let me emphasize that frequency is not the sole factor affecting the emergence of linguistic structure. There are other cognitive or psychological factors that influence the use and development of grammar. These factors may be divided into two basic types.

First, the emergence of linguistic structure is crucially influenced by analogy and related phenomena such as metaphor (cf. Itkonen, 2005). Analogy is a general psychological mechanism that plays an important role in both language acquisition and diachronic change (cf. Gentner, Holyoak, & Kokinov, 2001; Hock, 2003). In language acquisition, analogy accounts for the emergence of abstract linguistic representations (cf. Rattermann & Gentner, 1998), and in language change, analogy accounts for the extension of established patterns to novel expressions and the leveling of irregular forms (cf. Hock, 2003).

Second, the emergence of linguistic knowledge is influenced by particular communicative and cognitive pressures. For instance, it has been repeatedly argued that the grammatical category of subject is shaped (via the pragmatic category of topic) by information processing (cf. Givón, 1976; Shibatani, 1991), and that the cross-linguistic word order correlations described by Greenberg and others are motivated by competing pressures from syntactic parsing, information processing, and semantics (Hawkins, 1994, 2004; see also Diessel, 2001, 2005).

In general, grammar is an emergent phenomenon that is fundamentally grounded in language use. One factor that has an impact on the emergence of linguistic structure is frequency of occurrence, which is the driving force of several psychological mechanisms that are involved in using language. The frequency-based mechanisms interact with other psychological mechanisms such as analogy and information processing, which together shape linguistic structure in the process of language use.

References

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

Bod, R. (2003). *Beyond grammar. An experienced-based theory of language*. Stanford: CSLI Publications.

Bod, R., Hay, J., & Jannedy, S. (Eds.). (2003). *Probabilistic linguistics*. Cambridge, MA: MIT Press.

Bybee, J. (1985). *Morphology*. Amsterdam: John Benjamins.

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

Bybee, J. (2001). *Phonology and language use*. Cambridge, MA: Cambridge University Press.

Bybee, J. (2003). Mechanisms of change in grammaticization: the role of frequency. In B. D. Joseph, & R. D. Janda (Eds.), *The handbook of historical linguistics* (pp. 602–623). Oxford: Blackwell.

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

Bybee, J., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar. Tense, aspect, and modality in the languages of the world*. Chicago: The University of Chicago Press.

Bybee, J., & Scheibman, J. (1999). The effect of usage on degrees of constituency: The reduction of *don't* in English. *Linguistics*, *37*, 575–596.

Bybee, J., & Thompson, S. (1997). Three frequency effects in syntax. *Berkeley Linguistics Society*, *23*, 378–388.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Crain, S., & Pietroski, P. (2001). Nature, nurture and universal grammar. *Linguistics and Philosophy*, *24*, 139–186.

Croft, W. (2003). *Typology and universals* (2nd ed.). Cambridge, MA: Cambridge University Press.

Diessel, H. (2006). Demonstratives, joint attention, and the evolution of grammar. *Cognitive Linguistics*, *17*, 463–489.

Diessel, H. (1999). *Demonstratives. Form, function, and grammaticalization*. Amsterdam: John Benjamins.

Diessel, H. (2001). The ordering distribution of main and adverbial clauses: A typological study. *Language*, *77*, 343–365.

Diessel, H. (2004). *The acquisition of complex sentences*. Cambridge, MA: Cambridge University Press.

Diessel, H. (2005). Competing motivations for the ordering of main and adverbial clauses. *Linguistics*, *43*, 449–470.

Diessel, H., & Tomasello, M. (2005). A new look at the acquisition of relative clauses. *Language*, *81*, 1–25.

Dimmendaal, G. J. (1983). *The Turkana language*. Dordrecht: Foris.

Dryer, M. (2005). Coding of nominal plurality. In M. Haspelmath, M. Dryer, D. Gil, & B. Comrie (Eds.), *World atlas of linguistic structures* (pp. 138–141). Oxford: Oxford University Press.

DuBois, J. A. (1987). The discourse basis of ergativity. *Language*, *64*, 805–855.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness. A connectionist perspective on development*. Cambridge, MA: MIT Press.

Elman, J. L., & McClelland, J. L. (1984). Speech perception as a cognitive process: The interactive model. In N. Lass (Ed.), *Speech and language* (pp. 360–380). San Diego: Academic Press.

Ford, M., Bresan, J., & Kaplan, R. M. (1982). A competence-based theory of syntactic change. In J. Bresan (Ed.), *Mental representations and grammatical relations* (pp. 727–796). Cambridge, MA: MIT Press.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *7*, 58–93.

Gentner, D., Holyoak, K. J., & Kokinov, B. N. (Eds.). (2001). *The analogical mind. Perspectives from cognitive science* (pp. 437–469). Cambridge, MA: MIT Press.

Givón, T. (1976). Topic, pronoun, and grammatical agreement. In C. Li (Ed.), *Subject and topic* (pp. 149–188). New York: Academic Press.

Greenberg, J. H. (1966). *Language universals. With special reference to feature hierarchies*. The Hague: Mouton.

Greenberg, J. H. (1978). How does language acquire gender markers? In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language*, vol. 3 (pp. 47–82). Stanford: Stanford University Press.

Gregory, M., Raymond, W. D., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistic Society*, *35*, 151–166.

Hare, M., McRae, K., & Elman, J. L. (2004). Admitting that admitting verb sense into corpus analyses makes sense. *Language and Cognitive Processes*, *19*, 181–224.

Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of Linguistics*, *41*, 1–46.

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge, MA: Cambridge University Press.

Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Heine, B., Claudi, U., & Hünnemeyer, F. (1991). *Grammaticalization: A conceptual framework*. Chicago: Chicago University Press.

Heine, B., & Reh, M. (1984). *Grammaticalization and reanalysis in African languages*. Hamburg: Buske.

Hock, H. H. (2003). Analogical change. In B. D. Joseph, & R. D. Janda (Eds.), *The handbook of historical linguistics* (pp. 425–440). Malden: Blackwell.

Hooper, J. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In W. Christie (Ed.), *Current progress in historical linguistics* (pp. 96–105). Amsterdam: North Holland.

Hopper, P. J. (1987). Emergent grammar. *Berkeley Linguistics Society*, *13*, 139–157.

Hopper, P., & Traugott, E. C. (1993). *Grammaticalization*. Cambridge, MA: Cambridge University Press.

Itkonen, E. (2005). *Analogy as structure and process*. Amsterdam: John Benjamins.

Juliano, C. & Tanenhaus, M.K. (1993). Contingent frequency effects in syntactic ambiguity resolution. In *Proceedings of the 15th annual conference of the cognitive science society*, *COGSCI-94*, Boulder (pp. 593–598).

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

Jurafsky, D. (2003). *Probabilistic modelling in psycholinguistics: Linguistic comprehension and production*. In Bod et al. (Eds.), (pp. 39–95).

Jurafsky, D., Bell, A., Gregory, G., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee, & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–253). Amsterdam: John Benjamins.

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

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

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignment. *Psychological Review*, *99*, 349–364.

Krug, M. (1998). String frequency. A cognitive motivating factor in coalescence, language processing, and linguistic change. *Journal of English Linguistics*, *26*, 286–320.

Krug, M. (2000). *Emerging modals. A corpus-based study of grammaticalization*. Berlin: Mouton de Gruyter.

Krug, M. (2003). Frequency as a determinant of grammatical variation and change. In G. Rohdenburg, & B. Mondorf (Eds.), *Determinants of grammatical variation in English* (pp. 7–67). Berlin: Mouton de Gruyter.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English based on the British national corpus*. Harlow: Longman.

Lehmann, C. (1995). *Thoughts on grammaticalization*. Munich: Lincom Europa.

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

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, *191*, 676–703.

MacWhinney, B. (2000) (3rd ed.). *The CHILDES Project. Tools for analyzing talk, vol. 2, the database*. Hillsdale, NJ: Lawrence Erlbaum.

Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in the speech to young children. *Cognitive Science*, *26*, 393–424.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorization. *Cognition*, *96*, 143–182.

Pollack, I., & Pickett, J. M. (1964). Intelligibility of excerpts from fluent speech: Auditory vs. structural context. *Journal of Verbal Learning and Verbal Behavior*, *3*, 79–84.

Rattermann, M. J., & Gentner, D. (1998). More evidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. *Cognitive Development*, *13*, 453–478.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, *22*, 435–469.

Roland, D., Elman, J., & Ferreira, V. S. (2006). Why is *that*? Structural prediction and ambiguity resolution in a very large corpus of English sentences. *Cognition*, *98*, 245–272.

Roland, D., & Jurafsky, D. (2002). Verb sense and verb subcategorization probabilities. In P. Merlo, & S. Stevenson (Eds.), *Sentence processing and the lexicon: Formal, computational, and experimental perspectives* (pp. 325–346). Amsterdam: John Benjamins.

Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition*, *81*, 149–169.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-months-old infants. *Science*, *274*, 1926–1928.

Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition*, *69*, 105–134.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1–66.

Shibatani, M. (1991). Grammaticalization of topic into subject. In Traugott, & Heine (Eds.), *Approaches to grammaticalization* (pp. 93–134). Amsterdam: John Benjamins.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127–190.

Spivey-Knowlton, M. J., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55*, 227–267.

Tabor, W., Juliano, C., & Tanehnaus, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes*, *12*, 211–271.

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.

Tiersma, P. (1982). Local and general markedness. *Language*, *59*, 832–849.

Traugott, E. C. (1989). On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language*, *65*, 31–55.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology*, *36*, 417–427.

Wiechmann, D. (2006). Sense-contingent lexical preferences and early parsing decisions: Corpus-evidence from local NP/S-ambiguities. Manuscript, University of Jena.

Zipf, G. (1935). *The psycho-biology of language*. Boston: Houghton Mifflin.