

# USAGE-BASED GRAMMAR AND SECOND LANGUAGE ACQUISITION

Joan Bybee

## 1 Introduction and background

The structuralist theories that dominated the field of linguistics during the twentieth century advocated a separation of language use from the more abstract knowledge of language structure. For Saussure this distinction went under the labels of *langue* (knowledge of language) and *parole* (speech) (de Saussure, 1915/1966). In Chomskian theory the distinction is made between *competence* (tacit knowledge of language structure) and *performance* (the actual use of language) (Chomsky, 1965). Because knowledge of structure was taken to be the main object of study, there was little interest in the potential effects that usage might have on cognitive structures. However, common sense tells us that for second language learners, repeated exposure and practice are essential to the development of the cognitive structures that lead to fluent and grammatical speech.

More recently developed Usage-Based Theories of language recognize the impact of usage on the cognitive representation of language (Bybee, 1985, 2001, 2006; Langacker, 1987, 2000). Many empirical studies have now appeared relating frequency of use to various structural phenomena. In addition, many recent studies have revealed the extent to which language users retain specific information about their experience with language (for reviews see Bod, Hay, & Jannedy, 2003; Bybee & Hopper, 2001; Ellis, 2002). From these studies there is now emerging a new theory of language in which grammar is viewed as the cognitive organization of one's experience with language (Bybee, 2006). In this view, as users of language experience tokens of language use, they categorize them at varying degrees of abstractness. This categorization process creates a vast network of phonological, semantic and pragmatic associations that range over what

has traditionally been designated as lexicon and grammar. The network resulting from these categorized experiences is affected by repetition (frequency of use) in the ways that will be outlined below. In addition, it contains both specific and generalized information about form, meaning, and context of use of words and constructions.

In this chapter I will review this evidence and its theoretical implications both for native speakers' grammars and for a theory of second language acquisition. This theory, in which usage impacts grammar, has particular relevance for second language teaching and learning.

Some of the discussion will be phrased in terms of constructions, which are stored pairings of form and function that range over units at the level of the word up to and including complex sentences. Several versions of grammar in terms of constructions have been discussed in the literature with proposals being made by Fillmore and Kay (e.g. Fillmore, Kay & O'Connor, 1988; Fillmore & Kay, 1994), Goldberg (1995, 2003), Lakoff (1987), Langacker (1987) and Croft (2001). In all of these proposals, the following constitute constructions (see Goldberg, 2003):

- 1 simple lexical words: *table, decide, pretty*
- 2 grammatical morphemes and the items they appear with: VERB + Past Tense; *the* + NOUN.
- 3 Idioms with fixed lexical content: *go great guns*
- 4 Idioms that are partially filled: *jog* <someone's> *memory*
- 5 Constructions with some fixed material: *he made his way through the crowd*
- 6 Fully abstract constructions: *they gave him an award.*

Almost all constructions contain some explicit morphological material, tying them fairly concretely to specific words or morphemes (e.g. Past Tense *-ed*, the determiner *the*, or *way* and the possessive pronoun in (5)). The ditransitive construction in (6) contains no specific morphological material that identifies it as the ditransitive. Only the word order signals this. However, it should be noted that only a small class of verbs can occur in this construction so that it also has a grounding in particular lexical items.

Taking constructions as the basic unit of grammatical analysis has two advantages: on the one hand with constructions we can see the continuum from lexicon to grammar (as shown in (1) through (6)), and on the other hand constructions allow us to represent the interactions of specific lexical units with specific grammatical configurations (as seen in the relations between *jog* and *memory* or between *give* and the ditransitive construction).

When constructions are combined with a Usage-Based model the result is a theory that proposes that grammatical structures are built up

through experience with specific examples of constructions which are categorized in memory by a mapping process that matches strings for similarity and difference. The resulting cognitive representations are abstractions over one's cumulative experience with language. However, as mentioned above, there is evidence that knowledge of specific instances of language use is not entirely lost in this abstraction process and especially with reinforcement through repetition, specific instances of constructions can have memory representations. In this view of grammar, then, frequency of use plays an important role in determining cognitive structures. In the next section I will discuss the particular frequency effects that have been discovered in recent research.

## 2 Frequency effects

Recent research has shown that repetition of linguistic units has an impact on cognitive representations. To demonstrate this, a distinction must be made between token and type frequency. **Token frequency** counts the number of times a unit appears in running text. Any specific unit, such as a particular consonant [s], a syllable [ba], a word *dog* or *the*, a phrase *take a break*, or even a sentence such as *You know what I mean* can have a token frequency. **Type frequency** is a very different sort of count. Only patterns of language have type frequency because this refers to how many distinct items are represented by the pattern. Type frequency may apply to phonotactic sequences; it would be the count of how many words of the language begin with [sp] versus how many begin with [sf]. It may apply to morphological patterns, such as stem + affix combinations. For instance, the English past tense pattern exemplified by *know, knew; blow, blew* has a lower type frequency than the regular pattern of adding the *-ed* suffix. Syntactic patterns or constructions also have type frequencies: the ditransitive pattern in English, exemplified by *He gave me the change* is used with only a small set of verbs, while the alternate pattern *He gave the change to me* is possible with a large class of verbs (Goldberg, 1995). In this section we treat the three known effects of token frequency and the effect of type frequency, as well as their interaction.

### 2.1 Token frequency: three effects

The first effect of token frequency to be discussed, which we can call the **Conserving Effect**, depends upon the fact that repetition strengthens memory representations for linguistic forms and makes them more accessible. Accessibility in this sense refers to the fact that, in experiments where subjects are asked to say whether a string of letters or sounds is a word of their language, they respond much more quickly to high-frequency words than to low-frequency words. This greater accessibility

suggests that each token of use strengthens the memory representation for a word or phrase (Bybee, 1985).

The strength of representation of higher frequency forms explains why they resist reformation on the basis of analogy with other forms (thus the name, Conserving Effect). For instance, for English irregular verbs, there is a general trend diachronically towards regularization, a trend also witnessed in child language development. However, the higher-frequency verbs resist this trend; thus *keep*, *sleep*, *weep*, *leap*, and *creep* and other verbs of this shape acquired irregular past forms when the vowel was shortened in early Middle English, giving *kept*, *slept*, *wept*, *leapt*, and *crept*. Only the lower-frequency verbs of this class have subsequently developed regularized pasts *weeped*, *leaped*, and *creeped* (still used alongside the irregulars). The mechanism behind this type of change (analogical reformation) is that a new past form is created by accessing the base/present form and adding the suffix *-ed* (in this case its allomorph [t]) to it. For those verbs of high frequency, the greater accessibility of the irregular past makes such a reformation unlikely. For this reason, the lower-frequency paradigms tend to regularize before the higher-frequency paradigms.

This effect of repetition should play out in Second Language Acquisition in a very straightforward way: the more exposure a learner has to irregular forms, the greater the chance that he/she will produce them correctly. Less frequent irregular forms are more likely to be treated by the learner as regular.

The second effect of token frequency, **Autonomy**, can be thought of as an extreme case of the preceding effect. Sequences are autonomous when they are frequent enough to be learned by rote and are not associated with the units that comprise them. For instance, children may learn to say *gimme* without realizing that it consists of *give + me*. For second language learners, they may be autonomous if they are learned before the knowledge of the language has developed enough to allow analysis. For instance, I learned the Spanish word *ádale* which I assumed meant “hurry up” from exposure to Spanish speakers, but I figured out that it consisted of the verb *andar* “to walk” plus a pronoun *le* only after several years of formal study.

The third effect, the **Reducing Effect**, refers to the common observation that oft-repeated phrases, such as greetings (*God be with you* > *goodbye*, *how are you* > *hi*) and titles, tend to reduce phonetically. The same observation applies to much-used grammatical items, such as auxiliaries, modals, negatives, and pronouns, and phrases such as *be going to*. In addition, it has recently been fully documented that reductive sound change applies probabilistically across all frequency levels, affecting high-frequency items more quickly and radically than low-frequency items (Bybee, 2002b). The reason for this trend is that repetition of neuromotor sequences leads to

greater overlap and reduction of the component gestures. As articulation becomes more efficient, the output appears more and more to have been affected by assimilation and reduction.

The Reducing Effect presents an enormous challenge to second language learners in both production and perception. A phrase such as *going to* when used with a verb in its future sense can be produced as [gɔɪŋtu], [gɔ̃nə] or with “I” as [aɪmənə]. It is very difficult for students to acquire good approximations of these native variants as the phonetics of their native languages would most likely have different reduction and coarticulation effects.

## 2.2 *Grammar as automatized behavior*

These well-documented effects of token frequency point strongly to grammatical knowledge as automatized behavior. All language use involves procedural knowledge, as distinct from declarative knowledge, which is a knowledge of a fact, such as “whales are mammals.” Procedural knowledge is bound up with neuromotor events and is knowledge of how to do something, such as drive a car or stir pancake batter. While these kinds of actions are learned through the mastery of a sequence of actions, with practice they are repackaged into a single action. Once automatized, it is difficult to unpack them again. The linguistic routines that make up the grammar and lexicon of a language are like these procedures: they develop over time with practice, they are difficult to unpack and native speakers are often unaware of their component parts and, at times, even their meaning. In contrast, one might want to say that the meanings of lexical items, particularly nouns, is a matter of declarative knowledge, but while this is the case in part, procedural knowledge is also necessary to access lexical items and put them in an appropriate context.

Ellis (1996) discusses the importance of “chunking” for L2 acquisition. Chunking is a property of procedural knowledge; sequences of actions that occur together repeatedly are chunked into a single action. Chunking is possible with sequences of just two actions or it could occur with very long sequences of actions. However, the probability that a smaller number of actions recur is much greater than the probability that a long sequence will recur, so the degree of fusion within small chunks is greater than within larger ones. In Bybee (2002a) I argue that the embedding of small chunks within larger ones is what gives us the grammatical notion of hierarchies or tree structures. In Usage-Based Theory, constructions are chunks—neuromotor routines—with movable parts. They are established through practice and processed as single units. Since chunking occurs naturally and unconsciously with practice, even adults easily learn new chunks in their own language. It is an ability that is necessary and fortunately available for L2 learning. Of course, the difference between

L1 and L2 speakers is that the L1 speakers already have the necessary components of the chunks automatized and ready for re-use in new chunks.

The downside of the fact that repetition leads to chunking is fossilization in Second Language Acquisition: a learner repeats a sequence frequently but incorrectly and it is very difficult to change the internal structure of this chunk once it has become automatized (see Ellis, 2006a; Long, 2003).

### 2.3 Type frequency

As mentioned above, type frequency is a property of patterns or constructions and refers to the number of distinct items that can occur in the open slot of a construction or the number of items that exemplify a pattern, such as a phonotactic sequence. Type frequency is a major factor determining the degree of productivity of a construction (Bybee, 1985; Guillaume, 1927/1973; MacWhinney, 1978). That is, constructions that apply to a large number of distinct items also tend to be highly applicable to new items. In determining productivity, however, factors other than type frequency must also be taken into account: often the member items that occur with a construction must also belong to certain phonological or semantic categories. The verbs of the *string, strung* class must end in a nasal or a velar (Bybee & Moder, 1983); the adjectives that can be used in the construction [X drives me (or someone) ADJ], (as in *it drives me mad, it drives me crazy*) must suggest some degree of insanity, either literally or figuratively (Boas, 2003).

The contribution of type frequency to productivity is due to the fact that, when a construction is experienced with different items occupying a position, it enables the parsing of the construction. If *happiness* is learned by someone who knows no related words, there is no way to infer that it has two morphemes. If *happy* is also learned, then the learner could hypothesize that *-ness* is a suffix, but only if it occurs on other adjectives would its status as a suffix become established. Thus, a certain degree of type frequency is needed to uncover the structure of words and phrases. In addition, a higher type frequency also gives a construction a stronger representation, making it more available or accessible for novel uses. Hay and Baayen (2002) have proposed several other factors that facilitate parsing and thus influence productivity.

The implications of these findings for second language learning is obvious: if a learner is to apply a pattern productively to forms not necessarily encountered before, he/she must have encountered the pattern with a number of different items in it. We will refine this prediction in the next section.

## 2.4 Interactions of type and token frequency

One result of the interaction of type and token frequency has already been mentioned: a very high-frequency instance of a pattern might be learned and stored as if it were autonomous from the more general pattern. Moder (1992) experimented with the English irregular verb classes, such as *string/strung*, *fling/flung* and *write/wrote*, *ride/rode*, *steal/stole*. She made up nonce verbs that fit these patterns and asked subjects what their past tense forms would be. If she first presented them with a real member of the class that was medium-frequency, then that made the subjects more likely to give a past tense with a vowel change that fit the pattern. However, if she presented them with a very high-frequency verb in the class, they were less likely to respond with a form that fitted the pattern. A similar result on phonotactics is reported by Bailey and Hahn (2001): English-speaking subjects rated nonce phonotactic patterns for how acceptable they would be in English. They tended to prefer phonotactic patterns that were similar to dense groups of similar words (dense neighborhoods). However, the very high-frequency words and the very low-frequency words did not contribute to this effect.

A very different effect of token frequency on productivity of patterns is found with instances of constructions that are relatively high in frequency, but not so high as to cause the autonomy effect just discussed. As I said above, if a construction (either within a word, such as VERB + Past Tense or across words, as the ditransitive construction (*I gave Jim the rest of the stew*)) can occur with lots of different items (in both of these examples, many different verbs), then it is likely to be productive, that is, applicable to new items. One could thus conclude that, in acquiring productivity, exposure to many different types in a construction would be more helpful than exposure to many identical tokens. Actually, recent results of an artificial language learning task by Casenhiser and Goldberg (2005) shows that both type and token frequency are important to learning productive constructions (see also Goldberg, this volume; Goldberg, Casenhiser & Sethuraman, 2004). They taught English-speaking children, aged five to seven, novel verbs in a novel construction. The construction contained a subject, verb, and object, but the word order was non-English—the verb was final in the sentence. In one condition the subjects heard the sentences (while watching a video that represented the meaning of the sentences) with three verbs twice and two other verbs only once. In this condition, then, type frequency was emphasized. In the other condition, one nonce verb was presented in the construction four times while the others were presented only once. It was in the latter condition that the subjects performed better in learning the construction.

As Casenhiser and Goldberg point out, it is a general property of categorization that there will be one higher-frequency member of the category

and that this member will be considered prototypical. In a non-linguistic task, Kotovsky and Gentner (1996) have shown children do better extending a simple relational similarity to a more complex one if they have lots of practice on the simple relation. These studies suggest that while varying the types in a construction is important to learning, holding the type constant over some repetitions facilitates learning, probably by increasing general familiarity with the relations in the construction. Note that the mechanism of extension of a grammatical construction assumed here is analogy. Analogy as a basic mechanism for linguistic production has received new attention in the psycholinguistic and modeling literature recently, including connectionist models (Rumelhart & McClelland, 1986), as well as various analogical models (Eddington, 2000; Krott et al., 2001; Skousen, 1989). That is, we are assuming that on the basis of some instances of exposure a structure is built up that can then be used in a new context by analogy with the stored instances. Examples from morphology and syntax will be provided in the next two sections.

### 3 Morphology: how usage affects the structure of paradigms

The frequency effects discussed above reveal themselves in the structure of morphological paradigms; indeed, in my own work their importance was made clear to me through the cross-linguistic study of morphology (Bybee, 1985). A full treatment of these phenomena with respect to First and Second Language Acquisition can be found in Bybee (1991).

#### 3.1 Token frequency

First consider the Conserving Effect. Mańczak (1980) has pointed out that the forms of a paradigm that are most frequently used are the ones that resist change and serve as the basis of change when new forms are created. To take our example above, when *weep/wept* regularizes to *weep/weeped*, it is the infinitive or present form, *weep*, that is taken as the basis for the new form, *weeped*, rather than a potential but very implausible regularization that would use the past base, *wep*, to create a new present.

Because of this strong tendency across languages, it is often the case that the most frequent form of a paradigm either has no affix or has a simple and short affix, and many of the other forms of the paradigm can be derived from it. Consider a regular present indicative paradigm of Spanish verbs:

	<u>Singular</u>	<u>Plural "sing"</u>
1st	<i>cánto</i>	<i>cantámos</i>
2nd	<i>cántas</i>	<i>cantáis</i>
3rd	<i>cánta</i>	<i>cántan</i>



(I have marked the stress on these forms even though the orthographic conventions would not normally write them on any form but the 2nd plural. Stress will become important in our discussion of the preterite and imperfective forms.) Note here that the stem could be regarded as *canta* and from this form all the forms except 1st singular can be derived by adding an affix. It turns out that the 1st and 3rd singular of the present tense are the most frequently used verb forms in Spanish (Bybee & Brewer, 1980).

A similar set of relations is evident in the imperfect, which indicates past imperfective—a habitual or progressive situation in the past.

	<u>Singular</u>	<u>Plural "sing"</u>
1st	<i>cantába</i>	<i>cantábamos</i>
2nd	<i>cantábas</i>	<i>cantábais</i>
3rd	<i>cantába</i>	<i>cantában</i>

Here we see that both the 1st and 3rd singular are lacking in person/number marking and thus can serve as the base upon which other affixes are added.

Not all verbs in all languages follow this pattern, but the pattern reflects common tendencies towards change, that is, reformation of paradigms based on 1st or 3rd singular forms. Of course, languages are conventional systems and change only slowly, so not all systems nor all verbs would succumb to the pressures of change. The importance of the tendencies towards change is that they give evidence for the types of cognitive representations speakers have. In this case, the tendencies for change demonstrate that the more frequent forms of the paradigm have stronger representations in memory and thus can be used as the basis for constructing the other forms. The frequency skewing in the experience of speakers is reflected in cognitive representations: higher-frequency forms have relatively stronger representations that are easy to access; lower-frequency forms have weaker, less accessible representations.

Second language learners can benefit from the natural frequency skewing in the input in the same way. The higher-frequency forms will have stronger representations (or can be presented first) and the lower-frequency forms can be learned as derivations from the higher-frequency forms.

Special evidence for this property of morphology comes from cases where an unexpected form is highly frequent. In these cases, we find a greater tendency towards irregularity. Thus, the English nouns that have irregular plurals are for the most part nouns that designate entities often referred to in the pairs or groups: *feet*, *teeth*, *mice*, *geese*, *men*, etc.

Also from morphology we have the lesson that high-frequency items can preserve older patterns longer than low-frequency items. The nouns just mentioned represent only a few members of an older class of nouns

that had vowel changes (from earlier umlaut) in Old English: for instance, the older plural of *book* would be modern *beech*, *brother* had earlier *brethren*, *daughter* had a plural *dehter*, and *friend* had a singular *freond* with a plural *friend*. Thus, the surviving members of this class have preserved the archaic patterns.

As mentioned above, more extreme levels of frequency lead to high autonomy. Examples are the copula verbs in most languages, which are usually highly irregular: English *am*, *is*, *are*, *was*, *were*; Spanish *soy*, *eres*, *es*, *somos*, etc. In such cases, where each form is very highly frequent, each form is likely to be autonomous from every other form, though their meanings remain very closely related. Autonomy is evident in second language learners who use high-frequency irregular past tenses correctly despite not yet being at a stage of consistent application of the regular past tense (Bayley, 1994, 1996).

To summarize this point: token frequency plays a role in morphology by making the higher-frequency forms of a paradigm the anchoring points for the other forms. Lower-frequency forms can be analyzed and learned in terms of these more robust forms, creating a relationship of dependency. In paradigms where several or all forms are of high token frequency, we have less dependency inside the paradigm and more autonomy of individual forms. Since these patterns arise from the natural occurrence of inflected forms in discourse, exposure to natural patterns of use should set up native-like representations in the L2 learner. However, because L2 learners in the classroom have so much less exposure to the target language than children in a natural environment, some measures will likely be necessary to ensure that these paradigmatic relations that depend upon relative frequency are established.

### 3.2 *Type frequency*

As mentioned above, type frequency plays an important role in the determination of productivity. In order for the L2 learner to acquire native-like patterns of productivity, regular patterns must be taught by methods that mirror to some extent natural exposure to the L2 patterns. As mentioned above, Casenhiser and Goldberg (2005) found that while type frequency was important to the acquisition of a new pattern, repeated exposure to a single type also facilitated acquisition. In terms of teaching productive morphology, the most effective method would have two stages: first, focus on a single lexical paradigm that exemplifies the productive pattern, and, only after this is quite familiar, advance to a second stage in which the pattern is exemplified with other lexical types. The repeated type provides familiarity with the relations in the pattern, while the range of different types aids parsing and provides practice of the analogical extension of the pattern.

#### 4 Syntactic constructions: "become" verbs in Spanish

The effects of frequency just discussed in the context of morphology can be extended to syntax though work in this domain has only just begun (Bybee, 2006; Bybee & Thompson, 1997). For instance, we find the Conserving Effect in syntax in the preservation of older constructions within high-frequency collocations or with high frequency lexical items. For instance, the question *how goes it?* uses the inversion of the main verb with the subject which was the general pattern up until the mid-sixteenth century. Now the more general, productive pattern uses an auxiliary: *how does it go?* (which does not mean the same thing!).

The older pattern of negation evident in phrases such as *She saw no one, he does nothing, we have no aspirin* are retained only with certain high-frequency verbs and in certain collocations, such as *that proves nothing*. Otherwise we use the more productive pattern exemplified by *she didn't see anyone, he doesn't do anything, we don't have any aspirin, that doesn't prove anything* (Tottie, 1991). Often when two patterns are competing in a language, one is older and preserved only in certain contexts, often the more frequent contexts, and the other is newer and more productive, being used in a wider range of contexts, including those of lower frequency.

Usage-Based Theory holds that the properties of constructions can best be studied in corpora of naturally-occurring discourse. Very often, linguists' intuitions about the range of occurrence of constructions do not match what is actually found in language use (Biber, Conrad, & Reppen, 1998; Biber & Reppen, 2002; Sinclair, 1991). By studying the use of constructions in discourse we can learn more about the way they are represented in the speaker's mental grammar. What we find is that the interaction of type and token frequency discussed above also applies to syntactic constructions.

Languages have thousands of constructions, tens of thousands if we include inflected words and idioms. But even restricting our count to constructions that contain an open slot, there are thousands that have to be learned. For instance, in Spanish (as in many other languages) there are a large number of verb + preposition combinations that are conventionalized and have to be learned individually. Examples are *pensar en* + NP "think about NP," *empezar a* + verb "to begin to + verb," *sonar con* + NP "to dream about NP," and *insistir en* + NP "to insist on NP." Each of these can be considered a construction with an open position. Continuing with Spanish examples, there is a very general construction for the placement of an NP direct object (after the verb) and a different one for the placement of pronominal direct objects (before the inflected verb). The way constructions can be effectively taught may depend upon their distribution in natural discourse.

Consider a set of constructions whose distribution in natural corpora

has been studied for the purpose of understanding their cognitive representations (Bybee & Eddington, 2006). Spanish has various ways of expressing the notion of entering a state. In particular, there are four verbs that can be used with an adjective with roughly the meaning of "become + adjective." The verbs are all reflexive in form, and are given here with the glosses of the related non-reflexive form: *ponerse* "to put (reflexive)," *volverse* "to turn (reflexive)," *quedarse* "to remain (reflexive)," and *hacerse* "to make (reflexive)." Despite their etymological differences, all four verbs are now used with adjectives with a sense of becoming, for instance *ponerse nervioso* means "to get nervous" and *quedarse sorprendido* means "to be(come) surprised." The problem for the linguist and the learner of Spanish is which verb to use with which adjective. We hoped that a corpus study might shed some light on this question.

We studied these four constructions as used with animate subjects in a spoken corpus of 1.1 million words and a written corpus of just under one million words. One important finding of the study was that very few adjectives occurred with more than one verb. That is, the verbs were not used very often to give special nuances of meaning to the description of the process of becoming. Rather, verb + adjective combinations seemed to be conventionalized. For two of the verbs (*quedarse* and *ponerse*) a very interesting pattern emerged from the data. Certain combinations of verb + adjective had a high token frequency, i.e. they seemed to be conventionalized pairings or prefabs (Erman & Warren, 2000). Here are some examples:

(1) <i>quedarse solo</i>	to end up alone
<i>quedarse quieto</i>	to become still, quiet
<i>quedarse sorprendido</i>	to be surprised
<i>quedarse embarazada</i>	to get pregnant
<i>ponerse nervioso</i>	to get nervous
<i>ponerse furioso</i>	to get angry
<i>ponerse pesado</i>	to become annoying

These pairings seem somewhat arbitrary in the sense that, for example, it is difficult to find a semantic feature that characterizes all the collocations listed here with *quedarse*. Many authors have tried to find general features that characterize all the adjectives used with one verb, but these analyses have not been completely successful. Because each of these collocations occurred at least five times in the corpus, we took the view that they were conventionalized and stored in memory.

The other property of the corpus data that supported this view is that there were many single instances of verb + adjective combinations that could be grouped around these prefabs as semantically similar. Thus,

along with 28 instances of *quedarse solo* “to end up alone” we found three instances of *soltera* “single, unmarried,” two of *aislado* “isolated,” and one each of *a solas* “alone” and *sin novia* “without a girlfriend,” all used with *quedarse*. No adjectives with a sense of “alone” were used with any other verb. Thus, we argue that the conventionalized prefab *quedarse solo* serves as the central member of a set of tokens with similar meanings that also use *quedarse*.

The same type of pattern emerged for all the collocations or prefabs listed in (1). Here is another example: *ponerse furioso* occurred five times and *agresivo* “aggressive,” *bravo* “angry, aggressive,” and *enojadísimo* “very angry” occurred once with *ponerse*. In this case, *ponerse furioso* serves as the central member of the category.

What we propose is that the tokens that are more frequent have stronger representations in memory and serve as the analogical basis for forming novel instances of the construction. These novel instances are also represented in memory (although not so robustly) and may have an effect on the category, causing it to extend in new directions.

Note how the distribution in the corpus mirrors the distribution that Casenhiser and Goldberg (2005) found to be most effective in teaching children a novel construction. Their most effective condition had the construction used with several different verb types, but with one type repeated several times and the others presented only once. If one were learning the use of *quedarse* + adjective or *ponerse* + adjective from natural exposure, this is exactly the pattern that would be presented—certain instances would be of high frequency and other, semantically related instances of low frequency.

This convergence suggests that children’s learning patterns and adults’ usage patterns are both traceable to the nature of categorization: members of categories that are more frequent in the environment are taken to be the central members of categories; they are more accessible and in many cases more generalized in their meaning. When children are forming categories, frequent exposure helps them to set up the category with a central member. When adults are using language, the more accessible, more generalized member of the category is used more often.

So far the discussion has focused on two of the four verbs studied. The constructions with the other two verbs did not show the same type of categorization as that just discussed. In the case of *volverse* there is one very high frequency collocation, *volverse loco* “to go crazy,” but in the corpus we did not find that it had spawned other semantically related collocations. *Volverse loco* occurred 16 times in the corpora, but the other 14 adjectives that occurred once each did not form a semantically coherent set. *Hacerse* did not have any high-frequency uses (with animate subjects—it certainly has some conventionalized uses with inanimate subjects, such as *hacerse tarde* “to get late”). It was used with 18 different

adjectives and none more than three times. These adjectives form a miscellaneous set, e.g. *aburrido* “boring,” *cursi* “tacky,” *consciente* “aware of,” etc. Some of the adjectives used with *hacerse* were judged to be similar in an experiment we conducted. These were the positive adjectives *rico* “rich,” *bueno* “good,” *famoso* “famous,” and *fuerte* “strong,” which are not synonyms, but may be used together in some contexts. I do not know why the constructions with these two verbs behave differently from the first two we discussed. It appears that the latter two verbs may be used when the stronger patterns with *quedarse* and *ponerse* are not available, that is, with adjectives that do not fit into the more coherent categories. It is worth noting that the tokens that fit into the categories used with *quedarse* and *ponerse* constituted 87 percent of the 453 tokens found in the corpora.

The implications for L2 acquisition seem clear. Efficient learning should occur when the central members of categories are presented early and often. The extensions of the categories should also be modeled, but the learner’s natural ability to analogize can be relied on to help him/her produce combinations not experienced before, just as native speakers do. However, this method depends upon good corpus analyses of constructions. Without good information about the type and token frequency of constructions in natural language use, this method cannot be implemented with confidence.

## 5 Grammaticization

Another linguistic phenomenon that is highly usage-driven is the process of grammaticization (or grammaticalization). Grammaticization is a gradual process taking place in all languages at all times; it is the process by which new grammatical morphemes within grammatical constructions are developed out of lexical items or combinations of grammatical and lexical items (Meillet, 1912/1958; see Hopper & Traugott, 2003). Thus, the English perfect tenses, such as *have done*, *had done* developed from the possessive verb plus the past participle; the future auxiliary *will* was once a verb meaning “to want”; *be going to*, used as future, came from the progressive phrase with *go* and *to*. In French and Spanish, the auxiliary *habere* Latin “to have” after an infinitive form of the verb gives rise to the future and conditional suffixes, as in Spanish *cantaré* from *cantar* + *he*. Cross-linguistic studies have shown that grammaticization is the major force that creates grammar and the processes involved are very similar across all languages (Bybee et al., 1994).

As lexical items become grammatical markers, we see huge increases in frequency, since grammatical elements are much more frequent than lexical ones. This frequency increase is instrumental in many of processes that drive grammaticization. On the phonetic level, grammaticizing phrases

undergo massive reduction as the articulatory gestures are reduced and overlap. This produces the variation we find in *be going to*, which we mentioned before is pronounced as [gənə], or in the case of *I'm gonna* we often hear an even more reduced form [aimənə]. High-frequency words and phrases tend to reduce more than low-frequency ones, and material undergoing grammaticization usually reduces quite radically (Bybee, 2002b, 2003). As mentioned above, the reason is that neuromotor sequences that are repeated come to be processed as a single unit. The individual gestures are reinterpreted as part of a single routine, leading to the overlap and reduction of the movements. Interestingly, during the process of grammaticization a great deal of phonetic variation is tolerated.

On the semantic side, grammaticizing phrases are said to become semantically bleached—that is, they lose specific features of their earlier meaning. Thus, *have* in the perfect construction does not signal possession; *will* as a future marker shows only faint traces of its earlier meaning of volition (such traces remain in contexts such as *Give them the name of someone who will sign for it* or *If you'll help, we can finish sooner* where willingness is signaled). *Be going to* as a future no longer signals movement in space. Some of this bleaching is due to habituation—the more an item is used, the less force it has.

The other important source of meaning change in grammaticization is change by pragmatic inference (Traugott, 1989). When we communicate with one another, we are constantly making inferences about what the other party means. If a certain inference occurs frequently with a particular construction, then that inference can become conventionalized as part of the meaning of the construction. Thus, if we hear *I'm going to mail this letter* as a person walks out the door, the basic meaning might be movement in space, but a strong inference is that the person intends to mail the letter. The repetition of this inference with *be going to* + verb can lead to the meaning of intention as part of the meaning of this construction. This is apparently what happened around the time that Shakespeare was writing and that started the course of development that led to the intention and future meaning that the construction has today (Hopper & Traugott, 2003).

On the syntactic level, grammaticizing items change category and this affects the constituent structure. The main verb *willan* eventually became the auxiliary *will*; the main verb *have* also became an auxiliary in the perfect construction. *Be going to* has also changed from being a main verb. It will not take on the properties of the established auxiliaries, but it is already functioning like an auxiliary because it does not affect the selectional restrictions between the subject and the now main verb. Thus, entities that are immobile can be the subject of *be going to*, as in *That tree is going to need more fertilizer soon*. The change from main verb status also implies a change from a complex clause with two main verbs to a simple

clause with an auxiliary and a main verb. This downgrading in the status of the first verb is related to its loss of meaning.

Grammaticization helps us understand the ranges of variation that we see in both meaning and form and it helps us understand why languages have grammar. However, the variation found in phrases or constructions that are undergoing grammaticization present a real challenge to the L2 learner. On the one hand, such phrases are of very high frequency in the input, which helps with their acquisition. On the other hand, the more grammaticized a form becomes, the less salient it is in running discourse and the more redundant it is given the construction it appears in. In fact, highly grammaticized forms such as agreement markers, tense, aspect, and case inflections, have such a reduced communicative value that they seem to remain in the language largely for convention's sake. Thus, if an L2 learner's main goal is to communicate rather than assimilate the conventions of the society, such items seem dispensable, as evidenced by the fact that many fairly proficient L2 learners in a natural setting fail to use many of the grammatical markers of the target language (see Ellis, 2006b).

## 6 Low frequency: prefabs

Conventionalized sequences do not have to be of high frequency. Apparently, a few repetitions is enough to establish the conventionality of a sequence for native speakers. For instance, the prefabricated sequence *experience delays* is not very high in frequency, but it is the conventionalized way of expressing a certain notion. Pawley and Syder (1983), Schmitt (2004), Wray (2002) and others have noted the importance of prefabs for success with a second language. As Pawley and Syder point out, selection of pre-formed word sequences provides not only native-like word combinations but also enables native-like fluency since words are selected more efficiently in groups rather than individually.

From a theoretical point of view, prefabs represent an important argument for usage-based grammar. In order for a native speaker to know that a certain sequence of words is a prefab, she or he must remember that they have been experienced before. This means that even completely predictable word combinations such as *experience delays* or *choose one's words carefully* are registered in memory (Bybee, 1998, 2006; Erman & Warren, 2000). This fact points to a highly redundant, experienced-based cognitive storage and access system for language. This system must intertwine grammar and lexicon, as constructions contain both grammatical and lexical material. Such a theory stands in rather dramatic contrast to the generative proposals of an abstract grammatical system with a redundancy-free lexicon.

A number of studies have demonstrated that prefabs are quite common in both written and spoken language. Erman and Warren (2000) count



word choices in natural texts, both written and spoken, and find that 55 percent of word choices are determined lexically by occurrence in prefabs. Given the high level of use of prefabs in constructing native-like utterances, it makes sense to suggest that the learning of prefabs should constitute an important part of the SLA curriculum. The lexical approach, originated by Lewis (1993) and also described in Richards and Rodgers (2001, Chapter 12), focuses on these conventionalized word sequences as the center of SLA. If embedded in an approach that also teaches general morphosyntactic constructions, attention to prefabs seems quite appropriate (Biber, Conrad, & Cortes, 2004; Simpson & Ellis, 2005; Simpson & Mendis, 2003).

## 7 The role of the first language

As Usage-Based Theory views linguistic knowledge as a set of automatized patterns which are schematic to varying degrees, the first language must be viewed as both a help and a hindrance to Second Language Acquisition. To the extent that the constructions in the second language are similar to those of the first language, the L1 constructions can serve as the basis for the L2 constructions, with only the particular lexical or morphological material changed. However, since even similar constructions across languages are likely to differ in detail, the acquisition of the L2 pattern in all its detail is hindered by the L1 pattern (see Odlin, this volume, Robinson & Ellis, this volume).

Despite these difficulties, Usage-Based Theory would predict that with sufficient input and practice, any morphosyntactic pattern can be acquired, though specific productive lexical distributions will present persistent problems. Of course, in addition to exposure, motivation and ability, both of which differ from one situation to the next, are required. The child learner acquires language as part of learning how to be a human being in the culture in which he/she is being raised. Children assimilate the language, gestures, postures, facial expressions of their culture in minute detail. The L2 learner already is a human being and is trying to communicate. He or she may or may not want to fit into the target culture. In my view, this factor is a strong determinant of the degree to which the grammar of the target language is successfully acquired.

## 8 Usage and pedagogy: controlling frequency in the input

In the foregoing I have pointed out the numerous ways that experience with language impacts its cognitive representation. In some ways this is bad news for the L2 learner, because it is difficult for the adult learner or the classroom learner to get completely native-like exposure to the target

language. In addition, many adults lack the plasticity needed to set up native-like neuromotor routines for the new language. On the other hand, there is some good news. The natural frequency distributions do not have to be exactly reproduced in the classroom. In natural frequency distributions, the high-frequency items are much higher in frequency than they would have to be to be learned well and any language material the learner is exposed to will naturally contain this high-frequency material. Exposure to the lower-frequency constructions and prefabs, however, will take more effort.

The other piece of good news for the L2 learner that comes from Usage-Based Theory is that the analogical mechanisms that we use every day to produce and decode language are just what is needed to use the L2 productively. Similarly, the categorization mechanisms that we use every day for language, as well as for non-linguistic categorization, are available for use in the task of learning a new language. The only requirement is sufficient exposure to the categories of the L2. And finally, the chunking and automatization processes needed to gain fluency occur naturally with practice of both linguistic and non-linguistic tasks.

A great advantage of Usage-Based Theory over Generative Theories is that it does not rely on innateness to explain linguistic categories but rather proposes that much of grammar can be explained on the basis of the domain-general abilities of humans that were mentioned in the previous paragraph. Given these very generalized cognitive abilities, usage factors themselves become part of the explanation for the properties evident in human language.

### Bibliography

- Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language*, 44, 568–591.
- Bayley, R. (1994). Interlanguage variation and the quantitative paradigm: past tense marking in Chinese-English. In E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 157–181). Hillsdale, NJ: Lawrence Erlbaum.
- Bayley, R. (1996). Competing constraints on variation in the speech of adult Chinese learners of English. In R. Bayley, & D. R. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 97–120). Amsterdam: John Benjamins.
- Biber, D., Conrad, S., & Cortes, V. (2004). "If you look at . . .": Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25, 371–405.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: investigating language structure and use*. New York: Cambridge University Press.
- Biber, D., & Reppen, R. (2002). What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition*, 24(2), 199–208.
- Boas, H. C. (2003). *A constructional approach to resultatives*. Stanford: CSLI.

- Bod, R., Hay, J., & Jannedy, S. (Eds.) (2003). *Probabilistic linguistics*. Cambridge, MA: MIT Press.
- Bybee, J. L. (1985). *Morphology: a study of the relation between meaning and form*. Amsterdam, Philadelphia: John Benjamins.
- Bybee, J. L. (1991). Natural morphology: the organization of paradigms and language acquisition. In C. Ferguson, & T. Huebner (Eds.), *Second language acquisition and linguistic theory* (pp. 67–91). Amsterdam: John Benjamins.
- Bybee, J. L. (1998). The emergent lexicon. *Chicago Linguistic Society*, 34, 421–35.
- Bybee, J. L. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.
- Bybee, J. L. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14, 261–290.
- Bybee, J. L. (2002a). Sequentiality as the basis of constituent structure. In T. Givón & B. F. Malle (Eds.), *The evolution of language out of pre-language* (pp. 109–132). Amsterdam: John Benjamins.
- Bybee, J. L. (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. L. (2006). From usage to grammar: the mind's response to repetition. *Language*, 82, 711–733.
- Bybee, J. L., & Brewer, M. A. (1980). Explanation in morphophonemics: changes in Provençal and Spanish preterite forms. *Lingua*, 52, 201–242.
- Bybee, J. L., & Eddington, D. (2006). A usage-based exemplar model approach to Spanish verbs of “becoming.” *Language*, 82, 323–355.
- Bybee, J., & Hopper, P. (Eds.) (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.
- Bybee, J. L., & Moder, C. L. (1983). Morphological classes as natural categories. *Language*, 59, 251–270.
- Bybee, J. L., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar: tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.
- Bybee, J. L., & Thompson, S.A. (1997). Three frequency effects in syntax. *Berkeley Linguistics Society*, 23, 378–388.
- Casenhiser, D. & Goldberg, A. E. (2005). Fast mapping of a phrasal form and meaning. *Developmental Science*, 8(6), 500–508.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Croft, W. (2001). *Radical construction grammar: syntactic theory in typological perspective*. Oxford: Oxford University Press.
- De Saussure, F. (1915/1966). *Course in general linguistics*. New York: McGraw-Hill.
- Eddington, D. (2000). Spanish stress assignment within the analogical model of language. *Language*, 76, 92–109.
- Ellis, N. C. (1996). Sequencing in SLA: phonological memory, chunking and points of order. *Studies in Second Language Acquisition*, 18, 91–126.
- Ellis, N. C. (2002). Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188.
- Ellis, N. C. (2006a). Language acquisition as rational contingency learning. *Applied Linguistics*, 27, 1–24.

- Ellis, N. C. (2006b). Selective attention and transfer phenomena in SLA: Contingency, cue competition, salience, overshadowing, blocking and perceptual learning. *Applied Linguistics*, 27, 164–194.
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20, 29–62.
- Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: the case of *let alone*. *Language*, 64, 501–538.
- Goldberg, A. E. (1995). *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends in Cognitive Science*, 7, 219–224.
- Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics*, 15, 289–316.
- Guillaume, P. (1927/1973). The development of formal elements in the child's speech. In C. A. Ferguson and D. I. Slobin (Eds.), *Studies of Child Language Development* (pp. 240–251). New York: Holt, Rinehart and Winston.
- Hay, J., & Baayen, H. (2002). Parsing and productivity. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 2001* (pp. 203–235). Dordrecht: Kluwer Academic.
- Hopper, P., & Traugott, E. (2003). *Grammaticalization* (2nd ed.). Cambridge: Cambridge University Press.
- Kotovsky, L., and Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development*, 67, 2,797–2,822.
- Krott, A., Baayen, R. H., & Schreuder, R. (2001). Analogy in morphology: modeling the choice of linking morphemes in Dutch. *Linguistics*, 39, 51–93.
- Lakoff, G. (1987). *Women, fire and dangerous things*. Chicago: University of Chicago Press.
- Langacker, R. (1987). *Foundations of cognitive grammar*, Vol. 1: *Theoretical prerequisites*. Stanford: Stanford University Press.
- Langacker, R. (2000). A dynamic usage-based model. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 1–63). Stanford: CSLI.
- Lewis, M. (1993). *The Lexical Approach*. Hove: Language Teaching Publications.
- Long, M. H. (2003). Stabilization and fossilization in interlanguage development. In C. J. Doughty & M. H. Long (Eds.), *Handbook of Second Language Acquisition* (pp. 487–536). Oxford: Blackwell.
- MacWhinney, B. (1978). *The acquisition of morphophonology*. *Monographs of the Society for Research in Child Development*, no. 174, Vol. 43.
- Mańczak, W. (1980). Laws of analogy. In J. Fisiak (Ed.), *Historical morphology* (pp. 283–288). Berlin: Mouton.
- Meillet, A. (1912/1958). L'évolution des formes grammaticales. *Scientia* 12, No. 26, 6. Reprinted in Meillet 1958, *Linguistique historique et linguistique générale*. Paris: Champion.
- Moder, C. A. (1992). Productivity and categorization in morphological classes. Ph.D. diss., SUNY at Buffalo.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–226). Longman: London.

- Richards, J. C. and T. S. Rodgers. (2001). *Approaches and methods in language teaching*. (2nd ed.). Cambridge: Cambridge University Press.
- Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs: implicit rules or parallel distributed processing? In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: explorations in the microstructure of cognition* (pp. 216–271). Cambridge, MA: MIT Press.
- Schmitt, N. (Ed.) (2004). *Formulaic sequences*. Amsterdam: John Benjamins.
- Simpson, R., & Ellis, N. (2005). *An Academic Formulas List (AFL): extraction, validation, prioritization*. Paper presented at the Phraseology, Louvain-la-Neuve, Belgium.
- Simpson, R., & Mendis, D. (2003). A corpus-based study of idioms in academic speech. *TESOL Quarterly*, 3, 419–441.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Skousen, R. (1989). *Analogical modeling of language*. Kluwer: Dordrecht.
- Tottie, G. (1991). Lexical diffusion in syntactic change: frequency as a determinant of linguistic conservatism in the development of negation in English. In D. Kastovsky (Ed.), *Historical English syntax* (pp. 439–467). Berlin: Mouton de Gruyter.
- Traugott, E. C. (1989). On the rise of epistemic meanings in English: an example of subjectification in semantic change. *Language*, 65, 31–55.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.