CHAPTER 4

Sequentiality as the basis of constituent structure

Joan Bybee
University of New Mexico

1. Sequentiality and constituent structure

Recent investigations of the relation between grammar and language use, especially in the literature on grammaticization (Bybee *et al.* 1994, Hopper and Traugott 1993, among many others), have made it clear that grammar is not a fixed logical structure, but rather is constantly changing and evolving. Bybee (1998b) and Haiman (1994) have argued that the basic mechanisms of the grammaticization process are neither domain-specific nor species-specific. Thus the hope has arisen that the highly abstract and heretofore mysterious properties of grammar might be explainable in more general terms. In my earlier work I have addressed the process by which grammatical morphemes develop, emphasizing the role that repetition plays in the process. The current paper addresses the mechanism behind the hierarchical arrangement of linguistic elements into constituents, once again emphasizing the role of language use and repetition.

The existence of constituent structure and the hierarchical organization resulting from it has always been taken by linguists as prime evidence that linguistic behavior does not consist merely of linear strings of elements. It is further believed that the hierarchical organization of sentences is one of the most basic aspects of language, indeed, a defining feature of human language.

Linguists rarely ask why natural language has constituent structure;[1] they merely assume that it does, just as they assume that all phonologies will be organized into segments and features. In the spirit of Lindblom *et al.* (1984), I submit that structure can be explained, that form is emergent from substance, and that the larger design of language can be interpreted as the indirect consequence of local behavior (Lindblom *et al.* 1984: 186). My goal in this paper is to present and defend the hypothesis that sequentiality is basic to language and constituent structure emerges from sequentiality because elements that are frequently used together bind together into constituents.

This point is likely to seem rather obvious to some readers, and indeed follows directly from comments in Givón (1998, in this volume). To my knowledge, however, it has not been directly argued for or demonstrated. In fact, some linguists seem to be convinced that constituent structure is basic to language, and therefore might find the hypothesis surprising. Thus, towards the ultimate goal of identifying the mechanisms that create grammar and relating them as much as possible to domain-general cognitive processes, I undertake this demonstration.

2. Definition of constituent

A popular introductory textbook for linguistics, *Language Files*, has a nice discussion of the notion of constituent. It first begins with a discussion of linear order in sentences, and then points out that in addition to linear order, there is hierarchical organization. The book says that the semantically coherent groupings of words in sentences are constituents, thereby identifying both semantics and linear order as important to constituent structure. It goes on to give three tests for identifying constituents:

1. Constituents can sensibly be used alone, as in the answers to questions. (*Who chewed up your shoe? My new puppy.*)
2. Constituents can be replaced by pro-forms. (E.g. *I saw him do it* where *him* replaces *my new puppy* and *do it* replaces *chew up my shoe.*)
3. Constituents can occur in various places in the sentence, e.g. an NP can be the subject of the verb or object of the verb or a preposition; an NP can be extraposed (e.g. *My new puppy, he chewed up my shoe.*)[2]

3. Explanations for constituent structure

Of course various explanations for why language is organized into constituents are possible. One could claim that specific types of phrases, such as NP, VP and PP are innate and only need the presence of minimal input to trigger their language-specific organization. Or one could propose, as does Langacker (1987: 310), that 'hierarchy is fundamental to human cognition' making it a general cognitive attribute that can be applied to language. Certainly I agree with this view. However, we must also ask just what is organized hierarchically and how languages come to be organized in the particular way that they are.

Langacker's (1987, 1997) particular proposal is that constituency reflects semantic relations. Again, it is easy to agree with this position since many linguists have pointed out the iconic relation between the conventionalized structures of languages and their semantic relevance (e.g. Bybee 1985). However, as we will see below, constituency and meaning are not always in a strictly iconic relationship and the human mind can apparently find constituents in re-occurring meaningless syllables (Saffran *et al.* 1996). Thus we search for an additional factor to fully explain the existence of constituent structure.

My hypothesis is that semantics, and to some extent, pragmatics and our experience with the world, will determine what elements tend to occur together in sequences in an utterance, but repetition is the glue that binds constituents together. Thus I hypothesize that hierarchies of constituent structure are derivable from frequent sequential co-occurrence. In this view, the more often particular elements occur together, the tighter the constituent structure. Thus low-level constituents such as a determiner, *the*, and a noun, such as *puppy*, frequently co-occur, while higher-level constituents, such as an NP, *the puppy*, and verbs such as *ran, licked,* or *slept* occur together less often. Note that in this view constituent structure can be gradient and two constituents which seem to have the same structure may have different degrees of cohesion due to the differences in their co-occurrence patterns (Bybee and Scheibman 1999).

4. Knowledge of grammar is procedural knowledge

To understand the role that frequency or repetition plays in the creation of grammar it is important to recognize that language production is a neuromotor behavior based on procedural knowledge rather than propositional knowledge. Propositional knowledge is 'knowing that' or knowing facts such as 'Santa Fe is the capital of New Mexico'. Procedural knowledge is 'knowing how' and includes knowing how to tie shoelaces or how to drive a car. Propositional knowledge is conscious knowledge which is easy to report on. Procedural knowledge is usually below the level of conscious awareness and while subjects can carry out the procedures, it is much more difficult for them to report what the procedure is. This distinction has an interesting parallel in the difference between lexical and grammatical knowledge. While speakers are often able to report on the meanings of words or phrases, it is much more difficult for untrained speakers to explain the meanings of grammatical morphemes or grammatical constructions. Thus we might conclude that lexical items involve at least some propositional knowledge, while grammatical constructions are largely procedural.

This conclusion is further supported by the manner in which procedural knowledge develops, as outlined by Anderson (1993) and Boyland (1996). Two properties of the development of procedural knowledge are important for our understanding of the way grammar develops. First, frequently used actions become fluent more quickly; that is, repetition increases fluency. For the purposes of language, our common-sense notion of fluency can be applied here: in fluent speech, words are strung together without inappropriate pauses. We can also go beyond this sense of fluency and note that with high levels of repetition, articulatory gestures can overlap one another and individual gestures can be reduced both in duration and in displacement. Thus grammaticizing constructions that undergo extreme frequency increases also undergo extreme phonological fusion and reduction, as for example, when *going to* reduces to *gonna* (see Bybee, Perkins and Pagliuca 1994; Bybee 2001). The second property is in a sense the mechanism that makes the first one possible: 'recurring sequences of actions come to be represented at a higher level as single actions, thus increasing fluency' (Boyland 1996: 10). Thus repeated sequences become fluent because they become automated into a single chunk that can be accessed and executed as a unit.

5. Linguistic evidence for chunking

Several types of evidence for the chunking of high frequency sequences can be cited. First, one unit of a chunk primes or automates the other unit. In the plaza at Old Town in Albuquerque I watched a boy of about five years spot an antique cannon. He said 'hey, Dad, there's a cannonball. Can I climb on the cannonball?' The father responded, 'that's not a cannonball, it's a cannon.' The boy insisted, 'Dad, can I climb on the cannonball' and the exchange repeated itself. The boy had learned *cannon* only in the context of the compound *cannonball* and that was the only context in which he could access the word. Hearers have automated chunks as well, with analogous priming effects. In the US, upon hearing *supreme*, one can expect *court* as the next word; or upon hearing *sesame* one can expect *street*.

Second, inside frequently used chunks, internal structure tends to be lost. Thus *gonna* no longer consists of the three morphemes *go*, *ing* and *to*. Third, the morphemes or words inside a chunk become autonomous from other instances. For example, speakers probably do not associate *go* in *gonna* with the lexical movement verb anymore. Sosa and MacFarlane (to appear) show that subjects have difficulty identifying the word *of* when it occurs in frequent chunks such as *sort of* or *kind of*. And fourth, the components of a chunk become compressed and reduced phonologically, as illustrated by *gonna*, but also by other verbs that are frequently followed by *to*, such as *wanna*, *hafta*, *gotta*, and so on.

The hypothesis of this paper is: **Items that are used together fuse together.**

This could be called the Linear Fusion Hypothesis. The explanation for this phenomenon is two-fold: first (and in my mind, foremost) is the automation of production that is typical for procedures; second, the fact that items are predictable in context allows the speaker to unleash the reductive processes that increase fluency.

6. Further examples of the fusion of high frequency combinations

Boyland (1996) argues that the sequence *would have* fuses due to the high frequency with which the two units occur together. The misanalysis apparent in the non-standard spelling *would of* shows that the identity of the two units has been lost. Also, the context after a modal is the only place that *have* reduces down to [ə]. In the example of *be going to* the most fusion is among the parts that are invariant and therefore occur together more often, that is, the *going to* part. The *be*-verb part, which varies among *am, is, are, was* and *were*, reduces and fuses only in the high frequency first person singular: *I'm going to* becomes not just *I'm gonna* but also [aimə̃ɾə̃] in casual speech. Krug (1998) studies the rate of contraction of English auxiliaries with the subject and finds that the more frequent a combination, the more likely it is to show contraction. For instance, the most frequently occurring contraction is *I'm* and that combination is the most frequent in the corpus. The basis for the high frequency of *I'm* is the strong subjectivity of speech, which makes first person singular the most frequent pronoun and the fact that *am* cannot occur with any other subject (Scheibman 2002). Bybee and Scheibman (1999) study the reduction of *don't* in American English and find that it reduces most in the contexts in which it is used the most, i.e. after *I* and before certain verbs such as *know, think, mean, care, feel*, etc.

While it is not necessarily true in the cases just cited, it is usually the case that high frequency, grammaticizing items that are used together are in the same constituent in the traditional sense. For example, a determiner will tend to show reduction and fusion that depends upon the noun with which it is in construction. Thus the English article *the*, has two pronunciations, [ði] and [ðə] as does *an*/*a*. The second variant in each case developed as a reduction before the consonant in the noun or adjective of the same NP. The French masculine definite article *le* loses its vowel before a vowel-initial noun: *le* + *ami* > *l'ami* 'the friend', etc.

Postpositions fuse with the nouns they operate on to become case suffixes. Thus the Turkish word *değin* meaning 'as far as' has fused with the preceding noun to make a new case suffix, as in *köydek* 'as far as the village' (Kahr 1976). Auxiliaries that are part of tense and aspect constructions and that follow the verb fuse with the verb to become suffixes, as in the famous case in the Romance languages, where the infinitive + *habere* constructions yields the future tense: Old Spanish *cantar ha* becomes Modern Spanish *cantará* 'he will sing'. Examples such as these are abundant in the grammaticization literature.

7. Constituency with less frequent combinations

The examples just cited are the extreme cases, since they involve combinations of extremely high frequency. But more subtle effects can also be found in cases of co-occurrence that are less frequent, leading me to hypothesize that chunking and constituency relate directly to frequency of co-occurrence.

Gregory *et al.* (1999) examine three measures of shortening of content words in a large corpus of conversation (the Switchboard corpus, see below). They find that the tapping of a word-final /t/ or /d/ is highly affected by the probabilistic variable 'mutual information', which is a measure of the likelihood that two words will occur together. If the final /t/ or /d/ occurs before a word that is highly likely given the first word, the rate of tapping increases. For instance, tapping would be more common in the word pair *lot of* than in *out alone.* They write 'The effect of mutual information on tapping suggests that tapping is a process that may also preferentially apply internally to highly cohesive pairs' (Gregory *et al.* 1999:9). Word-final deletion of /t/ and /d/ and the duration of a word are also highly affected by mutual information as well as other measures of probability. Jurafsky *et al.* 2001 find similar effects of probability on function words.

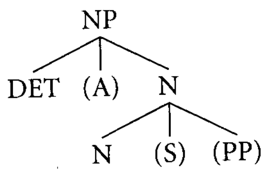
The usual interpretation of small phonological adjustments that occur between words is that they are due to on-line phonetic processes that apply as the words are strung together. However, if this is the case, it is difficult to explain why frequency of co-occurrence affects the probability that a phonological adjustment would take place. Part of the explanation is that the speaker, knowing that the next word has a high probability of occurring given the current word, allows reduction processes to occur. But for the speaker to know the probability of the occurrence of the next word, s/he must have a mental representation that includes knowledge about what words have occurred together in the past. Thus the occurrence of two words together strengthens the sequential link between them. It is this sequential link that is the local basis for the eventual emergence of constituent structure. That is, pairs of words with strong sequential links are regarded as occurring in the same constituent.

8. Sequentiality and constituency in the NP

To test the hypothesis that traditional notions of constituent structure correspond rather directly to frequency of co-occurrence, I elected to examine the English noun phrase in conversation. The noun phrase (especially one with a lexical noun) is a very good example of a constituent in English, as it has the same structure whether it is the subject or object of the verb, or object of the preposition. Further, it is sometimes used independently in conversation (Ono and Thompson 1994).[3]

The basic structure of an English NP was represented with the phrase structure tree in (1) in early models of generative grammar.

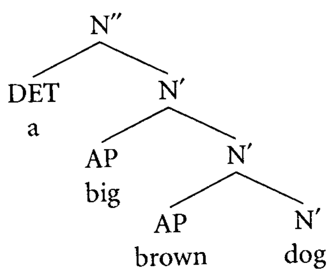
(1) NP structure:



The DET (determiner) and N are obligatorily present (although the DET can be a zero) and the other items are optional. The NP can contain an adjective which precedes the N, or a relative clause (S), or prepositional phrase (PP), which follows.

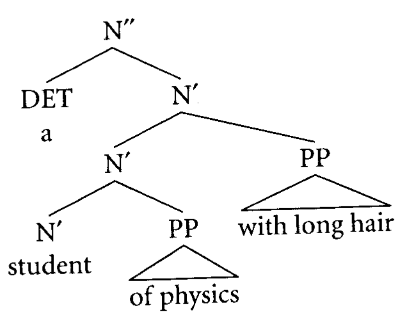
Another more recent representation of the NP in X-bar notation, provides an extra layer of constituency for each element added to the NP (see Radford 1988). Thus (2) shows an NP with two adjectives and (3) an NP with post-modifiers.

(2)



What predictions are made about levels of constituency by these phrase structure trees? In (1) the DET (A) N are all under the same node, NP, so they are will not be separated by any constituent boundaries except those that indicate that they are separate words. In (2) each adjective increases the distance between the determiner and the noun.

(3)



With post-modifiers in (1) or (3), there are two consequences for the constituent structure in the normal interpretation: first, the N will be separated from the DET (A) by an extra bracket, suggesting, perhaps erroneously, that an N with post-modifiers is not in the same constituent with the DET as the N is when it is final to the NP. Second, the post-modifier, being a constituent itself (as an S or PP) is separated from the N by several brackets, indicating a looser constituent structure.

While the second consequence seems fine, the first presents an odd prediction that a DET N combination has tighter constituent structure when no modifiers follow it than when they do. However, other mismatches between the bracketing generated by phrase structure rules and the surface constituency have been identified in the literature and a special component of rules, Readjustment Rules, were set up to fix these mismatches (Chomsky and Halle 1968). In addition, phonological phrasing, or the domain of phonological rules is sometimes at odds with the generated phrase structure and needs adjusting. Thus it is generally clear that the structures predicted by trees such as (1), (2) and (3) cannot be taken in a completely literal way. However, certain elements of the general structure presented in these trees are well-motivated, i.e. that the NP is a constituent, and that these elements, DET, pre-modifying adjectives and post-modifying PPs and Ss belong in the NP constituent. Thus I will compare this general structure to the quantitative facts of co-occurrence to determine whether or not generalizations about constituent structure can be said to be derived from multiple patterns of language use.

My hypothesis is that specific items that are used together frequently will form tighter bonds than items that occur together less often. The prediction is that items that occur within a traditional constituent are more likely to occur together in running discourse than are items that are in different constituents. One complication is the existence of optional items in a constituent. Strictly speaking, my hypothesis would predict that optional elements have weaker sequential links than obligatory elements, and thus looser constituency. While this prediction may be correct, I will not pursue it further here. Instead, I will focus on testing, in a general way, the predictions about co-occurrence made by the NP constituent. The question will simply be whether or not nouns occur more frequently with items inside or outside the NP.

To study the English NP, I selected eleven lexical nouns from the most frequently occurring nouns in the Switchboard corpus. This corpus consists of over 2.4 million words from recorded telephone conversations. The nouns used were: *husband, mother, computer, movie, school, car, house, money, idea, class* and *problem.*[4] These nouns were studied only in their singular forms. The goal of the study was to discover which items most frequently preceded and most frequently followed these nouns in the conversations and to compare the frequency of preceding vs. following items to see if the quantitative distributional facts correspond to traditional notions of constituency in the NP.

First consider where pauses occur in relation to these eleven nouns. Table 1 shows that the probability of a pause following the noun is much greater than that of a pause preceding the noun. In fact, more than one-third of the tokens of these nouns were followed by a pause, while fewer than one percent were preceded by a pause.[5] This suggests a much weaker constituency bond to the item following the noun, either because it is in another constituent or because it is a more loosely joined part of the NP. The reason for the low probability of a pause before the noun is presumably the obligatory presence of a determiner before the noun.

Table 1. Occurrence of pauses before and after eleven nouns from the Switchboard corpus (N=7870)

Before the noun	After the noun
74 (1%)	2696 (34%)

Table 2 shows the three most common linguistic items to precede and follow the noun. What is notable here is the greater predictability of the item preceding the noun compared to the one following it. In the Preceding column, *the* accounts for 17% of all tokens preceding the noun. Following the noun, *and* is the most common unit, but it only occurs after 7% of the tokens. Thus the co-occurrence patterns for X+N are stronger than for N+X, reflecting tighter constituency.[6]

Table 2. The three most frequent linguistic items preceding and following the eleven nouns

Preceding			Following		
the	1,348	17%	*and*	562	7%
my	958	12%	*that*	345	4%
a	767	10%	*is*	270	3%

In terms of understanding the constituent structure of the NP, it is instructive to view some of the most frequent items to precede and follow these eleven nouns. The items in Table 3 were selected by first counting the ten most frequent items to precede each noun, and then by selecting all the items that occurred in more than one list. Table 4 was constructed in the same way with attention to the items that followed the noun. Tables 3 and 4 are organized by grammatical function. The totals in the last column are followed by asterisks if they present apparent counter-examples to the hypothesis that elements following or preceding the noun are in the same constituent. These cases are discussed below.

Tables 3 and 4 confirm that the constituency of the NP as traditionally represented corresponds closely to the sequential co-occurrence patterns found in continuous speech. A large majority of the preceding and following items are part of the NP.

The position preceding the noun most frequently has a member of the class of determiners, including the high frequency articles, possessive pronouns (especially *my*), and quantifiers or demonstratives (a total of 47%). A few high frequency adjectives represent the adjective position in the NP. Finally, some items that are not in the NP made the list: the three prepositions *of*, *in* and *to* and the conjunction *and* with only six occurrences. A closer examination of the preposition + noun examples reveals that they occurred in particular phrases. *Of* occurred primarily with the mass noun *money* and was part of a quantifier phrase, e.g. *lot of* occurred 77 times, *amount(s) of* occurred 40 times, etc. Sinclair (1991:85) argues that in such phrases we do not have a head noun (*lot*) followed by a PP, but rather the second noun is the head of the phrase, and *lot of* or comparable phrases are modifiers to that noun. Under that analysis, which seems realistic for this data, *of* is not really a preposition and the phrase *lot of* is in the same constituent as the noun it modifies.

Table 3. Sets of items preceding the eleven nouns studied

Articles	*the*	134	3,703
	a/an	767	
Possessives	*my*	958	
	your	147	
	our	127	
	his	48	
	their	31	
	's	19	
Other determiners	*that*	137	
	any	53	
	no	76	
	one	11	
Prepositions	*of*	260	592*
	to	178	
	in	154	
Adjectives	*good*	152	248
	new	80	
	whole	16	
Conjunction	*and*	6	6

Table 4. Sets of items following the eleven nouns studied

Prepositions	*to*	120	445
	for	117	
	in	112	
	of	81	
	at	15	
Verbs	*is*	270	444*
	was	79	
	does	32	
	has	32	
	had	31	
Relative clause	*that*	345	409
	I	64	
N/NP Conjunction	*and*	342	342
S/VP-Conjunction	*and*	286	286*
Possessive	*'s*	145	145

The preposition *to* occurs with *school* in 170 out of 178 examples and with *class* in the other eight examples. *In* occurs with *school* in 139 out of 154 cases and with *class* and *computer* (e.g. *computer class)* in the other 15 cases. The usage suggests that *to/in school* and *in class* are locative adverbial rather than phrases involving full NPs. After all, it is quite idiosyncratic that these nouns lack a determiner in these expressions.

Even given that the prepositions *in* and *to* pose counter-evidence to my hypothesis, 93% of the items preceding the nouns investigated are in the same NP.

The position following the noun has no one category or item that is as strong as the articles or possessive pronouns in the preceding position. Prepositions, verbs and relative clauses follow the noun with approximately equal frequency. Prepositions are expected in this position and are considered part of the NP. Items that initiate relative clauses are also expected in this position, i.e. *that. I* turns out to be common at the beginning of a relative clause (as in *the movie I saw, the class I'm taking*). While the relative clause is in the NP, it also begins a new clause, so that it has a weaker constituent bond with the N than any items that precede the N. The rate of occurrence among these high frequency classes–relative clauses, prepositional phrases and high frequency verbs–is about the same, even though the verbs are not in the same constituent with the NP. This distribution suggests that the constituency bond between the noun and the modifying elements following the noun is weaker than with items preceding the noun. This relation is represented by additional constituency boundaries, as the relative clause would have a new S constituent and the PP itself is a constituent.

The fact that the frequency with which the N is followed by a verb, the first element in the other major constituent of the sentence, is comparable to that of the relative clause and PP, which are part of the NP, seems problematic for the hypothesis that frequency of co-occurrence corresponds to constituency. In particular, the high frequency of *is* would seem to be counter-evidence to this hypothesis. However, *is* and in some cases *has* and *had* are subject to contraction with the subject. Such cases of fusion where frequently co-occurring items are not in the same traditional constituents are the evidence I will use in sections 12 and 13 to argue that sequential co-occurrence is more basic than constituency.

None of the descriptions of the NP consulted before examining the conversational data prepares us for the high frequency with which the conjunction *and* follows a noun. In Table 4 I report a rough breakdown into those instances of *and* that conjoin another noun and those that introduce another clause or verb phrase. This breakdown was based on the item following *and.* Clause-conjunction was assumed if the following item was a conjunction, *then* or *so,* or an adverb. When a pronoun followed that was in the nominative case, the type of conjunction was assumed to be N-conjunction if the preceding noun was animate (*husband and I*) but clause-conjunction if the preceding noun was inanimate (*computer and we*). A following verb indicated verb or VP conjunction. All other cases could be reasonably assumed to be noun-conjunction. In a small number of cases these assumptions may have led to an erroneous assignment, but this breakdown allows us to get an idea of how many of these *and*'s introduce a major constituent break.

Overall the data suggest the predicted correspondence between sequential co-occurrence and traditional notions of constituency within the NP. They predict a stronger bond with the item preceding the noun than with the item following it, and most of the items found to precede or follow the noun are in the same traditional constituent. The primary counter-examples involve preceding prepositions, which seem to be restricted to certain high frequency phrases, and common verbs or auxiliaries following the noun.

In the next two sections, I discuss the cognitive mechanisms that underlie this general correspondence, and then in the following sections I proceed to the argument that the hierarchical structure of language is derivable from the more basic sequential nature of language.

9. Fragments and networks

Language learning in a natural setting involves the storage and processing of repeated fragments of speech. Such fragments might include parts of words,

words, or multiple words. Since human beings are sensitive to recurring sequences of stimuli and record them in memory, they learn to recognize repeated sequences of speech or speech-like stimuli (Saffran *et al.* 1996; Gomez and Gerken 1999). Similarly the automated neuromotor sequences that correspond to these perceptual units are recorded in memory and strengthened with use. Since the number of fragments of speech that must be stored in memory for a mature user of a language is extremely large, a tight organization of these fragments is necessary. I suggest for multi-word sequences, as I have for morphologically complex words, that whole complex units may be stored in memory, but they are typically associated with other units that are similar phonologically and semantically (Bybee 1985, 1995, 1998a). Figure 1 shows a simplified, schematic representation of a possible organization for some of the NPs we discussed above. The connecting lines between words indicate a relation of phonological and semantic identity.[7] Figure 2 shows a set of relations among NPs centering on the determiners rather than the N. The networks in Figures 1 and 2 interlock.

In this model representation is affected by language use. Each token of use of a word or sequence of words strengthens its representation and makes it more easily accessed. In addition, each instance of use further automates and increases the fluency of the sequence, leading to fusion of the units. Thus, as in morphology, high frequency combinations have a stronger representation. For instance, *my mother* (which occurred 182 times in the Switchboard corpus) has a stronger representation than *her house* (which occurred 22 times) or *this hospital* (which

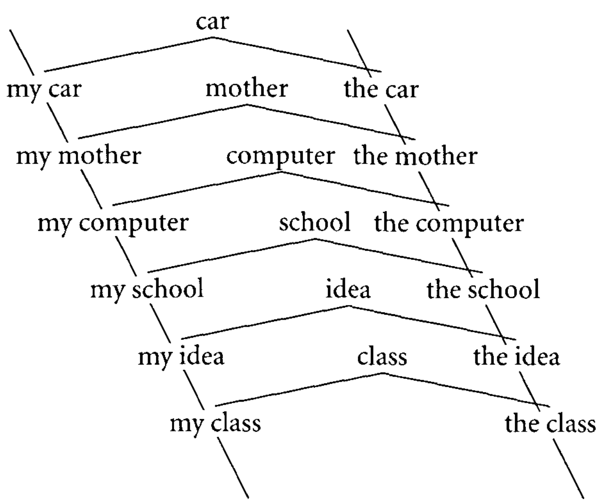


Figure 1. Possible organization of relations among NPs centering on frequently occurring nouns

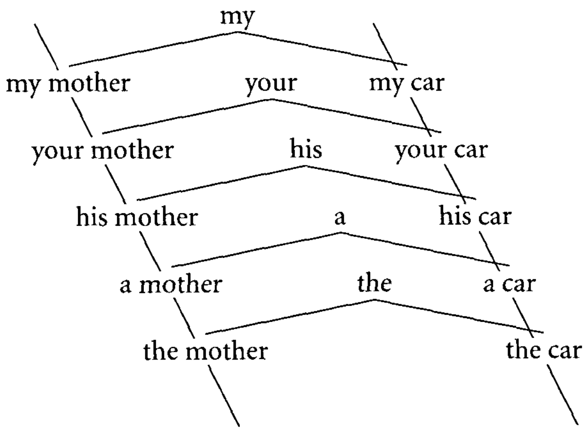


Figure 2. Possible organization of relations among NPs centering on dterminers

occurred only once). All of these NPs are interlocked in a network based on experience with language. From this network emerges the fact that any of these nouns, e.g., *mother, house* or *hospital* can occur with any of the set of determiners. Since particular combinations are also represented, we do not lose the information that some of these nouns occur more often with certain determiners (*my* with kinship terms, or *an* with *idea*) than others.

The hypothesis that very particular sequences of words from our experience are stored in lexical memory is supported by the findings of Erman and Warren 2000 that in the spoken and written texts they examined, 55% of the texts consisted of prefabricated units, that is, sequences of words that are conventionalized. Prefabricated sequences occur more frequently and have more idiomatically determined meaning than sequences that are put together afresh each time. In other words, prefabricated sequences have been experienced before. Pawley and Syder (1983) point out that native-like use of a language depends heavily upon knowing and using conventionalized word combinations.

10. Schemas emerge from the network

The organization apparent in Figures 1 and 2 is only possible if categories are established for strings or parts of strings. A relation between *my car* and *your car* is possible only if a category is established for *car*. Such a category is based on the similarity between tokens of use, which includes information not just about phonology and semantics, but also about contexts of use, both linguistic and non-linguistic. The linguistic context in which *car* is used will include information about what immediately precedes and follows it. Since, as we have seen, the preceding elements are more predictable, that is, the same ones occur more often, a category that includes the frequently occurring preceding element could be formed. The storage of multi-word strings and the categorization of their component elements leads to the formation of constructions. Constructions are schematic generalizations over sequences with shared parts. I agree with Langacker (1987) that schemas are formed at various levels of abstraction. For instance, for NP in English, we might find the following levels of abstraction:

1. Very specific: *my mother, my computer, the car, a problem, an idea*
2. Partially general: [*my* + NOUN], [POSS PRO + *mother*]
3. More general: [POSSESSIVE + NOUN]
4. Fully general: [DETERMINER + NOUN]

These levels of abstraction involve categorization at differing levels of generality. *Mother* is a collection of all the exemplars of this word encountered so far. POSSESSIVE PRONOUN is a level of categorization that ranges over specific words such as *my, your, our, his*, etc. NOUN obviously has an even broader range. Grammatical constructions arise, then, from the storage of frequently repeated sequences and the categorization of their parts at different levels of abstraction.

11. Sequentiality is more basic than hierarchy

In the previous sections I hope to have demonstrated that there is a correspondence between frequency of co-occurrence and traditionally-established notions of constituent structure. Now the question arises as to the directionality of that correspondence: do elements occur together frequently because they are in the same constituents or are constituents derived from clusters of items that occur together frequently? In this section I argue that sequentiality is basic and that constituents and hierarchies arise because frequently-used strings are chunked as single units.

First, observe that learning of sequences of behavior is part of implicit learning in other domains and that such sequences naturally fall into hierarchies with the more frequently-used sequences recombined into larger units to form higher-level constituents (Fentress 1983). Driving a car involves a number of automatic sequences such as shifting into reverse or drive, braking to slow down or stop, putting on a turn signal, turning right or left. On a frequently-driven path, such as from one's home to one's office, these chunks of behavior are sequenced in a particular way and constitute a larger constituent that itself can be automated, as evidenced by the mistake of following this frequent path even when the destination is elsewhere. Also, any of the subunits of the whole sequence, such as braking, accelerating, turning left or right, can be recombined infinitely so that one can drive anywhere one wants to go. For the experienced driver, such recombinations occur with great ease and fluency, just as the native speaker recombines the automated chunks of language to produce a sentence s/he has never produced before. Thus the creation of hierarchy out of automated sequences of behavior is a domain-general cognitive process.

It is by now well-established that Broca's area deals not just with language, but also with sequential neuromotor behavior (Greenfield 1991; Roland 1985). In addition, many researchers have stressed the importance of left-hemisphere dominance for both motor control and language (Armstrong, Stokoe and Wilcox 1995; Corballis 1989 and Kimura 1979, 1993). Thus the hierarchy in automated motor activities and the hierarchy in grammar could stem from the same neurological source. Moreover, recent research into the perceptual processing of predictable visual stimuli suggests that Wernicke's area processes predictable events in time and may not be exclusively associated with language (Bischoff-Grethe *et al.* 2000). Since predictability is the perceptual side of sequentiality, it may turn out that the mechanisms behind the ability to perceive linguistic sequences and perhaps group them into constituents may also be domain-general.

Humans from 12 months to adulthood can learn repeated sequences of meaningless syllables, as shown by Saffran *et al.* 1996; Gomez and Gerken 1999, 2000. Moreover, Gomez has recently shown that both babies and adults can learn sequences of two nonce words that are separated by a third 'word' chosen from a large class (Gomez 2001). Thus meaning is not necessarily involved in learning sequences, suggesting that the basis for constituent structure may be recurring sequences and not just semantics. (See also Santelmann and Jusczyk 1998.)

In addition, there is purely linguistic evidence for the dominance of sequentiality over hierarchy in cases where fusion between elements in different constituents occurs because the two elements occur frequently in sequence. Examples of such cases will be discussed in the next two sections (sections 12 and 13). Linguistic cases of the opposite type, that is, those that appear to show hierarchy dominating sequentiality, as when constituents are discontinuous are discussed in section 14, where it is argued that the linear constraints on discontinuous constituents, such as the Heavy NP Constraint and the Distance Principle, demonstrate that sequential organization underlies even discontinuous constituents.

12. Chunking in violation of 'constituent structure'

Linguistic units chunk together according to repeated sequences, sometimes in violation of usual notions of constituent structure (Bybee and Scheibman 1999). A very robust example is English auxiliary contraction, which occurs in *I'm, I've, I'd,* *I'll, he's, he'll, he'd*, etc. The auxiliary, which is the first element in the VP, contracts with the subject NP, which is in the other major constituent of the clause. Thus contraction occurs across the primary constituent boundary of the clause. The reason for this is that the pronoun + auxiliary or NP + auxiliary combinations are highly frequent. In a large corpus of British English, Krug (1998) finds that contraction occurs in the most frequently-occurring combinations. The most common contraction is between *I* and *am* and that is also the most frequent sequence.

But the auxiliary also occurs right before the verb, so why doesn't the auxiliary fuse with the following verb, where it belongs semantically and syntactically? In order to answer this question I examined the distribution of auxiliaries in spoken American English, using the Switchboard corpus again. In Table 5 we see the token count of the ten most frequent items to precede and follow the auxiliary *will*, in its full and reduced version, *'ll*.

Table 5. Ten most frequent items occurring before and after *will* and *'ll*. (Switchboard corpus)

Preceding		Following	
I	918	*be*	466
they	471	,	244
we	368	*have*	199
it	256	*get*	130
you	200	*go*	119
that	183	*do*	103
he	122	*probably*	90
she	53	*just*	81
,	47	*tell*	75
people	38	.	42

What we see is that, indeed, the most frequent items to precede *will* or *'ll* are pronouns and the most frequent items to follow are verbs. What governs the contraction is the asymmetry in the frequency of the preceding versus the following items. Note that the most frequent pronoun (*I*) preceding *will* is twice as frequent as the most frequent verb (*be*) following *will*. A similar pattern is found for all contracted auxiliaries in the Switchboard corpus. Not only are the fused items in different major constituents, but also they have no semantic relevance to one another. The fusion seems due entirely to frequency of co-occurrence.

Contraction is recorded in Switchboard for all the items listed here as preceding *will* except for *people*. The only other preceding items showing contraction are *there, this* and *who*. In other words, contraction only occurs between the most frequent combinations in the case of *will*. Apparently contraction can generalize from the most frequent items to a general class of items, as demonstrated by the contraction of *has* to *'s*, which appears in the data with a large class of full NPs.

Other details of the distribution of auxiliaries support the Linear Fusion Hypothesis. For all the modal auxiliaries except *can* and *can't*, the most frequent items to follow are *be*, *have* and the negative.

(i) When *have* is part of the Perfect construction, it contracts with the preceding modal, as in *could've, should've, would've, must've, might've*.
(ii) Of course the negative *not* also contracts in some cases, e.g, *couldn't, shouldn't, wouldn't*.
(iii) Interestingly, *be* also forms a unit with some of these modals, but of a different sort. The combinations *could be* and *maybe* become lexicalized units that serve as independent epistemics, while *would be* has become an adjective.

Thus we have ample evidence from the auxiliaries that high frequency sequences become chunked into units.

This paper is not the first to observe that the combination of subject pronouns or nouns plus the auxiliary in English behaves like a constituent. Halliday and Hasan (1976: 197) call this complex the Modal Element and distinguish it from the Propositional Element (roughly the remainder of the clause). Halliday (1985) regards this collocation of units to be a constituent in the structure imposed on the clause by the interpersonal metafunction where again the Modal Element (subject + finite element of verb group) is distinguished from the Propositional Element. This level of constituent analysis coincides, then, with the frequency of co-occurrence found in conversational data and evidence from contraction that we have just considered.

The distribution of pronouns and auxiliaries and verbs, along with the Linear Fusion Hypothesis, explain why English auxiliaries contract with a preceding pronoun (and in some cases, nouns) rather than becoming prefixes on the verb. If similar distributions occur in other languages it may explain why in languages where the auxiliary follows the verb (SOV languages) there is massive suffixation, while in languages where the auxiliary precedes the verb (VO languages) there is not a comparable trend toward prefixation (Bybee, Pagliuca and Perkins 1990).

13. Other cases of non-constituents fusing.

Another robust phenomenon demonstrable across a wide range of languages is the binding of a verb and preposition into a unit. Reh (1986) discusses this phenomenon in African languages as an explanation for why case affixes are almost always suffixes; that is, why prepositions do not tend to become case prefixes.

Reh (1986) points out that in several languages of the Southern Lwo family, a preposition following a verb becomes a suffix on the verb. This can be seen by examining the dative/benefactive preposition in Dhuluo, which is *nɪ* as shown in examples (4) and (5), where it is positioned between the verb and the NP that is its object.

Dhuluo (Southern Lwo; Western Nilotic)
(4) Otieno o-kelo nɪ Odhiambo kitabu.
O. PERF-bring DAT/BEN O. book
'Otieno has brought a book to Odhiambo.'

(5) Onyangɔ tiyo nɪ japuonj.
O. IMPF:work DAT/BEN teacher
'Onyango works for the teacher.'

In Lango, the cognate preposition and the pronoun that was its object have fused with the verb to form 'the benefactive stem.' Thus *-kèlò* 'bring' + *nɪ* > *-kèllì* 'bring for someone', as shown in (6).

(6) Lango
Ò-kèllì dàkô.
3SG-bring-BEN woman
'She brought it for the woman.'

In this case, then, the former preposition has fused with the preceding verb rather than with the following noun. I propose that the explanation for this is that that particular prepositions would tend to occur with certain verbs, such as verbs meaning 'bring' or 'give', while the noun that follows is much less predictable, presumably being drawn from the entire class of human nouns and perhaps some non-human ones as well.

Analogous situations, but usually without affixation, can be found in European languages. Second-language learners of Spanish and French must learn lists of verb + preposition combinations, as particular verbs select particular prepositions. For instance, Spanish *pensar en* 'to think about', *acabar de* 'to finish' and *comenzar a* 'to begin to'. Again, the verb + preposition sequence would be more frequent than any particular preposition + noun or preposition + infinitive sequence.

Another common fusion across constituent boundaries is the fusion of prepositions and determiners, European languages, e.g. Spanish and French. For instance Spanish: *a* 'to, at' and *el* 'the (MASC.SG) > *al*, *de* 'of, from' + *el* > *del*. In this case, as in the others, it is plausible to assume that the frequent co-occurrence of these grammatical items leads to their fusion. Note that there is no particular semantic relevance (in the sense of Bybee 1985) or semantic affinity between the meaning of a preposition and that of a determiner. This appears to be a case of pure sequentiality.

14. Discontinuous constituents and discontinuous dependencies

A major argument that the utterances of a language are not just linear strings of words is the fact that non-adjacent elements can be in the same construction or constituent. Thus in the English verb + particle combinations, the particle can be separated from its verb by a pronoun or short NP, as in *look the number up; look it up*. However, the separation of the constituents does not necessarily mean that their connection is not still linear or sequential. Other types of neuromotor behavior can be suspended at certain points and then resumed with the association still being sequential. Furthermore, the predictability of the second element from the first can still be maintained across intervening items, as when the phrase *look the number* leads to the expectation of hearing *up*. This is analogous to waiting for the other shoe to drop.

A purely linguistic argument for the importance of sequentiality even in these cases is the well-documented existence of constraints on the material intervening between the two parts of the constituent. The Heavy NP Constraint describes the fact that, for instance, an NP with a lot of modifiers does not do well between the verb and its particle. Chen (1986) shows that the separation of the verb and the particle is constrained by the number of syllables in the direct object. Separation of the verb and particle is practically non-existence in both spoken and written language for direct objects of more than five syllables. Thus (7) and (8) would be very rare or non-existent in discourse and thus have a very awkward feel:

(7) I need to look the number that I lost up.

(8) I need to look a word that I can't remember how to spell up.

Other kinds of discontinuous dependencies also rely on sequentiality. For example, in French certain main clause verbs can have Subjunctive verbs in their subordinate clauses. In Canadian French, the use of Subjunctive or Indicative is variable, and apparently is not meaningful. Poplack (1992) has studied the variables that affect the mood choice and finds that certain high frequency main clause verbs in combination with certain high frequency subordinate clause verbs are more likely to have the Subjunctive.

Examples (9) and (10) illustrate this variability. (*Faut* is a reduction of *il faut que*, which means 'it is necessary'.)

(9) Bien certain, faut qu'ils *aient* (S) une place eux-autres aussi pour vivre.
'Well, of course, they should have a place to live, too.'

(10) Faut *j'aille* (*S*) voir pour de l'ouvrage.
'I have to go look for a job.'

I would argue that such dependencies are basically sequential. In Poplack's analysis of Canadian French, it turns out that if a parenthetical expression intervenes between the main verb and the subordinate verb, the Subjunctive verb form is much less likely to appear. In other words, intervening material can disrupt the dependency, suggesting that the use of Subjunctive results from the selection of an automated chunk.

As mentioned above, infants are sensitive to sequential regularities in input even in the absence of semantic factors. Santelmann and Jusczyk (1998) found that 18-month-old infants can discriminate between grammatically correct English discontinuous constituents and ungrammatical ones. The construction in question was the Progressive, in which the finite form of the verb *to be* forms a construction with the suffix *-ing* with an open class of verbs intervening. Thus the infants demonstrated Head Turn Preference for the natural passages such as *Everyone is trying to be careful* vs. the unnatural one, *Everyone can trying to be careful.* When extra words were inserted between the discontinuous constituents, the infants no longer preferred the natural passages, suggesting that this is indeed a sequential effect that can be disrupted by intervening words.

Gomez (2001) also tested 18-month-olds using sequences of three nonce 'words'. The infants learned to discriminate sequences they had heard before even though other 'words' from a large set always came between the first and third word. Since the nonce words are meaningless, this experiment demonstrates that sequential dependencies can be detected and learned in the absence of meaning.

The sequentiality hypothesis makes general predictions about center embedding, i.e. that center embedding would be constrained by the length of the embedded unit. To my knowledge studies of the length of center embedded clauses in natural discourse have not been undertaken.

15. Conclusion

Linguists have been accustomed to viewing language as emanating from a mental structure that is autonomous from actual usage events. A more explanatory view is afforded by recent functionalist views of language as highly contextualized and embodied (Fox 2001). Most tokens of language use are routine events that respond to the environment — both social and physical, i.e., the people encountered and the utterances they produce. These responses are partially automatic, though they do involve an assessment of the environment and the choice of an appropriate response, as do other fine-tuned neuromotor behaviors. As with other neuromotor skills, language responds to practice. Perceptual skill also improves with repetition. Thus we have every reason to believe that repetition could be the main factor that builds up structure in language (Haiman 1994). Its importance in grammaticization has been well documented. Here I suggest that repetition of sequences of units is the main factor in the creation of linguistic patterns that have been identified as constituent structure.

My proposal is as follows. Constituents of the type proposed for generative grammar which are described by phrase structure trees do not exist. Instead, units of language (words or morphemes) are combined into chunks as a result of frequent repetition. Most of the time the units of these chunks bear a semantic and/or pragmatic relation to one another allowing them to fulfill the grammatical criteria for constituency: they can sensibly be used alone, as in the answers to questions; they can be replaced by a pro-form; and they can be used in various positions in a sentence (see the examples in section 2). In such cases, where frequency of co-occurrence corresponds to semantic relevance, we have traditional constituents. Indeed, the semantic coherence of such units may facilitate their establishment as chunks. However, other types of chunks also exist, as I have demonstrated in this paper, showing that frequency of co-occurrence is an independent factor. Thus pronoun + auxiliary, preposition + determiner, and verb + preposition sequences can form chunks but are difficult to describe in traditional frameworks since they do not meet the criterion of semantic relevance. For this reason, too, they do not fulfill the grammatical criteria of occurring alone or being replaceable by a pro-form. Thus constituency in this view is the convergence of two other factors and is itself not a basic structure. It is an emergent property of language.

A second point is that this emergent constituency differs from traditional constituency in that it can be gradient, since the factors determining it are themselves gradient. Gradience in constituency refers to the fact that different items of the same putative category might fuse less with one another. If frequency of co-occurrence is a major determinant of emergent constituency, then the two units in *my mother* are in a tighter constituent bond than the two units in *my appointment*. This difference has no overt consequences in English, but in some languages, it would be manifest as a difference between alienable and inalienable possession, where the latter always has a more fused expression (Nichols 1988). Other gradient differences in frequency of co-occurrence do have overt consequences, as seen in the fact that *I don't* is more fused than *they don't* and *hit'im* is more fused than *hit the ball*. These relations cannot be captured by using the same phrase structure tree for all instances of the same traditional constituent.

Considering now the evolution of language, the development of grammar may be easier to explain in terms of domain-general abilities than many researchers have supposed (e.g. Pinker and Bloom 1990). If constituent structure is epiphenomenal, then a theory of the evolution of language need not account for it directly, but rather by reference to the processes that underlie it and these appear to be domain-general. The abilities include (i) highly advanced motor skills, fine motor control and the associated neurological capabilities including the ability to compose, store and access relatively long sequences of behavior; (ii) the ability to combine concepts into communicatively coherent sequences which in turn is based on (iii) extensive categorization capacity that is applied to both phonological form and meaning; and (iv) the ability to store and categorize vast quantities of prefabricated sequences. These abilities interact in that one may facilitate the others. In particular, the semantic coherence of units in sequence may make it possible to compose longer sequences more fluently. And, as I emphasized throughout, the automation of lower-level sequences makes the composition of hierarchically complex sequences possible. Thus abilities that are neither domain-specific nor species-specific interact in current language processing to create the apparent structure that is grammar. As these abilities evolved from a more primitive to a more advanced stage, language might also have evolved from a set of relatively short utterances consisting of first one, then two units, to much longer utterances with apparent hierarchical structure via the concatenation of preformed chunks.

Notes

1. Langacker (1997), which is a discussion of the iconic relations between conceptual structure and syntactic structure, is an exception.

2. Interestingly, the best examples of syntactic constituents in English seem to be noun phrases. Noun phrases also have the best support as constituents in the discourse literature (Ono and Thompson 1994). Verbs and verb phrases and prepositional phrases present certain problems, as we will see below.

3. On the difficulties of identifying the VP as a constituent, see Givón (1995).

4. In selecting these nouns, there was a conscious bias towards count nouns. In addition, it was immediately obvious that kinship terms, such as *mother* and *husband* had the special property of occurring more with possessive pronouns, and thus only two of them were included. Bleached or grammmaticized nouns such as *thing, couple* and *stuff* were also passed over.

5. The Switchboard transcriptions distinguish several types of pauses, but I have added them all together for the purposes of Table 1.

6. One might expect a higher percentage for *the* as the item to precede a noun. Table 3 shows that there are quite a number of frequently-occurring determiners competing with *the*. In addition, some nouns frequently occur in compounds, so that the item preceding them is another noun. For instance, the item to most frequently precede *school* is *high*.

7. In the model developed in Bybee (1985), semantic and phonological connections can be made independently, but when they coincide a morphological relation is indicated.

References

Anderson, J.R. 1993. *Rules of the Mind.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Armstrong, D., Stokoe W., and Wilcox, S. 1995. *Gesture and the Nature of Language.* Cambridge: Cambridge University Press.

Bischoff-Grethe A., Proper S.M., Mao H., Daniel K.A., and Berns G. 2000. "Conscious and unconscious processing of nonverbal predictability in Wernicke's area". *The Journal of Neuroscience* 20: 1975–81.

Boyland, J.T. 1996. *Morphosyntactic Change in Progress: A Psycholinguistic Approach.* Dissertation: University of California.

Bybee, J.L. 1985. *Morphology: A Study of the Relation between Meaning and Form.* Amsterdam: John Benjamins.

Bybee, J.L. 1995. "Regular morphology and the lexicon". *Language and Cognitive Processes* 10: 425–55.

Bybee, J.L. 1998a. "The emergent lexicon". *CLS 34: The Panels,* 421–35. University of Chicago: Chicago Linguistic Society.

Bybee, J.L. 1998b. "A functionalist approach to grammar and its evolution". *Evolution of Communication* 2: 249–78.

Bybee, J.L. 2001. *Phonology and Language Use.* Cambridge: Cambridge University Press.

Bybee, J.L., Pagliuca W., and Perkins, R. 1990. "On the asymmetries in the affixation of grammatical material". In *Studies in Diachronic Typology for Joseph H. Greenberg,* W. Croft, K. Denning, and S. Kemmer (eds), 1–42. Amsterdam: John Benjamins.

Bybee, J.L., Pagliuca W., and Perkins, R. 1994. *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World.* Chicago: University of Chicago Press.

Bybee, J.L. and Scheibman J. 1999. "The effect of usage on degree of constituency: The reduction of *don't* in American English". *Linguistics* 37: 575–96.

Chen, P. 1986. "Discourse and particle movement in English". *Studies in Language* 10 (1): 79–95.

Chomsky, N. and Halle M. 1968. *The Sound Pattern of English.* New York: Harper and Row.

Cipollone, N., Keiser, S.H., and Vasishth S. (eds). 1998. *Language Files: Materials for an Introduction to Language and Linguistics,* 7th edition. Columbus: Ohio State University Press.

Corballis, M.C. 1989. "Laterality and human evolution". *Psychological Review* 96: 492–505.

Erman, B. and Warren B. 2000. "The Idiom Principle and the Open Choice Principle". *Text* 20: 29–62.

Fentress, J.C. 1983. "Hierarchical motor control". In *The Psychobiology of Language,* M. Studdert-Kennedy (ed), 40–61. Cambridge, Mass.: MIT Press.

Fox, B. 2001. "On the embodied nature of grammar: Embodied being-in-the-world". In *Complex Sentences in Grammar and Discourse: Essays in Honor of Sandra A. Thompson,* J.L. Bybee and M. Noonan (eds). Amsterdam: John Benjamins.

Givón, T. 1995. *Functionalism and Grammar.* Amsterdam: John Benjamins.

Givón, T. 1998. "Toward a neurology of language". *Behavioral and Brain Sciences* 21: 154–5.

Givón, T. This volume. "The visual information-processing system as an evolutionary precursor of human language".
Gomez, R. and Gerken, L. 1999. "Artificial grammar learning by 1-year olds leads to specific and abstract knowledge". *Cognition* 70: 109–35.
Gomez, R. and Gerken L. 2000. "Infant artificial language learning and language acquisition". *Trends in Cognitive Sciences* 4: 178–86.
Gomez, R.L. 2001. "Finding structure in language: Sometimes more variability is better." Presentation given at the annual meeting of the AAAS in the symposium: Tools Infants Might Use to Learn Language. San Francisco, February.
Greenfield, P.M. 1991. "Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior". *Behavioral and Brain Sciences* 14: 531–95.
Gregory, M., Raymond, W.D., Bell A., Fosler-Lussier E., and Jurafsky, D. 1999. "The effects of collocational strength and contextual predictability in lexical production". *CLS* 35.
Haiman, J. 1994. "Ritualization and the development of language". In *Perspectives on Grammaticalization*, W. Pagliuca (ed), 3–28. Amsterdam: John Benjamins.
Halliday, M.A.K. 1985. *An Introduction to Functional Grammar.* London: Edward Arnold.
Halliday, M.A.K. and Hasan, R. 1976. *Cohesion in English.* London: Longman.
Hopper, P. and Traugott, E. 1993. *Grammaticalization.* Cambridge: Cambridge University Press.
Jurafsky, D., Bell A., Gregory M. and Raymond, W.D. 2001. "Probabilistic relations between words: Evidence from reduction in lexical production". In *Frequency and the Emergence of Linguistic Structure*, J.L. Bybee and P. Hopper (eds), 229–54. Amsterdam: John Benjamins.
Kahr, J.C. 1976. "The renewal of case morphology: sources and constraints". *Working Papers on Language Universals* 20: 107–51. Stanford, Calif.: Stanford University Press.
Kimura, D. 1979. "Neuromotor mechanisms in the evolution of human communication". In *Neurobiology of Social Communication in Primates*, H.D. Steklis and M.J. Raleigh, (eds), 197–219. San Diego, Calif.: Academic Press.
Kimura, D. 1993. *Neuromotor Mechanisms in Human Communication.* Oxford: Oxford University Press.
Krug, M. 1998. "String frequency: A cognitive motivating factor in coalescence, language processing and linguistic change". *Journal of English Linguistics* 26: 286–320.
Langacker, R. 1987. *Foundations of Cognitive Grammar*, Volume I. *Theoretical Prerequisites.* Stanford: Stanford University Press.
Langacker, R. 1997. "Constituency, dependency and conceptual grouping". *Cognitive Linguistics* 88: 1–32.
Lindblom, B., MacNeilage P., and Studdert-Kennedy, M. 1984. "Self-organizing processes and the explanation of phonological universals". In *Explanations for Language Universals*, B. Butterworth, B. Comrie, and Ö. Dahl, (eds), 181–203. Berlin: Mouton de Gruyter.
Nichols, J. 1988. "On alienable and inalienable possession". In *In Honor of Mary Haas: From the Haas Festival Conference on Native American Linguistics*, W. Shipley (ed), 557–609. Berlin: Mouton de Gruyter.

Ono, Y. and Thompson, S. 1994. "Unattached NPs in English conversation". *BLS* 20.

Pawley, A., and Syder, F.H. 1983. "Two puzzles for linguistic theory: Nativelike selection and nativelike fluency". In *Language and Communication*, J.C. Richards and R.W. Schmidt (eds), 191–226. Longman: London.

Poplack, S. 1992. "The inherent variability of the French subjunctive". In *Theoretical Analyses in Romance Linguistics*, C. Laeufer and T.A. Morgan (eds), 235–63.

Pinker, S. and Bloom, P. 1990. "Natural language and natural selection". *Brain and Behavioral Sciences* 13:707–84.

Radford, A. 1988. *Transformational Grammar: A First Course.* Cambridge: Cambridge University Press.

Reh, M. 1986. "Where have all the prefixes gone?" *Africanistische Arbeitspapiere* 5:121–34.

Roland, P.E. 1985. "Cortical organization of voluntary behavior in man". *Human Neurobiology* 4:155–67.

Saffran, J.R., Alsin, R.N., and Newport, E.L. 1996. "Statistical learning by 8-month-old infants". *Science* 274:1926–1928.

Santelmann, L. and Juczyk, P.W. 1998. "Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space". *Cognition* 69:105–34.

Scheibman, J. 2002. *Pont of View and Grammar: Structural patterns of subjectivity in American-English conversation.* Amsterdam: John Benjamins.

Sinclair, J. 1991. *Corpus, Concordance, Collocation.* Oxford: Oxford University Press.

Sosa, A.V. and MacFarlane, J. To appear. "Evidence for frequency-based constituents in the mental lexicon: collocations involving the word *of*". *Brain and Language.*