CHAPTER 32

USAGE-BASED THEORY

JOAN L. BYBEE

CLAY BECKNER

32.1 Statement of Goals of the Theory

Usage-based theory takes language to be an embodied and social human behavior and seeks explanations in that context. As the name indicates, this theoretical perspective incorporates the basic insight that usage has an effect on linguistic structure. It thus contrasts with the generative paradigm's focus on competence to the exclusion of performance and rather looks to evidence from usage for the understanding of the cognitive organization of language. Thus usage patterns, frequency of occurrence, variation, and change are all taken to provide direct evidence about cognitive representation. No relevant methods for gaining evidence about language are excluded; studies of corpora, large and small, diachronic data, psycholinguistic experiments, cross-linguistic comparison and child language development all provide essential data for constructing a comprehensive theory of language.

32.2 Background

Usage-based theory has its source in a confluence of a variety of research perspectives which consider the effect that usage might have on linguistic representation.[1] One practice that unites many of these researchers is a methodological one: it is common now to address theoretical issues through the examination of bodies of naturally-occurring language use. This practice has been in place for decades in the work of those who examine the use of grammar in discourse with an eye toward determining how discourse use shapes grammar, notably Givón, Thompson, Hopper, and DuBois (e.g., DuBois 1985; Givón 1979; Hopper and Thompson 1980; Ono et al. 2000; Thompson and Hopper 2001). In addition, researchers in sociolinguistic variation, such as Labov, Sankoff, and Poplack (e.g., Labov 1972; Poplack 2001; Poplack and Tagliamonte 1999, 2001; Sankoff and Brown 1976) have always relied on natural discourse to study the inherent variation in language use.

Usage and text-based research, always central to traditional historical linguistics, is especially emphasized in functionalist work on grammaticalization, e.g., Bybee (2003*a*, 2003*b*), Hopper and Traugott (2003), and Poplack and Tagliamonte (1999). In fact, the study of grammaticalization has played a central role in emphasizing the point that both grammatical meaning and grammatical form come into being through repeated instances of language use (see section 32.7.3).

Of course, one major impetus for the shift to analysis of natural language use is the recent availability of large electronic corpora and means for accessing particular items and patterns in such corpora. Through the work of corpus linguists, such as John Sinclair (1991), computational linguists, such as Dan Jurafsky and colleagues (e.g., Jurafsky et al. 2001; Gregory et al. 1999), and those who are proposing probabilistic or stochastic grammar, such as Janet Pierrehumbert (e.g., 2001), Rens Bod (1998; this volume), access to the nature and range of experience an average speaker has with language is now within our grasp. Studies of words, phrases, and constructions in such large corpora present a varying topography of distribution and frequency that can be quite different from what our intuitions have suggested. In addition, the use of large corpora for phonetic analysis provides a better understanding of the role of token frequency as well as specific words and collocations in phonetic variation.

At the same time a compatible view of language acquisition has been developing. The uneven distribution of words and constructions in speech to children is mirrored somewhat in the course of acquisition: children often produce their first instances of grammatical constructions only in the context of specific lexical items and later generalize them to other lexical items, leading eventually to productive use by the child; see Tomasello, Lieven, and their colleagues (e.g., Lieven et al. 2003; Tomasello 2003; Savage et al., 2003; Dąbrowska and Lieven 2005).

[1] The term "usage-based" comes from Langacker (1987*b*; 1988); see Barlow and Kemmer (2000).

32.3 Domain-general processes

Usage-based theory postulates that the units and structure of language emerge out of specific communicative events (section 32.4), and strives to avoid relying on innate knowledge specific to the domain of language. A usage-based model thus takes as its null hypothesis the view that language is an extension of other cognitive domains. Elman and Bates (1997: 1,180) write that "language evolved through quantitative changes in social, perceptual, and cognitive abilities, including statistical learning, that exist in other species. These abilities have been recruited for language, but they continue to do nonlinguistic work (that is, they have kept their 'day jobs')". Along these lines, usage-based theory seeks to derive the mechanisms of language from more general and basic capacities of the human brain, including sequential and statistical learning, chunking, and categorization.

32.3.1 Repetition, chunking, and knowledge of usage

A general characteristic of cognition is that repetition of an activity has a cumulative effect on future behavior. In the domain of motor skills—as in learning to ride a bicycle, for instance—an initially deliberate, difficult task can be automatized with practice, eventually becoming an unconscious routine (McCrone 1999). Repetition of an activity causes us to develop "procedural knowledge", that is, implicit knowledge about *how* to do something, in contrast with explicit, declarative knowledge (Anderson 1993). Across domains, learning involves a feedback loop: the human cognitive system produces actions while also monitoring and updating itself on the basis of these actions. With respect to motor activity and other cognitive processes, experiments show that repeatedly engaging in a task leads to the formation of a representation of that process in long-term memory (Shadmehr and Brashers-Krug 1997), and "chunks" the process into useful sub-routines (Simon 1974; Graybiel 1998; Sakai et al. 2004).

In general, it seems that our cognitive systems track any behaviors that keep occurring, improving performance by rendering the activity into chunks that make processing more efficient (Haiman 1994). This principle seems to hold across domains, both for events presented in isolation, and for multiple events that co-occur or occur in sequence. People are quite good at learning when two (or more) events tend to co-occur, or when one event tends to predict another; such abilities are indeed shared by animals other than humans (Kelly and Martin 1994; Bush 2001). A variety of artificial grammar studies in recent years (e.g., Saffran et al. 1996) have demonstrated that people are strikingly skilled at detecting patterns, and inferring units of co-occurrence, based on transitional probabilities, on the basis of relatively little input. Both small children and adults learn such patterns relatively automatically, whether the input consists of language-like syllables, or unlanguage-like stimuli like tones or shapes (Saffran et al 1999; Fiser and Aslin 2002). This pattern detection is a domain-general process of the human mind: we pursue it without conscious effort, and whether or not there is a communicative reason to do so.

Experience thus has an ongoing effect on mental representation. With respect to this broad principle, usage-based theory holds that there is no reason to claim that language is different from any other cognitive domain. A speaker's knowledge of language incorporates a large body of implicit, procedural knowledge, including knowledge of frequency and statistical patterns (Bybee 1998; 2002*a*; Bybee and Hopper 2001*a*; Gahl and Garnsey 2004). A usage-based view holds, further, that there is little reason to claim that knowledge gathered from ongoing experience is fundamentally separate from core knowledge of the language (e.g., "competence" or I-language; see Lightfoot 2006; Newmeyer 2003; 2006 as examples of the generative view).

It would in fact be surprising if experiential knowledge needed, for some reason, to be quarantined from the rest of linguistic knowledge. Anderson (1978: 273) observes that "well-designed systems tend to have special representations for the kinds of information they have to process frequently",[2] drawing examples from visual and auditory processing, and human-designed systems in computer science. The online demands of processing language, both in perception and production, are not trivial. Compared with a static generative model, a system in which mental representations are updated on the basis of incoming information (e.g., a usage-based system) would seem to be more likely to operate smoothly in the face of such demands, and is more in line with what we know about other areas of cognition.

As we will see below (sections 32.5, 32.6, and 32.7), there are further reasons to believe that knowledge of usage is a core part of linguistic knowledge, given that

[2] We wish to be cautious in interpreting the term "well-designed" in the present context. We certainly make no claim that language (or any other cognitive capacity) is externally designed, nor that it is maximally optimized. Our perspective instead is that language is a self-organizing system (Camazine et al. 2001) that exhibits certain apparent "design features". One of these emergent design features is that the grammar is rendered more efficient by encoding frequency information, resulting from a domain-general pattern in which mental representations are updated rather than remaining static.

procedural knowledge is implicated in lexicon and grammar change. Frequency along with other usage-based factors must be incorporated into the grammar, because repetition is necessary to the operation of the common mechanisms of language change (Haiman 1994; Bybee 2006*a*).

32.3.2 Categorization

Categorization represents another domain-general capacity which is of central importance in usage-based theory. We have noted that mental representations are continually shaped by the repetition of events, but for repetition to be recognized, people must sometimes consider two events to be "the same" despite some differences (Haiman 1997). Across domains, categorization allows us to map continuously varied input into "equivalence classes" in some context, on the basis of shared properties (Bruner et al. 1956; Pierrehumbert 2001). For instance, people can quickly learn to classify visual stimuli on the basis of examples they are exposed to in an experimental setting (e.g., Posner and Keele 1968, Medin and Schaffer 1978, Notman et al. 2005), and rapid visual classification of certain complex scenes can occur seemingly effortlessly even without focused attention (Li et al. 2002).

Although we are indeed able to group together input having varied properties, category membership is a gradient, rather than an absolute, phenomenon. There is little evidence for the classical model of categorization, in which categories are defined by necessary and sufficient conditions (Rosch 1978). One category is not sharply defined from the next, but rather the boundaries are gradient (Labov 1973). Moreover, in contrast with the classical theory, categories have an internal structure, and some members are "better members" than others. For instance, with respect to identifying members of the conceptual category "fruit", American participants find that especially good examples are *apple*, *orange*, and *banana*, but less central examples might be *watermelon*, *raspberry*, and *mango*. Such internal category structures become evident via a variety of converging methodologies, including typicality ratings, response times for classification tasks, and the order in which items are listed in a production study (e.g., Battig and Montague 1969; Rosch 1975; Van Overschelde et al. 2004).

In one framework, these category-internal structures are said to derive from relations to a category prototype that encapsulates a central tendency (Rosch 1978). Degrees of category membership then extend outward from the prototype in a network of partially-shared features, resulting in a category with a "family resemblance" structure (Rosch and Mervis 1975). Yet further study has led to the finding that our knowledge of categories cannot just be based on an abstract summary but must include representations for individually experienced tokens.[3] For example,

[3] A full discussion of the merits of prototype vs. exemplar models is not possible in the present chapter; see Medin and Schaffer (1978) and Ross and Makin (1999). The evidence indicates that

Posner and Keele (1968) studied subjects' classification of visual dot patterns under different training conditions in which the central tendency for the category was held constant. Subjects who learned the category on the basis of a more variable training sample were better at classifying noisier variations on the category, compared with subjects who learned under a less variable condition. Since both groups were presumed to have the same category prototype (and the low-variability group should have learned that prototype better), this result is not expected unless learners maintain knowledge about individual exemplars. Another bit of evidence for extensive exemplar storage comes from the finding that people are aware when certain features tend to co-occur *within* a particular category. For instance, people implicitly know that if a bird sings, it is much more likely to be a small bird than a large bird (Malt and Smith 1984). This detailed, intra-category knowledge is not explainable if people only represent the category using an abstract "bird" prototype, while discarding knowledge of individual exemplars.

Evidence such as the foregoing implies that we do not reduce categories to minimal abstractions but rather maintain representations for both coincidental features and highly predictable traits for the category. Moreover, it seems that people retain memories of individual members of a category, since the structure of categories is known to be influenced by the frequency with which particular items are experienced (Nosofsky 1988). All these findings are crucial in usage-based theory, which holds that in language and other domains, specific instances of learning are retained in memory alongside the generalizations that gradually emerge from them (see section 32.4). Moreover, usage-based theory maintains that linguistic categories are just like categories from any other cognitive domain; there is a rich, item-specific internal structure to categories in phonology (Miller 1994), morphology (Bybee and Moder 1983), and grammatical constructions (Goldberg and Giudice 2005; Bybee and Eddington 2006). As we argue below (section 32.4.2), linguistic units are gradient categories that have no fixed properties but rather are formed on the basis of experienced tokens.

32.4 The formal apparatus: Exemplars, networks, and constructions

In contrast to earlier theories that assume limited memory capacity and thus attempt to separate the predictable from the idiosyncratic by representing the latter

exemplar models can fully account for prototype effects, and can explain some findings that are not predicted by prototype models.

in the lexicon and the former in rules, usage-based theory takes a nonreductive, non-minimalist approach to linguistic representation (Bolinger 1976; Langacker 1987*b*, 1988; Booij, this volume). We take into account the extensive evidence that speakers maintain "rich memory representations" in which experiences with language in all their glorious detail are stored in exemplars (Tenpenny 1995; Goldinger 1996; 2000; K. Johnson 1997). In addition to specific exemplars of experienced language, categorization of these exemplars provides more abstract generalizations or schemas. While generative theories emphasize the abstractions, in the current framework we are interested in how the specific experiences speakers have with language combine to yield more general patterns, and how the specific and general interact in acquisition, processing, and language change.

In an exemplar model every token of use impacts cognitive representation. In phonetic perception and decoding, if an input token is the same as an existing exemplar, it is mapped onto that exemplar, strengthening it. If it is not similar enough for a mapping to an existing exemplar, a new exemplar is established, positioned in a metaphorical space close to similar exemplars (Bybee 2001*a*; Pierrehumbert 2001). Thus for every word in a speaker's lexicon, there is a cloud or cluster of phonetic exemplars representing all the phonetic variants of word with information about their linguistic context and further indexes to the social context (Foulkes and Docherty 2006). In speaking, one of these exemplars is chosen for production (Pierrehumbert 2001; 2002). The meaning of the word is also represented by a cluster of exemplars which represent the context and meaning for each token of a word. It is proposed that memory for linguistic objects is the same as for non-linguistic objects, which means that memories can also decay. Particular exemplars that are marginal and not reinforced may be lost, keeping word (and other) categories centered in both their form and meaning (Pierrehumbert 2002; Wedel 2006).

Although every token of experience does affect the system in an exemplar model, we should also take note that not every token produces sweeping change! In fact, as a general rule, in an exemplar model new input either further reinforces an existing pattern, or produces a relatively small change in the system's probabilities. Exemplar models thus provide a framework in which usage-based theory can explain both diachronic and synchronic regularities in language—a necessity for an adequate linguistic theory since we must account for the fact that communities exhibit quasi-stability in speech conventions over time, in addition to the fact that languages *do* change in certain ways. On the one hand, with experience speakers accrue a store of exemplars which may lead to progressively advanced entrenchment via an ongoing production-perception feedback loop. In a population of speakers, stability may be further encouraged by the collective weight of accrued conventions multiplied out over an entire speech community. On the other hand, system equilibrium is anything but inevitable in an exemplar model. New exemplars (involving any combination of phonological, morphosyntactic, or semantic-pragmatic traits) may

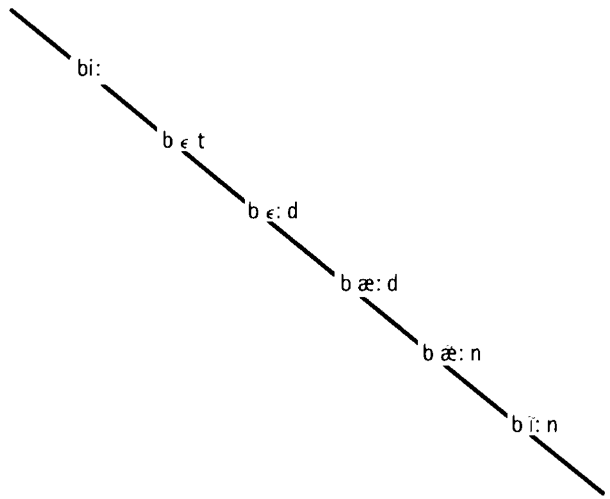

Figure 32.1. Lexical connections for the [b] in *bee, bet, bed, bad, ban, bin*

filter into the system via dialect contact or via the internal mechanisms of reduction, chunking, categorization, analogy and inference discussed below.

32.4.1 Networks

Similarities among words and even longer strings are represented in networks. Through these networks, units of language on various levels emerge. Networks arise through categorization; when tokens of linguistic experience share properties with established exemplars, but also differ in some way, then their shared properties are linked or located close by in mental "space". In the diagrams we will use for illustration—following Bybee (1985*a*) such links are shown as lines—solid lines for identity and broken lines for similarity.[4] These links establish units smaller than the word. Figure 32.1 shows phonological connections; Figure 32.2 shows parallel phonological connections and semantic connections that occur across a number of items; in this case we can speak of affixes emerging from the categorization. Figure 32.3 shows how the internal structure of a complex word emerges through the comparison with related words.

Considerable evidence has been presented in recent work to show that multiword phrases can also be stored in memory. In the case of idioms, which have meanings

[4] The schemas of Booij (this volume) may be interpreted as generalizations that capture the types of relationships expressed by network diagrams. In the network diagram convention, generalizations are understood to be implicit and emergent from the network, rather than being overtly notated in the diagram. (See also Bybee 2001*a*: 22.)

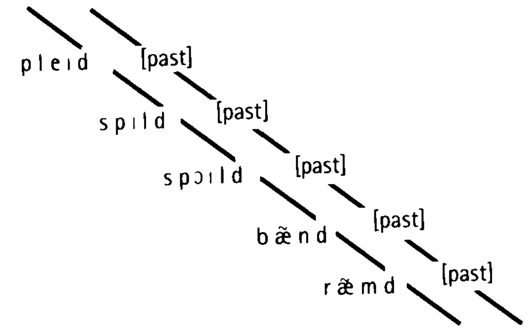

Figure 32.2. Phonological and semantic connections yield Past in *played, spilled, spoiled, banned, rammed*

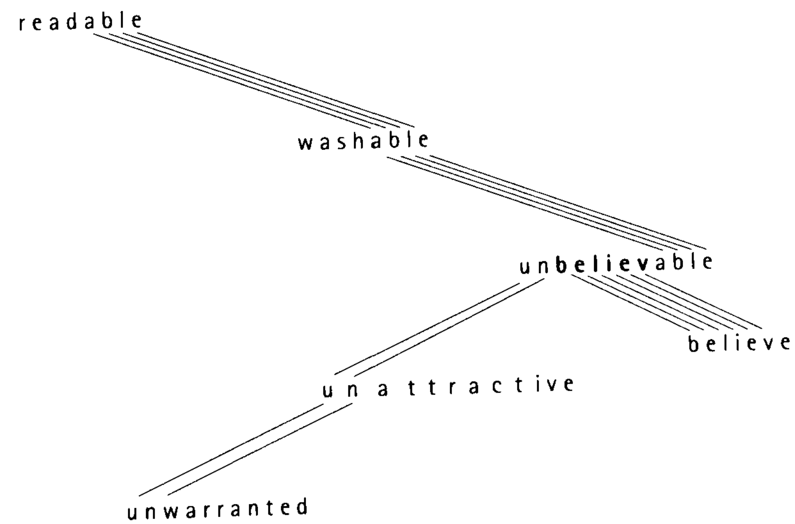

Figure 32.3. The internal structure of *unbelievable* emerges from connections to related words

that are not transparently compositional, such storage is necessary by traditional standards. However, the existence of other collocations, commonly referred to as "prefabs", which do not necessarily have any idiosyncrasies of meaning or form but are conventionalized expressions and known to speakers as such, argue for more extensive storage of multiword sequences (Pawley and Syder 1983; Erman and Warren 2000). Thus for instance, while *pull strings* as in *he pulled strings*

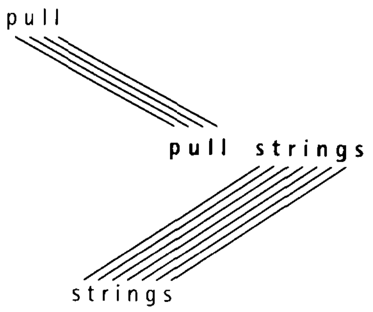

Figure 32.4. The connections between an idiom and its component words

to get that job has a metaphorical meaning, the phrases *for some reason* or *dark night* are transparently compositional in form and meaning and yet represent the conventional way of expressing certain notions. Knowledge about the conventionality of all these sequences must be represented somehow in the grammar, since fluent speakers do not produce (or accept) the full range of utterances permitted by combinatoric syntactic rules. (Compare the non-conventionalized and rather awkward *by some reason, for some cause*, and *black night.*) In the case of idioms and prefabs, their representation in memory does not preclude the speaker knowing what the constituent words are, nor does it preclude access to their meanings and other uses (Nunberg et al. 1994). From a usage-based perspective, there is no need to choose between storage of an unanalyzable unit and compositional assembly, since speakers may in fact have a rich and multifaceted representation for a sequence. A network representation is quite appropriate as it allows access to the sequence as a whole, while maintaining the links that identify the component parts, as illustrated in Figure 32.4.

32.4.2 Units and levels as emergent

All of the units of language—segments, phonemes, morphemes, words, phrases, constituents—can be arrived at by the simple categorization process described above. They do not have to be postulated as a part of the innate universal grammar because they can be arrived at by speakers based on the input and the domain-general process of categorization. The strings of linguistic material that are experienced by the learner are stored in memory (perhaps imperfectly at first) and the brain automatically searches for similarity among such stored experiences, placing them in networks based on these similarities. Whatever repeated units occur in

the experience of the learner will emerge in the networks. Thus if the child hears *ice cream* in different linguistic contexts, such as *I like ice cream* and *do you want some ice cream*, the string *ice cream* will emerge from comparisons of similarity. Note that the speaker/learner is also registering in memory the extra-linguistic contexts in which the linguistic material occurs; in this way semantic and pragmatic representations are also set up. Thus, given certain constraints (such as token and type frequency, see section 32.5), the learner will find the regularities that occur in the input.

This theory raises the question of why languages have units such as segments, affixes, stems, words, and constructions. Rather than postulating such units as givens (innate in the language learner), usage-based theory leaves open the possibility of actually explaining why languages have such units and how they differ across languages (see Lindblom et al. 1984 for an early expression of this view). This explanation will look to diachronic processes to explain current language states. It is worth noting that the postulation of linguistic units as innate universals does not stand up well given the real facts of language, which show, as we will see in the next section, that distinctions between unit types are blurred by both gradience and variability. These facts indicate that dynamic processes rather than static universals are creating regularities.

32.4.3 Gradience

Gradience refers to the fact that the boundaries of many categories of grammar are difficult to distinguish, usually because change occurs over time in a gradual way, moving an element along a continuum from one category to another. Continua such as those between function words and affixes, between regular and irregular patterns, and between productive and unproductive constructions illustrate this gradience.

To demonstrate how the exemplar cum network representation allows for gradience, let us consider some examples, starting with morphemes. Morphemes are traditionally considered to be form–meaning pairings, but problems with the premise that all strings are exhaustively dividable into morphemes have been long noted in the literature. Two types of problems occur. First, dividing words into morphemes sometimes yield leftover bits that are not themselves morphemes. Dubbed "cranberry" morphs by structural linguists, *cran* is one of them because *berry* is obviously a morpheme, but what is *cran*? Other examples are the *Tues-* and *Wednes-* of the days of the week, where the morpheme *-day* is recognizable but the front part of the word is not. These are not problematic for the network model because whole words are stored in the lexicon and there is no requirement that all parts of a word be connected to some other word. Thus *berry* and *day* can have their links while the other parts of the word can be unconnected (Bybee 1988*a*).

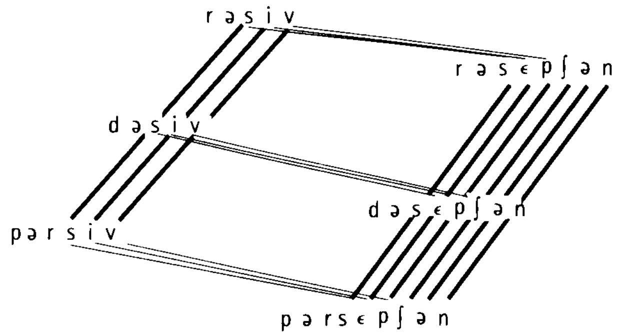

Figure 32.5. How recurring word parts are recognized

Second, some parts of words appear to be repeated across words, but they lack a discernible meaning. For instance, *-ceive* (with its alternate, *-cep-*) occurs in words such as *receive, reception; deceive, deception; conceive, conception; perceive, perception*, revealing a minor pattern of which most speakers probably are aware, yet this Latin stem has no meaning in English. Thus the category of "morpheme" shows gradience in that formal bits of language exist that are recognizable but not fully meaningful. In the network, the phonological associations are made, as in Figure 32.5, but no semantic connections are made, except those between the base and the derived form, and those relating to the categories of verb and noun.

Another gradient dimension in morphology ranges between regular vs. irregular morphological formations. In some theories (generative theories and Pinker's (1991) dual processing model) regulars and irregulars are thought to constitute discrete types and to be processed in totally different ways. In the usage-based model, differences arise in the way complex words are processed due to differences in token and type frequency (see section 32.5). One argument for not drawing a strict line between regulars and irregulars is that there can be substantial overlap between the two types. Thus McClelland and Patterson (2002) point out that many so-called irregular verbs in English have the *t/d* suffix of the regulars, e.g., *slept, brought, went*, etc.[5]

Other instances of gradience in grammar concern the degree of grammaticalization, which of course changes over time (see Heine and Narrog, this volume) and which gradually moves units from independent words, to function words, to affixes. The gradualness of linguistic change means that at any given moment in a synchronic grammar, there will not only be variation, but also gradience in the sense that some units will not fall squarely into the linguist's categories of word, clitic, or affix.

[5] For a presentation of the usage-based approach as applied to phonological phenomena, see Bybee (2001*a*).

32.4.4 Larger units

Much of our subsequent discussion will focus on the gradient properties of larger syntactic units—in this framework, constructions, which we will discuss more thoroughly in sections 32.6 and 32.7. Constructions are conventionalized sequences of morphemes or words that contain a position that can be filled by more than one item. Consider, for example, the expression *drive someone crazy, mad, insane, nuts, up the wall*, etc. This is a construction that contains the specific verb *drive* (in any of its inflected forms), an object pronoun, and an adjective denoting a state ranging from true insanity to extreme irritation. Such a construction can emerge from a network via exposure to specific tokens. *Drive* + object pronoun is the anchor, i.e., the most stable part of the expression, and the adjective slot is more open, though it is semantically constrained. This analysis is arrived at again by categorization based on similarity of form and meaning for *drive* and meaning only for the adjectives.

Like the other units we have discussed, constructions exhibit both gradience and variation, since they can vary in their degree of grammaticalization, productivity, schematicity and their appropriate contexts of use, as we will see below.

32.5 The Role of Repetition: Effects and Mechanisms

We have already seen that exemplar models register variation and change while it is ongoing. In addition, exemplars are strengthened by repetition, so that frequency is naturally represented in cognition. The network of connections is also sensitive to frequency of use. In this section we consider both TOKEN FREQUENCY—the number of times an item or string occurs in running text, and TYPE FREQUENCY—the number of distinct items that can be used in a pattern.

32.5.1 The reducing effect of high-token frequency

An extensive body of literature has shown that high-frequency words and phrases undergo phonetic reduction at a faster rate than low- and mid-frequency sequences (Schuchardt 1885; Fidelholtz 1975; Hooper 1976; Bybee and Scheibman 1999; Bybee 2000*b*; 2001*a*). This REDUCING EFFECT applies to phrases of extreme high frequency such as *I don't know*, which shows the highest rate of *don't* reduction (Bybee and Scheibman 1999), and also to words of all frequency levels undergoing gradual sound change, such as English final t/d deletion or Spanish [ð] deletion, both of which affect high-frequency words earlier than low-frequency words (Bybee 2001*a*; 2002*b*; Gregory et al. 1999). This effect of repetition is the result of the domain-general processes discussed in section 32.3.1. Words and phrases represent "chunks" of neuromotor behavior. With repetition, their execution becomes more fluent as the articulatory gestures involved reduce in magnitude and overlap adjacent gestures (Pagliuca and Mowrey 1987; Browman and Goldstein 1992).

32.5.2 Entrenchment and autonomy: The conserving effect of high token frequency

Alongside the Reducing Effect, words and phrases with high-token frequency are also subject to the CONSERVING EFFECT, meaning that high-frequency items are more resistant to reformations based on productive patterns in the language. These two effects may seem paradoxical at first glance, but they are caused by two different cognitive mechanisms which respond to token frequency: in addition to increasing fluency, high token frequency has the effect of strengthening memory representations. This strength is reflected in easier lexical access and, in complex words and strings, resistance to reformation. For any given string that consists of more than one meaningful element, there can be at least two ways of accessing it: either as a single unit or as a set of units that are then combined into a whole. For instance, the word *insane* can either be accessed as a unit, or it can be built up by combining the prefix *in-* with the stem *sane* (Hay 2001).[6] The higher the token frequency of the sequence, the more likely it will be to be stored and accessed whole (Bybee 1985*a*; but see also Hay 2001). Thus high-frequency sequences are more entrenched in their morpho-syntactic structure and therefore resist change on the basis of more productive patterns. Among English irregular verbs the low-frequency verbs are more likely to regularize (*weep, weeped*) while the high-frequency verbs maintain their irregularity (*keep, kept*). The reason is that frequency strengthens the memory representations of words or phrases, making them easier to access whole and thus less likely to be subject to reformation on the basis of more productive patterns (Hooper 1976; Bybee 1985*a*). This effect applies to syntactic sequences as well, allowing higher-frequency exemplars to maintain a more conservative structure (Bybee and Thompson 1997). In section 32.7.2 we discuss several examples that show the maintenance of the older constructions in high-frequency contexts.

As we said in our discussion of networks, chunks that are stored whole can also maintain their associations with other instances of their component parts. In cases of extreme high frequency, however, a morphologically complex form (or string of words) can lose its internal structure and become autonomous from etymologically related forms (Bybee 1985*a*). This can be seen, for example, in the way that words with derivational affixes become less transparently related to their base forms as they become more frequent (Bybee 1985*a*; Hay 2001). Hay (2001) argues that the semantic opacity of words such as *dislocate*, etc. is due to the fact that their complex forms are more frequent than the bases from which they were originally derived. The effect applies to inflection only in cases of extreme high frequency where it leads to suppletion. Thus *went* was formerly the past tense of *wend* but (for unknown reasons) it increased in frequency and moved away from *wend*, joining *go* to become the past tense of that verb. This effect also applies in grammaticalization when sequences that are originally complex (such as *be going to*) lose their semantic and syntactic transparency and move away from other instances of the Progressive, *go*, and *to*.

[6] Building up strings of morphemes of words vs. accessing them already assembled do not actually constitute two mutually exclusive means of access; rather they represent two poles of a gradient, which is the extent to which the separate components of a string are activated when the whole string is activated.

32.5.3 Type frequency, schematicity, and productivity

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. For instance, the regular English Past Tense inflection with *-ed* applies to thousands of verbs and thus has a very high type frequency. In contrast, the vowel-change pattern exemplified by *string, strung; fling, flung; stink, stunk* applies to some eighteen English verbs and thus has a lower type frequency. Taking a phonotactic example, the word-initial sequence *sp-*, as in *spark, spot, spin*, etc., has a much higher type frequency that the cluster *sf-*, as in *sphinx* and *sphere*.

Type frequency is the main factor that determines the degree of productivity of a construction (Guillaume 1973 [1927]; MacWhinney 1978; Bybee 1985*a*). That is, patterns or constructions that apply to a high 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. For instance, the verbs of the *string, strung* class must end in a nasal or a velar (Bybee and Moder 1983). The open slots in constructions are often semantically restricted, as 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). Thus productivity is a matter of degree, determined by an interaction of type frequency with schematicity—the degree to which the category is open or restricted.

The contribution of type frequency to productivity comes about when a construction is experienced with different items occupying a position, which enables the parsing of the construction (Hay and Baayen 2002). 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. Schematicity contributes to productivity in that highly schematic categories are more easily extended to new items. Since there are no phonological or semantic restrictions on the regular English Past Tense suffix *-ed*, it is free to apply to any verb.

Thus productivity and schematicity are highly related to categorization since the application of a construction depends upon the properties of the category formed for the open position. Both types and tokens contribute to categorization. The *properties* of the types included in a category establishes its range or schematicity while the *number* of types relates to the degree of productivity of the construction referring to the category. In research into exemplar models (in which the category consists of the experienced exemplars), token frequency has been shown to influence the perception of the center of the category, as well as its boundaries (Nosofsky 1988). In phonetic categorization, high-frequency exemplars tend to be maintained while low-frequency ones are marginalized and lost (Bybee 2001*a*; Pierrehumbert 2001). In semantic categorization a similar phenomenon occurs; in a corpus and experimental study of the pairing of verbs meaning "become" with adjectives in Spanish, it was found that the high-frequency pairs served as the center of some of the most productive categories (Bybee and Eddington 2006). Similarly, Casenhiser and Goldberg (2005) show that children and adults learn a new construction faster if they are exposed to one higher-frequency token as well as several types exemplifying the construction.

32.6 Constructions: Form–Meaning Pairings

For the purpose of syntactic description, the usage-based model adopts constructions as the basic unit of form–meaning correspondence (Fillmore et al. 1988; Goldberg 1995; 2006; Croft 2001). We regard any conventionalized string of words or morphemes as a construction, but our focus for an understanding of syntactic productivity is on strings that include at least one schematic position—a position in which more than one word or morpheme may appear. What we regard as the

grammar of a language is a collection of constructions, organized into networks by the same criteria that words are—by their formal and semantic similarity.

An important property of a grammar based on constructions is that it reflects the deep intertwining of lexical items with grammatical structure. Most constructions contain very specific lexical material, such as the verb *drive* in the *drive someone crazy* construction mentioned above, or *-ed* (and its allomorphs) in the regular Past Tense construction. In addition to having fixed linguistic material, most constructions restrict the set of lexical items that can fill the open position, as when *drive someone* ____ must contain an adjective or prepositional phrase meaning "crazy". The fact that a certain lexical item (in this case *crazy* for American English) occurs more often in this slot than any other lexical item is recorded in the exemplar representation as important information for the category of items occurring there. In other words, in an exemplar model constructions are not abstract grammatical patterns but rather they are sets of experienced exemplars arranged in cognitive space to reflect their similarity in form and meaning.

Consider in more detail the *drive someone crazy* construction, as studied by Boas (2003) (cf. a set of "become" constructions in Spanish as analyzed by Bybee and Eddington 2006). This construction uses the verb *drive* with an adjective or prepositional phrase expressing a meaning such as "drive crazy". Particular tokens found in the British National Corpus (BNC) include:

(1) It drives me crazy.

(2) He was going to drive her crazy if she wasn't careful.

(3) That old thing, it's just driving us crazy.

(4) They drive you mad.

(5) The death of his wife the following year drove him mad.

(6) It drove the producer mad.

(7) A couple of channels that used to drive her up the wall...

(8) This room drives me up the wall.

For illustration, the eight tokens represented above could each be considered exemplars which are grouped together with their identical parts mapped onto one another and their schematic parts forming categories as in (9). The adjectives illustrated here are *crazy, mad*, and *up the wall*; the others that occur in the BNC are semantically related to these (see Boas 2003).

(9)

SUBJECT [DRIVE] { *me* / *us* / *you* / *him* / *her* / *the producer* } { *mad* / *crazy* / *up the wall* / . . . }

The category of SUBJECT has not been represented with actual exemplars because it appears to take any NP. Presumably NP is a category that can be developed on the basis of the exemplars that occur in other constructions (Croft 2001). [DRIVE] is a notation intended to show that any inflected form of the verb *drive* may appear, in addition to any of the other auxiliary or emerging auxiliary constructions (e.g., *used to, gonna*...). The enlarged font of [DRIVE] represents the strength it acquires by occurring in all instances of the construction. *Mad* and *crazy* are similarly shown enlarged because of their high frequency in the construction. The experiencer slot is usually a pronoun, but is always animate and usually human. The final position, which can be an adjective or prepositional phrase, has a strong semantic character. Most of the fillers for this slot found in Boas' study of the BNC were synonyms with "crazy", though there were also slightly more distantly related senses such as *to desperation*, or *to suicide*. Note that the category of adjectives and prepositional phrases is not represented as an abstraction but rather by specific items, since these exemplars are retained in memory along with knowledge of their respective frequencies. Novel additions to this category are made on the basis of analogy with existing exemplars. We propose, following the evidence in Bybee and Eddington (2006), that the most frequent members of this category serve as the center of the category; not only are they more likely to be chosen for subsequent productions but they also serve more often than any others as the basis for analogy.

Most of the constructions discussed in the literature are somewhat specific, as is the one discussed here. For this reason, some researchers doubt that a construction-based account can ratchet up to a full account of syntactic phenomena. For instance, Jackendoff (2002) accepts constructions as necessary in a grammar, but in addition maintains phrase structure rules. In contrast, we are confident that there are no empirical data of morphosyntax that cannot be adequately described via constructions and networks of constructions. This is a pressing issue for further research.

As further evidence for exemplar representation of constructions, consider the fact that such representations allow the association not just of form and meaning but also of pragmatic implications and social contexts of use, which we know from studies of change as well as variation are important parts of the knowledge that speakers have about their language (Traugott and Dasher 2002; Torres-Cacoullos 2001). This topic is treated in more detail in section 32.7.

Finally, we note briefly that the construction-and-exemplar framework we have sketched out in this section further fits into a unified usage-based model that incorporates language acquisition. Recent usage-based accounts of acquisition (for instance, Tomasello 2003; Goldberg 2006) view constructions as a basic building block in learning a language, as children learn verbs in the context of particular sequences that pair form and function. Children first comprehend (Roberts 1983; Akhtar and Tomasello 1997) and produce (Tomasello 2000) particular verbs only in highly specific contexts, gradually expanding on these to arrive at more abstract syntactic representations.[7] Dense corpus studies of child–parent interactions also find that children are very conservative learners who are guided by particular exemplars they have learned. Lieven et al. (2003) found that a majority of the utterances (63%) by a two-year-old child consisted of exact repetitions of utterances that occurred earlier in the corpus. Moreover, among the utterances that *were* novel, 74% needed only a single operation (such as adding or removing a word) to match a particular previous utterance, or even a whole class of related utterances that permit a variable slot. The overall picture that emerges is that language learners slowly generalize item-specific sequences to permit open slots, progressively linking these constructions in a network and allowing different constructions to be combined systematically (Tomasello 2000; Dąbrowska and Lieven 2005). Within usage-based theory, there is no need to assume that knowledge about particular items is purged from memory as soon as the language learner forms generalizations (Langacker 1987*b*). Indeed, we have evidence that adult speakers maintain detailed knowledge of the internal structure of constructional categories, including a sensitivity to frequency (Bybee and Eddington 2006).

32.7 Variation and Change

As we have seen, gradience and variability are built into an exemplar model: cognitive representations will reflect any new variants or ongoing changes in the distribution and frequency of variants. In this section we show that exemplar representation of constructions also provides a means to understand the creation of new constructions, the competition between constructions, and the grammaticalization of constructions.

[7] For additional evidence regarding the item-based nature of syntactic acquisition, see Lieven et al. (1997), Wilson (2003), Savage et al. (2003), and Dąbrowska and Lieven (2005), among others.

32.7.1 New constructions arise from specific exemplars of established constructions

The search for explanations for grammar in general and specific constructions in particular takes a diachronic perspective in this framework. If we want to know why a language has a particular feature, it is instructive to examine how it acquired that feature (Bybee 1988*a*). Thus we can take specific constructions and ask how they achieved that status in a particular language.

Consider a construction studied by Kay and Fillmore (1999) and C. Johnson (1997); they call it the WXDY? construction. It is exemplified in the famous joke, shown in (10) (also discussed in Bybee 2006*a*):

(10) Diner: Waiter, *what's this fly doing in my soup?*
Waiter: Why, madam, I believe that's the backstroke.
(From Kay and Fillmore 1999)

The joke shows the ambiguity of the sequence in italics. The usual interpretation of "what is X doing Y?" is one of surprise at incongruity accompanied by more than a hint of disapproval. Because it is syntactically indistinct from the construction from which it arose—a *what* question with *do* in the progressive—it gives the clever waiter license to interpret it as a literal question about what the fly is doing.

Interestingly, there is nothing in the form which explicitly suggests a meaning of incongruity, but the strong implication is nonetheless there. We can ask, then, how did an ordinary Wh-question with *doing* and a locative phrase acquire these implications? The answer must be that these implications arise from language use in context. The question of *what are you doing?* itself often has negative connotations. In a phone conversation, one may legitimately ask an addressee *what are you doing?*, but in a face-to-face situation the answer to the literal question should be available via visual inspection. Thus the question implies that the speaker wants some explanation not just of what the addressee is doing but *why* she or he is doing it. Similarly when this construction has a locative element, as in (11), there is the possibility of ambiguity, but the first reading is probably more common.

(11) What are you doing with that knife = 'why do you have that knife?'
or the literal meaning = 'what are you doing with it?'

The implication of disapproval, which is a subjective interpretation made in context, must have come from multiple instances of use with this negative nuance. As we have pointed out earlier, each exemplar of a morphosyntactic construction includes information about the contexts of use and this would include the inferences made in this context. We know from studies of grammaticalization that inferences can become part of the meaning of a construction (Traugott 1989; see section 32.7.3). The only way this could happen would be if language users were recording in memory the inferences in each situation and, at a point at which certain inferences become strong in certain contexts, they become part of the meaning of a construction.

The important point to note from this discussion is that new constructions arise out of specific exemplars of old constructions (Bybee 2003*b*; 2006*a*). This fact tells us much about how new constructions arise and it also provides evidence that cognitive representations of grammar include specific information about contexts of use of exemplars and their meaning and implications in these contexts.

32.7.2 Old and new constructions compete

Languages quite often have two or more ways of expressing the same or very similar meaning. Consider these examples: some English verbs express Past Tense by vowel changes (*blow, blew; write, wrote*, etc.) while other express the same meaning with a suffix (*chugged, hissed*); sentence negation has two alternate forms in cases where indefinites occur in the clause, for instance, *there was nothing to drink* and *there wasn't anything to drink*; English also has infinitives marked with *to* and unmarked infinitives that occur after modal auxiliaries.

Consider first the English Past Tense. We know that the ablauting process for forming the Past goes back thousands of years in Germanic, while the suffixation process is more recent. Also, it is well known that suffixation, with its high type frequency and productivity, has been gradually supplanting the ablauting process for more than a thousand years. The ablauting verbs that remain in the language are all of fairly high frequency, which is the main factor in their preservation (see section 32.5.2 above). Thus we can conclude that when older and newer constructions exist side by side in a language, it will commonly be the case that the older construction is preserved primarily in high-frequency contexts.

This principle can be applied to syntactic constructions as well. For instance the two ways that negation affects indefinite items within its scope consists of an older and a newer construction. The newer construction is the one with *not* and its contraction, as in *there wasn't anything to drink*. The older construction negates just the indefinites, as in *there was nothing to drink*. In a corpus-based study of cases where these two constructions have the same meaning and implications, Tottie (1991*b*) shows that the older (NEG-incorporation) construction is mostly used with high-frequency constructions such as existential *be* as in (12), stative *have* as in (13), and copular *be* as in (14):

(12) By the time they got to summer there was no more work to do.

(13) The Fellowship had no funds.

(14) As a nation we are not doing well enough. This is no new discovery.

The use of this type of negation with lexical verbs is much less common and tends to center around high-frequency verbs such as *know, do, give*, and *make*. The construction with *not* is much less restricted.

A third example concerns the marking on infinitives in English. Most infinitives use *to* as a marker, but after modal auxiliaries, the infinitive has no marker. Thus we contrast *I want to go* with *I can go*. The unmarked infinitive derives historically from a form with a suffix: Old English marked infinitives with the suffix *-an* and its variants. This suffix was eroded to -ən and later to -ə and then it was completely lost. At the same time, the *to* as infinitive marker had started out in purpose clauses, and was appearing in more and more constructions. However, the construction of modal auxiliary plus infinitive verb was already established in late Old English and had become quite frequent by the time the *to*-infinitive was spreading to more constructions. Because of the conserving effect of token frequency, the *to* has never been able to make its way into the modal auxiliary construction.

32.7.3 Grammaticalization of constructions requires frequency of use

Grammaticalization (see Heine and Narrog, this volume) is a central phenomenon of usage-based linguistics because it is the principal mechanism (or set of mechanisms) by which grammar is created, and it requires language use to take place. As we saw in section 32.7.1, new constructions arise out of exemplars of existing constructions. In grammaticalization, a further step is taken in that a lexical item within this construction takes on grammatical status. A recent example in the history of English is the development of the future marking periphrasis *be going to*. This developed out of a purposive construction meaning "to go somewhere to do something". It is important to note that uses of *go* in other constructions do not grammaticalize into futures. As recently as Shakespeare's time such a construction had its literal meaning. It was just one exemplar—but the most frequent exemplar—of the more general purpose construction exemplified by these sentences from Shakespeare:

(15) Don Alphonso,
With other gentlemen of good esteem,
Are journeying to salute the emperor
And to commend their service to his will. (*Two Gentlemen of Verona* I.3)

(16) ...the kings
and the princes, our kindred, are going to see the queen's picture.
(*Winter's Tale* V.2)

Note that in both (15) and (16) the subjects are actually moving in space. In contemporary English *we're gonna see the queen's picture* can be interpreted simply as expression of future time.

Grammaticalization takes place as language is used. Grammaticalizing constructions make huge gains in token frequency and thus undergo the effects of high token frequency. As argued in Bybee (2003*b*), the changes that take place in grammaticalization are conditioned at least in part by high frequency of use. The following is a brief explanation of how frequency of use helps to condition the changes that took place in this construction. Note that all of these changes are intricately interrelated.

First, as we saw above, phonological reduction takes place when words and phrases are often repeated. Thus the increasing token frequency of *be going to* leads to the creation of a neuromotor routine that is processed as a single unit and can undergo phonological reduction to the form spelled *gonna*. Indeed, the highest frequency expression involving *be going to* is *I'm going to*, which is often produced as [aɪmənə].

Second, the autonomy of a new construction is conditioned by frequency as explained in section 32.5.2. That is, as a particular string grows more frequent, it comes to be processed as a unit rather than by its individual parts. As it is accessed more and more as a unit, it grows autonomous from the construction that originally gave rise to it. It loses its association with the purpose construction and also with the other instances of the verb *go*.

Third, the loss of the specific meaning of movement in space and addition of inferential meaning from the context also relies on frequency of use. The *be going to* construction in many contexts carried the pragmatic inference of intention, as shown in the following exchange from *Two Gentlemen of Verona* as cited in Hopper and Traugott (2003).

(17) Duke Sir Valentine, whither away so fast?
Val. Please it your grace, there is a messenger
That stays in to bear my letters to my friends,
And I am going to deliver them.
(1595, Shakespeare, *Two Gentlemen of Verona* III.i.51)

In this example, the Duke's literal question is "where are you going?" Valentine's answer does not specify location but rather intention. Interestingly, that is actually what the Duke wanted to know. The inference of intention often accompanies the use of this construction. Repeated instances such as this one make "intention" part of the meaning of the construction. The meaning and contextual implications of a construction form an exemplar cluster much as the phonetic variants do. These clusters are susceptible to the same sort of reorganization we have discussed with respect to phonetics: high-frequency semantic/pragmatic exemplars come to dominate the cluster and lower frequency exemplars may be lost, bringing about gradual semantic change.

Example (17) shows how the meaning of "intention" becomes associated with *be going to*; this interpretation is still available today. However, a further inferential change has also taken place: the expression of intention can give rise to the inference of prediction about a future event (see Bybee et al. 1994).

Finally, because items that are used together frequently come to be processed together as a unit, changes in constituency and category can take place. Thus *going to* as the constant part of this construction becomes a single unit not just phonologically but also syntactically. As the construction acquires new nuances of meaning and loses its motion sense, the following verb is taken to be the main verb. This process, known as "reanalysis", is viewed in a usage-based perspective as being gradual, that is, as consisting of a gradual change in the exemplar cluster (Beckner and Bybee 2009; Haspelmath 1998).

Thus the study of grammaticalization provides the explanatory basis for grammar as an emergent phenomenon; it also provides us with an understanding of the semantic categories of grammar and how they evolve, and an explanation for the correspondence between behavioral properties of grammatical elements and their meanings or functions (Bybee et al. 1994).

32.8 Language acquisition and adult-based change

As described in the previous sections, usage-based theory is fundamentally concerned with diachronic change, insofar as language use shapes language structure in an ongoing and dynamic fashion. A usage-based model assigns a central role to usage by *adult speakers* in accounting for language change, in contrast with the traditional generative approach, in which language change is introduced via acquisition across generations, as learners deduce a new grammar on the basis of adult speech (see Halle 1962; Lightfoot 2006, among others). In this section we note the weaknesses in the theory that allows change to occur only in the acquisition process and note the many arguments in favor of the proposal that adults can also change language.

First, a model in which children innovate via imperfect learning is unable to account for known diachronic regularities. As shown in section 32.5.1, high-frequency words and word sequences undergo the greatest degree of phonetic reduction. Such a pattern is fundamentally at odds with an imperfect learning mechanism, which predicts that children will be more likely to change low-frequency items. As we have noted, the reduction of frequent items in fact

arises out of expert fluency, when well-practiced articulatory routines lead to the diminishment and overlap of speech gestures. Such reductive changes may actually result in forms that are articulatorily more complex, and harder to acquire, than the non-reduced forms. For example, consider English contracted auxiliary/negation sequences such as *did not* > *didn't* and *could not* > *couldn't*. This evidence does not point toward young language learners as the originators of the change but rather indicates that reductive change originates in usage-based factors.

Additionally, small children are unlikely to be the instigators of changes involving domains that are cognitively accessible only to older speakers. For instance, complementation is not acquired early by children, and young speakers strongly disfavor the use of overt complementizers (Radford 1990; Adamson 1992; Penke 2001). Historically, complementizers originate from a variety of lexical sources, including demonstratives (e.g., English *that*), dative-allative particles, and the verb meaning "say" (Heine and Kuteva 2002; Hopper and Traugott 2003). It seems unlikely that first language learners would misapprehend these forms as complementizers, given that young children in fact struggle to master complementation as a feature of adult language. Similarly, small children are unlikely to contribute to the grammaticalization of epistemic markers from deontics, since epistemicity cannot be fully acquired until children develop basic social competencies, in addition to attaining cognitive milestones such as a theory of mind (Barbieri and Bascelli 2000; Aksu-Koç and Alici 2000; Resches and Pereira 2007). In general, adults have more sophisticated social and cognitive abilities than children, and they face a broader range of domains in which they must communicate. Adults bring to this task a full set of capacities that can influence language change via usage, including the ability to invite and comprehend conversational inferences (Traugott 1989; see section 32.7.1 and 32.7.3). It is not plausible to assume that adults are stuck with using only the grammatical structures and conventions that children have managed to innovate. More reasonably, adult speakers are capable of extending existing patterns, which with repetition may then lead to new grammatical conventions.

Finally, there is the mismatch between children's innovation and documented diachronic changes that has been often noted in the past. Children often produce words with consonant harmony, while adult languages never have such a process (Drachman 1978; Vihman 1980); children's morphological formations at times reflect possible historical changes, but at times do not (Bybee and Slobin 1982). The fact is that children's innovations typically do not influence language because there is no social mechanism for the propagation of these innovations, given that children copy adults rather than the other way around.

The usage-based model, as described in section 32.3, proposes that even in adulthood our experiences with language continue to affect mental representations, just as in other experiential domains. It is indeed the case that adults are less influenced by new input than children, due to the cumulative effect of past tokens of experience. Moreover, early exposure to language, and early exposure to particular language features, affords learners the greatest opportunity to process language fluently (Morford 2003). However, this is not the same as saying that adult grammar is "frozen" beyond some critical period cutoff. Rather, there is a "sensitive period" that leads to a gradual decline in receptivity, but the system nonetheless never becomes completely static (Newport 1991; Morford 2002).

Contrary to claims that adults cannot adjust their grammar (Newmeyer 1998; Lightfoot 2006), we now have considerable evidence that adults continue to learn across all domains of language. A number of studies have found that speakers can adopt ongoing phonetic, even phonemic, changes in their language, long after the speaker enters adulthood (Harrington et al. 2000; Sankoff 2004; Harrington 2006; Sankoff and Blondeau 2007). Moreover, adults are not just capable of generalizing constructions to new items but must do so to use language productively. Subjects in an experiment by Kaschak and Glenburg (2004) learned an unfamiliar construction (*The meal needs cooked*) and quickly generalized this construction to new verbs. Likewise, Goldberg et al. (2004) found that, with three minutes of training, adult English speakers were able to learn an SOV construction and extend its semantics to new verbs.

In sum, we find that adult speakers are capable of participating in language change, and in some cases, adult speakers *must* be the originators of change. However, this is not to say that acquisition plays no role at all in diachronic processes. Changes such as the regularization or loss of infrequent forms may plausibly be influenced by usage (due to speakers' inability to retrieve weakly-represented variants) and by acquisition (due to children's insufficient exposure to rare variants) (Bybee and Slobin 1982). As we discuss below, a usage-based model considers contributions from multiple interacting factors in an emergentist account of language, and our catalog of language change mechanisms should be inclusive where appropriate.

32.9 Language as a Complex Adaptive System

In the usage-based framework, properties of languages and their grammars are viewed as emergent, i.e., not given a priori, but coming about through language use and the way the brain responds to the experience of language use (Hopper 1987; Lindblom et al. 1984; Larsen-Freeman 1997; Ellis and Larsen-Freeman 2006). Emergence is a feature of complex adaptive systems—systems in which a few causal mechanisms interact iteratively to produce what appears to be structure (Holland 1998; Camazine et al. 2001). Waves on water and dunes of sand are examples: while we perceive the structure in the waves or the dunes of sand, we know that it is not given a priori that waves or dunes should have a certain structure but rather a result of the physical properties of water and sand interacting iteratively over time and space with the bottom of the sea, the wind, and so on. It might also be noted that waves and dunes show much variability and gradience and, while we can recognize them when we see them, it might be difficult to give a firm description of their apparent structure.

We have tried to make the case in this chapter that what we perceive as language structure comes about through the application of a handful of common mechanisms that recur when human beings use language. The domain-general processes of sequential learning, chunking, categorization, and inference-making, along with the effect of partial or complete repetition, lead to the establishment and conventionalization of the categories and structures we find in languages. This bottom-up and emergentist perspective, we argue, may turn out to be indispensable to our understanding of linguistic processes and structure. Here it is helpful to draw a parallel with what is perhaps the best-studied complex adaptive system, namely, biological evolution. In the oft-cited slogan of Theodosius Dobzhansky (1973), "Nothing in biology makes sense except in the light of evolution". To truly understand the modern-day diversity of biological species, it is essential to take note of a range of simple interactions that contribute to causal mechanisms such as natural selection, sexual selection, and genetic drift. In the domain of language, the mechanisms of change are quite different, but in describing linguistic phenomena we must likewise take account of the interaction of simple elements, along with considering diachronic processes as a source of explanations. Ignoring such considerations and defaulting to a nativist, top-down explanation runs the risk that we will overlook important regularities that emerge from diachrony.

As Greenberg has argued (1969; 1978*a*; 1978*b*) the source of structure in phonology and grammar and the explanation for their similarities across languages is the set of diachronic processes that are common cross-linguistically. Commonly-occurring sound changes create phonemic systems and the cross-linguistic markedness patterns they exhibit (Greenberg 1969; and for a more work in this tradition, Bybee 2001*a*; Blevins 2004). The major source of grammatical structure is the set of processes that constitute grammaticalization. A striking characteristic of grammaticalization is that very similar developments take place in different, unrelated languages (Bybee et al. 1994; Heine and Kuteva 2002). For instance, Bybee et al. (1994) found instances of a future marker developed from a movement verb in seventeen languages out of seventy-six languages chosen to be maximally unrelated. So the development of English *be going to* into a future is not an isolated occurrence but rather reflects a very strong tendency across languages. It is possible with these and similar results to construct diachronic paths of change that are cross-linguistically similar though not always identical, perhaps in the way that one sand dune resembles another without being identical to it. An historical approach holds the most promise for explaining the complex patterns we find across the world's languages: absolute synchronic universals are rather rare (Croft 2003; Bybee 2008), but there is indeed patterned variation, in the form of statistical tendencies and recurrent diachronic shifts.

Specific unidirectional paths of change for the grammaticalization of tense, aspect, modality, voice, definites and indefinites, and many other categories have been hypothesized based on both diachronic and cross-linguistic data. While much has been written about such paths of change (Greenberg 1978*a*; Givón 1979; Bybee et al. 1994; Heine and Kuteva 2002), in our theoretical perspective, they are not at all the end of the story. More important are the mechanisms that create these paths and they are precisely the domain-general processes we mentioned earlier—chunking, categorization, inference-making, generalization (Bybee 2006*b*). So when it comes to understanding how languages are alike and how they are different, it is important once again to take a diachronic perspective and to see how the processes that create the units and structures of language interact to give us the full range of types of human language.

Grammaticalization paths and the paths of phonological change (i.e., sound change) can be thought of as substantive universals, as they refer directly to linguistic substance of phonetic form and meaning. There are also universal paths that create structural or formal universals, such as Structure Preservation (Kiparsky 1985). While this does not seem to be an absolute universal, it does express a strong tendency, which is that segments involved in lexical or morphological alternations are phonemes in their own right. Bybee (2008) demonstrates that this tendency is a result of parallel developments along several paths of change, including the unidirectional tendency for phonetic changes to become associated with morphology and lexicon. Again, the paths of change themselves are the result of the application of a handful of mechanisms that operate as language is used in context.

Usage-based theory and a complex adaptive systems approach also allows us to find explanations for correspondences that are not incorporated into other theories. For instance, the observations that grammatical morphemes (function words and affixes) are usually short (comprised of fewer segments than lexical items in the same language) and highly frequent are both observations that have a direct explanation when the usage factors in grammaticalization are taken into account (see section 32.7.3). Grammaticalization does not take place without extreme frequency increases; these same frequency increases lead to phonetic reduction.

Thus usage-based theory views language as fluid and dynamic, changing through the interaction of social usage events with the cognitive processes characteristic of the human brain in general. We have tried to show here how fundamental

cognitive processes apply to linguistic experience to create the range of units and categories exhibited in human language, those structural properties that have intrigued linguists for centuries. The basic elements we have sketched, incorporating gradient categories, exemplar storage, and a non-static representational system, can account for the striking dual nature of language, as a system that undergoes change in systematic ways, while also exhibiting sufficient stability to allow communication.