Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition[1]

JOAN BYBEE AND JAMES L. MCCLELLAND

Abstract

It is argued that the principles needed to explain linguistic behavior are domain-general and based on the impact that specific experiences have on the mental organization and representation of language. This organization must be sensitive to both specific information and generalized patterns. In addition, knowledge of language is highly sensitive to frequency of use: frequently-used linguistic sequences become more frequent, more accessible and better integrated. The evidence adduced is mainly from phonology and morphology and addresses the issue of gradience and specificity found in postulated units, categories, and dichotomies such as regular and irregular, but the points apply to all levels of linguistic analysis including the syntactic, semantic, and discourse levels. Appropriate models for representing such phenomena are considered, including exemplar models and connectionist models, which are evolving to achieve a better fit with linguistic data. The major criticism of connectionist models often raised from within the combinatorial paradigm of much existing linguistic theory – that they do not capture 'free combination' to the extent that rule-based systems do, is regarded as a strength rather than a weakness. Recent connectionist models exhibit greater productivity and systematicity than earlier variants, but still show less uniformity of generalization than combinatorial models do. The remaining non-uniformity that the connectionist models show is appropriate, given that such non-uniformity is the rule in language structure and language behavior.

1. This article reflects initial stages of a dialogue between the authors, and the order of authorship is alphabetical. JLM is indebted to Catherine Harris for valuable preliminary conversations. Supported in part by NSF grant BCS 0079044 and by NIH grant MH 64445.

The Linguistic Review 22 (2005), 381–410 0167–6318/05/022-0381

1. Introduction

There is a range of views about the relationship between language as an abstract system and what people actually say when they talk – what de Saussure (1916) called 'langue' and 'parole'. One common view is that language has an essential and unique inner structure that conforms to a universal ideal, and what people say is a potentially imperfect reflection of this inner essence, muddied by performance factors. According to an opposing view – the one we will espouse – language use has a major impact on language structure. The experience that users have with language shapes cognitive representations, which are built up through the application of general principles of human cognition to linguistic input. The structure that appears to underlie language use reflects the operation of these principles as they shape how individual speakers and hearers represent form and meaning and adapt these forms and meanings as they speak.

In this article we will consider the relationship between psychology and linguistics from the perspective provided by this second view. We will ask not what linguistics can do for psychology, but what psychology can do for linguistics. If, as we claim, the characteristics of natural language reflect robust and general properties of human behavior and cognition, then efforts to understand the use and structure of natural language will be informed by the discovery of these domain-general principles and the key properties of the mechanisms that embody them.

We focus on two very general principles applicable to all forms of action and cognition, including speech and language:

(1) Our knowledge and our behavior reflect sensitivity to both general and specific information. Every thought and every action reflects a blend of general and specific influences. The weights of these influences vary, depending on form, function, meaning, and context.

(2) As sequences of actions are performed repeatedly, they become more fluent and integrated. Repetition alters the actions themselves and their similarity relations to other actions, giving rise to changes in representation and action over time within and across individuals.

From the vantage point of these principles and of the characteristics of linguistic forms that reflect their operation, we will consider a fundamental paradigmatic assumption that lies at the heart of a great deal of theorizing in linguistics. This is the assumption that language is composed of constituent units chosen from a finite taxonomy that can be combined relatively freely (subject mainly to categorical rules or constraints) to produce a potential infinity of more complex forms, including, for example 'Colorless green ideas sleep

furiously' (Chomsky 1957). This view is espoused by Jackendoff in his monograph and in other discussions of the essential nature of language, including Chomsky (1957) and Fodor and Pylyshyn (1988), and it is reflected in virtually every formal system for representing linguistic structure.

We will suggest that this assumption must ultimately be viewed as a descriptive approximation that applies with varying degrees of adequacy to actual language forms, and does not provide an accurate characterization of their general nature. We will argue instead for other approaches in wide use to model detailed properties of human cognition and behavior: exemplar-based mechanisms that preserve fine detail of specific linguistic forms and combine information across them according to their similarity to a given situation; and connectionist models that acquire knowledge through the modification of connections among simple processing units in ways that are sensitive to both general and specific information.

Our argument will be presented as follows. We will briefly discuss the two principles above with reference to domains other than language, to illustrate their generality. We will then consider trends within the field of linguistics that indicate a trend toward increasing exploration of alternatives to combinatorial approaches, across all branches of the field. In the following section, which forms the main body of our article, we will consider in some detail evidence from phonology and morphology that exhibits the above-mentioned principles in action, and points toward alternatives in these areas of linguistics to approaches based on the traditional paradigm of linguistic theory. Finally we will consider mechanisms that can capture this detailed evidence, reviewing some applications of exemplar-based and connectionist approaches, and some of the current challenges they face.

2. Two general principles of cognition and behavior

2.1. Combined influence of general and specific information

An important experiment in the psychological literature by Posner and Keele (1968) showed that human subjects were sensitive to both general and specific information. They presented subjects with random dot displays. Each display was derived from a prototype random dot pattern by perturbing the positions of the dots in the prototype; the prototype itself was never seen. Other similar experiments have been done using many different sorts of items, including strings of consonants, or line drawings formed by combining distinct facial elements (Whittlesea 1983; Medin and Shaffer 1978). What is found in these experiments is that participants are sensitive both to the particular items they have seen and to the central tendency of the items. That is, the probability that

they will say that a new pattern comes from the same category as those previously seen is influenced by the familiarity of the item itself, by its similarity to other items already seen, and by its similarity to the prototype, never seen. When the distortions are large and each is shown repeatedly, sensitivity to the individual patterns is relatively high; when the distortions are small and each is shown only once, sensitivity to the individual patterns is relatively low and sensitivity to the prototype or central tendency is higher. The essential point is that performance with each item reflects both specific and general information.

We submit that the same situation arises in our thoughts and actions about natural objects encountered in nature. When we encounter, for example, a specific dog, what we know about other physical objects, other animals, other types of dogs, other dogs of the same or very similar breeds, and of this dog itself, all enter into our representations of and actions toward this dog. Crucially, what research in conceptual knowledge has shown is that category membership, and thus inheritance of properties shared with other category members, is a matter of degree (Rosch 1975). This is both a statement about the world – there are vast differences among dogs, birds, trees, etc in the degree to which they share properties with other dogs, birds, trees, etc – and about our cognitive representations, which reflect the graded category membership exhibited by objects in the world. For more frequent and less typical objects, people show lesser influence of knowledge of other related objects than for less frequent and more typical objects.

2.2. Repetition leads to fluency and integration

Psychologists have studied the effects of practice and repetition on human performance across a wide range of specific skills, including things as diverse as typing, arithmetic, formal proof checking, and cigar rolling. In cases where it's possible to measure rate of performance, there is a well-known relationship between speed and practice called *the power law of practice* (Blackburn 1936; Newell and Rosenbloom 1981): The time it takes to perform the act (e.g., roll a single cigar, or type a fixed number of words) decreases with the number of repetitions, taken to a small power. Qualitatively what this relationship means is that performance is always improving, although the rate of improvement diminishes rapidly (a given percentage of improvement takes an increasing number of trials to achieve).

A corollary of overall improvement with practice is increased efficiency and integration. Instead of a complex act proceeding as a series of discrete and distinct actions, the act becomes more integrated with practice so that the individual elements lose their discrete identity. This process even occurs in actions that seem inherently to be discrete, as in typing of a highly familiar word (Rumel-

hart and Norman 1982). Here one can see, for example, in the typing of the frequent word *very*, that a skilled typist will adopt a specialized hand posture for typing this word, slightly twisting the whole left hand to make it easier for the fingers to access the letters *v, e*, and *r*, while the right hand prepares itself simultaneously with the onset of the typing of the *v* to be ready to strike the *y* right after the *r*.

The above review is necessarily brief but hopefully illustrates how, outside the domain of language, the two key principles that we will appeal to in what follows are reflected.

3. Related trends within the discipline of linguistics

We will focus our discussion of the relevance of these principles for language in a detailed discussion below of their relevance to phonology and morphology, but we view these as just examples. In our view, all aspects of language structure reflect the principles of experience- and usage-dependence discussed above. Before considering phonology and morphology in some detail, we take note of broader trends in the field of linguistics, within which principles like those above have begun to assume a more prominent status.

Many researchers outside of linguistics attend mostly to the Chomskyan paradigm, which is perhaps the epitome of the combinatorial paradigm. The work of Jackendoff (2002) falls squarely within this paradigm, and it continues to have many adherents. It is important, however, to note that alternative perspectives have been pursued for several decades. In the early seventies, after a decade or so of active investigation of the formal properties of syntax initiated by Chomsky (1957), the approach hit substantial roadblocks for many researchers, leading them to explore alternative sources of linguistic explanation. A unifying theme in the four avenues of exploration that we will mention below is the expansion of the empirical database for linguistics beyond the sentence, beyond sentences made up by the researcher, beyond English and beyond the purely synchronic realm. A comprehensive treatment of these approaches directed at psychologists can be found in Tomasello's volumes, *The New Psychology of Language, Vols. I* (2001) *and II* (2003). We will mention here four issues that are treated differently in these approaches, which together are now being called *usage-based approaches* to language.

Chomsky's work attracted attention to syntactic phenomena, which had not been deeply studied by his predecessors. But the excitement about his approach to syntax died away for many researchers when they discovered that formal treatments alone were not able to explain why certain sentences worked in some contexts but not in others. Discourse-based studies set out to look at language beyond the scope of single sentences and the results called into question

the abstract formal structures that were being touted in generative grammar. The empirical results of discourse studies show that across many languages, the formal properties of lexical categories such as nouns, verbs and adjectives, the transitivity of clauses and the placement of subordinate clauses are strongly influenced by their use in natural discourse (Hopper and Thompson 1980, 1984; Thompson 1988).

For instance, Hopper and Thompson (1984) argue that the categories of noun and verb emerge from their prototypical discourse functions which are: for nouns, to introduce a referent into the discourse; and for verbs, to move the narrative sequence along. The morpho-syntactic properties that nouns and verbs display in different contexts relate directly to whether they are playing their prototypical roles. For instance, a noun used in a compound, such as *fox* in *foxhunting* does not take a plural marking or a determiner, since it is not functioning to introduce a referent into the discourse. Similarly, a verb used in a subordinate clause often lacks person/number or tense and aspect marking (i.e., it might occur in an infinitive or participial form) since it is not playing its prototypical role in the discourse. Other research on the occurrence of sentence structures in the wild have shown that real clauses rarely have two full NPs; it is much more common for transitive subjects to be pronominal while transitive objects and intransitive subjects are more likely to be full lexical NPs (DuBois 1987). This tendency explains why subject agreement is more common than object agreement (the subject pronouns for transitive verbs tend to develop into agreement markers) and how ergative marking can arise (since ergative systems group transitive object and intransitive subjects together). In fact, discourse distributions of categories and constructions provide explanations for their properties of a sort that is not even imaginable in a formal grammar that divorces language use from language structure.

Many researchers have also found it limiting to consider languages as if they were static and unchanging. Languages are the product of their evolution, and they are continually changing. Any explanation for linguistic structure must address not just how current speakers internalize this structure as it is, but also how it came into being in the first place. Joseph Greenberg, Tom Givon and others have discovered that languages change in very similar ways, both in terms of changes in structure and changes in meaning. Thus universals of language can be formulated in terms of possible trajectories of change and where a particular structure in a particular language is on one of these trajectories (Givon 1979; Greenberg 1978; Bybee and Dahl 1989). For instance, it is widely documented that phrases meaning 'be going to' or 'want' develop into markers of future; demonstratives become definite articles, and so on (Bybee 2003). This approach is applicable to syntactic change such as the development of subordinate clauses (Hopper and Traugott 1993), changes in word order, and the development of new constructions with their attendant grammat-

ical morphemes (grammaticization) (Givon 1979; Heine and Reh 1984; Bybee et al. 1994). A diachronic perspective on language shows languages to be constantly changing in a gradual and often imperceptible manner. Such gradual ongoing change is the main source of the gradience and variation that we are now finding to be so characteristic of language when studied via its actual use in context. Careful diachronic studies show that new constructions come into being and spread by gradually increasing their frequency of use over time. The nature of this spread is important evidence for the way grammar is represented cognitively. The probabilistic nature of the distribution of constructions strongly suggests that categories are gradient and processes of selection of structures to be used are dependent upon many factors, including the social situation, the particular lexical items involved, and the particular discourse context.

A third major trend is the development of cognitive linguistics, in which the cognitive structures that underlie the meaning of grammatical constructions and lexical items have been studied (Langacker 1987; Lakoff 1987). In this approach notions of categorization that have been developed in psychology (especially prototype and family resemblance structure) have been applied to linguistic categorization of word meaning, change of word meaning (Geeraerts 1997) and to the more abstract linguistic categories such as noun, verb, or relations such as 'subject of'. Spatial relations and metaphorical extensions have been of particular interest (Talmy 1983; Sweetser 1990; Lakoff 1987). The use of body part terms to represent spatial relations (e.g., *face* and *head* in spatial relations is a widespread tendency) and the extension of spatial relations into temporal domains (*before* and *after* both originated as spatial and have become temporal) have been used to argue that expression within language is highly embodied. Within this general framework, approaches to grammar utilize constructions, which are direct pairings of word or morpheme patterns with meaning (Fillmore et al. 1988; Goldberg 1995, 2003). This approach emphasizes that while constructions share many properties, such as having a subject-verb-object word order in English, a speaker's knowledge of his or her language involves a large number of constructions which have some idiosyncratic properties, such as the constructions exemplified by *Sam joked his way into the meeting.* This construction requires the use of the word *way,* a possessive pronoun coreferential with the subject, a directional phrase and verbs of certain types. The construction has a certain semantic construal associated with it that is not available without all these features (Goldberg 1995). The heavy use of such constructions in natural discourse, the knowledge of their conditions of appropriateness and how they can be expanded argues for great specificity in a speaker's cognitive representation of his or her native language. Recent work in child language acquisition (Lieven et al. 1997; Diessel and Tomasello 2000; Tomasello 2000) have shown that acquisition proceeds through the mastery of

very specific instances of constructions that then are expanded upon to reach full productivity.

A fourth way linguistic inquiry has been expanded in recent years is in the study of the multiple ways in which frequency of use impacts cognitive representations (Bybee and Hopper 2001). We mentioned above the *power law of practice* by which repeated sequences can be performed more quickly and fluently. For language structure, this principle has (at least) two consequences. First, repeated sequences, such as *in other words, take a break, break a habit, started to, I don't know* come to be processed as single units rather than morpheme-by-morpheme, much as the individual actions of a highly practiced sequence come to be performed as a single unit. One effect of frequency, then, is that sequences become highly integrated rendering them less likely to undergo the analysis that would make them susceptible to reformation on the basis of new patterns. Thus in such repeated, formulaic sequences, we find fossilized syntax and morphology from earlier periods of English (Bybee and Thompson 1997). This includes not just irregular morphology, such as *break, broke; take, took,* which resist regularization, but also whole phrases, such as *far be it from me, be that as it may, how goes it?*, *I know nothing, that proves nothing,* which preserve older word order and forms.

It appears that such sequences need not be of extreme high frequency to be treated like single lexical items, thus preserving their older structure. With much higher levels of repetition, however, the second effect of frequency emerges. High frequency words and phrases undergo more phonological reduction than low frequency sequences. This reduction is the manifestation of the increased fluency that results from practice. It can be seen in the fact that certain consonant deletion processes, such as the loss of final t/d in American English, occurs earlier in high frequency words (such as *just, went* and *and)* than in lower frequency words (such as *important, attract* and *resist*) (Bybee 2000). Reduction is especially apparent in the process of grammaticization where whole phrases are ground down to mere syllables (*going to* > [gənə]) and in the development of discourse markers (a type of grammaticization) in which phrases such as *I don't know* come to serve as markers of interaction and undergo extreme reduction, e.g. [aɪrəno] (Bybee and Scheibman 1999). In grammaticization the reduction and change due to frequency affects the meanings of the construction as well (Bybee 2003).

Given these four well-developed trends, many linguistic phenomena can be viewed quite differently than they would be in a theory that focuses on structure and disregards experience with language. In the following sections we will consider studies in phonology and morphology that reflect this orientation towards the effect of use on grammar and towards the interaction of the specific with the general. The studies mentioned here are only a few out of many recent studies that address these issues. These studies were chosen to illustrate the

interaction of specific information with general information, the effect of frequent use on representation, and (where applicable) the interaction of meaning with form.

4. Applications to phonology and morphology

4.1. Phonology

Phonotactic constraints are constraints on the sequences of consonants and vowels that can occur in a language. While it has been traditionally assumed that such constraints are categorical because speakers can distinguish between acceptable and unacceptable sequences, recent experiments have demonstrated that subjects can reliably distinguish degrees of acceptability even among occurring sequences and even judge some sequences that actually occur in the language as low in acceptability. One important factor in these judgments is the type frequency of the sequences, that is, the number of existing words that contain the sequence. Vitevitch et al. 1997 demonstrate this effect for disyllabic words by asking subjects to rate the wordlikeliness for English of nonce words. They find that subjects prefer words such as [fʌltʃʌn], which contain high frequency sequences, to words like [daibdʒaiz], which contain lower frequency combinations that are nonetheless attested sequences in English. Their results and those of others (Bailey and Hahn 2001) indicate that phonotactic judgments covary with the number of similar words. Nonce words that have many neighbors tend to be judged as very wordlike, while the nonce words with few or no neighbors tend to be judged as quite unwordlike.

Another area where similar results are found is in stress placement in Spanish. For nouns and adjectives there is a generalization that applies correctly to 95 % of the lexicon: vowel-final words have penultimate stress (*boníto* 'pretty', *cálle* 'street') while consonant-final words (excluding the plural marker *-s*) have final stress (*corazón* 'heart', *iguál* 'equal). Aske 1990 notes that nouns and adjectives ending in /n/ generally follow this rule and are stressed on the final syllable. However, a small class of exceptions are words ending in *-en* which tend to have penultimate stress (*orígen* 'origin', *crímen* 'crime'). In his experiment Aske asked Spanish speakers to pronounce nonce words occurring in noun and adjective positions in sentences; the subject read the sentences which were printed in all capital letters (so that no accent marks would occur). One might expect the subjects to follow the general rule and put final stress on all the items ending in *-n,* but instead they demonstrated knowledge of the lexical distribution. They assigned final stress to words ending in the other vowels plus *-n* in more than 96 % of the cases, but to words ending in *-en* in only 56% of the cases. (Other studies which show that speakers use knowledge of lexical

distribution in the assignment of phonological responses are Eddington 1996 and Knott et al. 2001.)

The study of phonological change in progress provides another domain in which the specific is seen to interact with the general in a way that demonstrates that specific knowledge is not completely disregarded in favor of generalizations. For reductive sound change, a frequency effect has been long observed: high frequency words undergo change at a faster rate than low frequency words. Since sound change occurs gradually and creates synchronic variation, we can observe that words have distinct ranges of variation in their phonetic implementation. For instance, high frequency words such as *just, went* or *don't* are more likely to have a final /t/ or /d/ deleted than lower frequency words such as *innocent, interest,* or *attract* (Bybee 2000). Similarly, English words that have an unstressed schwa followed by /r/ or /l/ and a final unstressed syllable have different propensities to delete the schwa, which is correlated with their token frequency. *Every* has become a two syllable word, while low frequency words such as *mammary* and *summery* are typically three syllables. Importantly, mid-frequency words such as *memory* and *family* have a range of variation between these other forms; they can be pronounced anywhere along a continuum from two to three syllables. Such detail is subphonemic and yet tied to particular lexical items. These data directly challenge the phonemic principle, which asserts that all subphonemic phonetic information is predictable from general rules and independent of individual lexical items.

There are two kinds of processes that contribute to more reduction in higher frequency items. The first process is the phonetic reduction that occurs in the production process: articulatory reduction and overlap is omnipresent in natural speech. This on-line modulation of production is governed by factors grouped under the term 'predictability' by Jurafsky et al. (2001), which includes transition probability (how likely one word is to follow another), repetition within a discourse, semantic predictability within the context, and so on. Secondly, in addition to the on-line reduction, we suggest that the underlying cognitive representations change with repetition. That is, each on-line reduction event exerts a small and cumulating influence on the speaker's representation of a word or phrase. While the factors grouped under 'predictability' clearly play an important role in phonetic change, it is their cumulative effect over repeated usage that results in the gradual reduction and eventual loss of phonemes, syllables, and boundaries between morphemes or words. This process has important implications for models of language structure and language change, which we will consider after reviewing additional relevant findings from morphology.

4.2. Morphology

4.2.1. Inflectional morphology. An alternative to rule-based, modular grammars arose from Bybee and Slobin's 1982 study of the English Past Tense formation and the connectionist model of the English Past tense by Rumelhart and McClelland (1986). Rumelhart and McClelland based their simulation on the findings of Bybee and Slobin, which pointed to the importance of token frequency in maintaining irregularity, the importance of type frequency and graded similarity in determining productivity, and the overlap between the regular and irregular past tense. Rumelhart and McClelland demonstrated that a connectionist model could produce a learning sequence similar to a child's reaction to regular and irregular English verbs without formulating a specific rule, but rather by adjusting weights of connections among simple processing units in response to the characteristics of forms provided to the model as input. This stimulated an intense debate, with Pinker and Prince (1988) weighing in on the short-comings of the connectionist model and others responding vigorously with counter-arguments and several models that improved on shortcomings of the Rumelhart and McClelland network. More crucial than the details of particular models, however, is the nature of the past tense itself. Pinker (1991) maintains the strict distinction between a regular past tense system in which past tense forms are created using a categorical symbolic rule, on the one hand, and a completely separate system for irregular verbs, within which frequency, family resemblance and graded similarity play a role, on the other.

A central element of our argument is that the claimed distinction between regular and irregular forms is not categorical, but forms instead a continuum. A close examination of the irregular verbs of English shows that many of them have properties that overlap substantially with the regular verbs (McClelland and Patterson 2002). Of the 150 or so irregular verbs of English that are in common use (from the list provided in Bloch 1947), 35 add a /t/ or /d/ in the past tense. Some examples are *feel, felt; lose, lost; tell, told; bring, brought; have, had; make, made.* In the rule-based model, affixation is accomplished by rule, and irregulars are selected from the lexicon, so there is no way to represent this generalization except by duplicating the affixation rule in the lexicon. In the connectionist rendering, where regulars and irregulars are not separate from one another, this similarity between regulars and irregulars will be discovered and exploited by the system (McClelland and Patterson 2002).

As pointed out in McClelland and Patterson (2002) a further interesting fact about the English irregulars is that many of them (about 60 out of 150) end in a /t/ or /d/ in both the base form and the past form. One class utilizes a vowel change to signal the difference between present and past, as in *find, found; feed, fed; get, got; ride, rode; stand, stood; write, wrote.* Another large class

of verbs that end in /t/ or /d/ and have lax vowels undergo no change in the past (for example, *bet, bid, cost, cut, rid),* and a few verbs change a /d/ to /t/ in the past (*bend, bent; build, built; send, sent*). Thus 95 out of 150 irregular past forms end in /t/ or /d/. A further generalization concerns the vowels of past tense forms. As Lupyan and McClelland (2003) have pointed out, some of the highest frequency past tense forms also have a lax vowel before /d/ (e.g., *had, did,* and *said).* In fact, of the 95 out of 150 irregulars that end in /t/ or /d/, 79 of them have a lax vowel. If the signal for past tense was conceived more as a template than an affixation rule, the generalization would be that past tense ends in /t/ or /d/ preceded by a lax vowel. This generalization is probabilistic: not all words that end in /t/ or /d/ are past tense and not all past tenses end in /t/ or /d/, and yet it is the type of generalization that real speakers work with (Bybee and Slobin 1982).

These facts demonstrate that there is not a clear line between regular and irregular past tense verbs, nor between specific and general knowledge. Some 'irregulars' such as *feel, felt,* and *keep, kept* are less irregular than others, such as *sing, sang* or *know, knew*. Moreover, it is clear that speakers and learners take advantage of these subregularities: pre-school children rarely overgeneralize the verbs that end in /t/ or /d/ but undergo no change in the past (Bybee and Slobin 1982; Menn and MacWhinney 1984).

In addition, even among the so-called regular verbs there are varying degrees of regularity. Losiewicz (1992) reported that when speakers produce past tense verbs, there is more reduction in the pronunciation of the final consonant and suffix in high frequency verbs than in low frequency verbs: *needed* is more reduced than *kneaded*. Reduction appears to affect less frequent verbs to some degree, in that they are less compositional in the way they are produced than nonce-verbs. Thus the articulation of the /pt/ in *wrapped* is more reduced than the production of the corresponding cluster in *vapped.*

One further important observation is that in addition to the more familiar regularization processes, language change also involves the development of new irregular patterns. In Old English (OE), there was a strong and a weak verb system with a far richer inflectional morphology than we have in present day English. Others have noted previously that as the system collapsed, the weak verb system coalesced into the current 'regular' past tense (Hare and Elman 1995), and the strong system gradually eroded. Many verbs from that period dropped out of the language, and have since been replaced by others of French or Latin origin, all of which adopted the regular pattern. Of the Old English (OE) verbs that remain, some have become regular, and many models explain this. Less well known, and far less frequently explained, is the fact that many OE weak verbs have ended up as irregulars. For example, OE *macian*, a weak verb with past tense *macode,* became modern English *make/made*. In one analysis (Lupyan, personal communication), 32 out of 141 OE weak verbs

still in the language were found to be among the so-called 'irregular' verbs of modern English.

One can see in the English past-tense system a process at work in which forms appear to be constrained both by a pressure to be short and a pressure to be like other past-tense forms. This pressure results in regularizations, but it also results in compression effects, reducing vowels and eliminating some consonants. The resulting forms, like *made* and *kept,* reflect the regular past tense (by ending in /d/ or /t/) and also the effects of compression. (As Burzio 2002 has noted, irregular past tense verbs tend to be regular phonotactically; for example, regular *loved* ends in a cluster not attested in monomorphemic English words, while exceptional *had* avoids this phonotactic irregularity.) In our view, any reasonable theory of language structure and language change will have to address these effects.

4.2.2. Derivational morphology. In derivational morphology we also find continua, but in this domain the continua are based on not just form and frequency but also on meaning. Bybee (1985) points out that words produced by derivational processes, particularly affixation, over time tend to become semantically different from the base words from which they were created. Often phonological changes accompany the semantic differentiation. Thus in derivational morphology we have situations in which words vary by degrees in the extent to which they can be decomposed into the semantic and phonological units that originally comprised them. For instance, consider the following words with an etymological prefix *pre-* and note the different vowel quality in the prefix.

(3)
a.	*president*	[ɛ]
b.	*prediction*	[ɪ]
c.	*predecease*	[i]

Not only do these words have different renderings of the vowel in *pre-*, they also have correspondingly different degrees of semantic compositionality. *President* is the most semantically opaque with regard to the identification of a morpheme *pre-* meaning 'before', while *prediction* has some sense of 'before' but the remainder of the word *-diction* is not highly associated with the derived word. In contrast *predecease* has a rather transparent composition of *pre* + *decease.* It turns out that an underlying factor in these differentiations is the frequency of use. Pagliuca (1976) shows that token frequency of the derived form corresponds to both the phonological reduction of the vowel and the semantic opacity of the word.

Hay (2001) offers a refinement of this hypothesis by proposing that the derived words are more likely to differentiate semantically from their bases if they become more frequent than their bases. Her account is that a derived word

whose base is more frequent will continue to activate the base when it is used, while a derived word that is more frequent than its base can be accessed independently of the base. In one experiment subjects were trained to consider words that could be broken down into smaller parts as 'more complex' than those that cannot. Then they were asked to rate pairs of words such as *refurbish* and *rekindle* and *diagonally* and *eternally* (the first in each pair is more frequent than its base). More than 65 % of the responses rated the words that are more frequent than their bases as 'less complex.' In another experiment she found that affixed words that were more frequent than their bases were less likely to have dictionary definitions that mentioned the bases than affixed words that were less frequent than their bases. For example, the definition of *insane* (which is more frequent than *sane*) is less likely to use the word *sane* than the definition of *inaccurate* is likely to use the word *accurate*.

Words formed through derivational morphology demonstrate a number of continua that are not compatible with a strict rule-based approach. First, the degree to which the affix is identifiable with other instances of the affix is variable; second, the degree to which the base is identifiable is also variable, perhaps independently of the transparency of the affix; third, the extent to which the semantics differs from the semantics of etymologically related forms is variable; and fourth, the degree to which the phonology is predictable from the base can vary.

Strictly rule-based approaches have failed to account for the variation found in derivational morphology. A well-known attempt is Level Ordered Phonology (Siegel 1979; Kiparsky 1982, 1985), wherein morphological rules and phonological rules are organized into groupings such that groups of affixes are associated with groups of phonological rules. These levels are intended to account for both the phonological changes triggered by affixes and the order in which the affixes occur when two or more occur together. The development of the formalist approach to affixation ran into obstacles when it emerged that some affixes were not content to remain on a single level. In addition, no semantic theory was ever devised to go along with the formal account. This is a serious shortcoming, since it is not possible to provide a complete account of morphology without taking meaning into account. Bybee (1985) has shown that meaning interacts with form in a number of important ways in the languages of the world, e.g., that meaning determines the order of inflectional affixes and the degree to which they condition phonological changes in stems (see also Burzio 2002).

Hay (2002) has proposed an alternative account of the facts in English that led to the proposal of Level Ordered Phonology. She proposes to rate morphologically complex words on how decomposable they are. Two factors are considered: the relative frequency of the derived word to the base and the phonotactics of the affix-base combination. As mentioned above, derived words that

are more frequent than their base are less decomposable. It follows then that suffixes that are found in a relatively high proportion of words that are less frequent than their bases will in general be more separable than suffixes occurring in more words that are more frequent than their bases. Hay illustrates this difference with the suffix *-ish*, which tends to occur in words that are less frequent than their bases, such as *grayish*, compared to the suffix *-ic*, which tends to occur in words that are more frequent than their bases, such as *scenic*.

The second factor considered is phonotactics. Hay's first hypothesis about phonotactics is that words containing the same suffix will be more decomposable if they contain a low probability phonotactic transition than if they do not. Thus *pipeful* is more decomposable than *bowlful.* Her second hypothesis is that suffixes beginning with consonants will tend to be more separable than suffixes beginning with vowels (e.g., *-ness* is more separable than *-ess*). Finally, as an alternative to level ordering, she proposes that more separable affixes will occur outside of less separable affixes.

Two experiments lend support to these hypotheses. In the first, subjects were asked to pick which of two words suffixed with *-al* sounds more like an English word. All of the words had the suffix *-ment* which sometimes takes *-al* and sometimes does not. In each pair, one word was more frequent than its base (*investment*) and one was less frequent (*arrangement*). Thus subjects were asked to compare words such as *investmental* and *arrangemental* and choose the one that sounds more like English. Given a large number of such pairs, the subjects preferred the affixation to forms that were more frequent than their bases to a significant degree; in other words, to forms that were in other experiments judged as less decomposable.

In the second experiment, the effects of probable vs. improbable phonotactics were explored. Two types of words with the suffix *-ment* were used: those in which the word ending and *-ment* formed a highly probably sequence, e.g., *requirement* and those in which the same sequence had a low probability, e.g., *improvement*. Pairs of such words with *-al* added to them were presented to the subjects (*requiremental* vs. *improvemental)*, who were asked to choose which word sounded more like an English word. The subjects chose the words with the highly probable phonotactic sequences significantly more often than those with the low probability sequences, supporting Hay's hypothesis that words seem less decomposable if they have probable sequences within them and that less separable affixes allow other affixes to occur with them more readily.

In contrast to the rule-based account, which has never been able to deal with the many exceptions to level-ordering, Hay's account follows from the individual characteristics of each word and each affix and does not assume that all derivational affixes can be separated into two discrete categories. Her approach depends heavily upon reference to relative frequency and processing effects on form and meaning and thus argues that grammar is highly affected

by experience with language. She posits no abstract categories or rule types, but refers only to what happens in processing. Even phonotactic generalizations that speakers are hypothesized to reference in decomposing words are based on experience with existing words of the language.

The processing mechanisms that are necessary to support the properties of derivational and inflectional morphology mentioned here require of language users only some very basic abilities that appear to be necessary under any account of language organization:

(4) Processing mechanisms

a. Speakers are attuned to the detailed phonetic relationship between an experienced input and previously experienced inputs in a way that is sensitive both to the frequency of occurrence of particular inputs and to partial correspondences for form and meaning across inputs.
b. Repeated exposure to a particular phonological pattern (be it one we classically call a morpheme, a word, or even a sequence of words) increases speed and fluency of processing of the pattern.
c. As this process is repeated, any tendency toward compositionality within the pattern is gradually reduced, leading to words and word sequences losing their compositionality if they are of high absolute or relative frequency.

All of these processes occur in real time and affect the cognitive representation of language. Because they are highly sensitive to input and the input is highly variable, the result will be representations that vary by speaker, as well as by item.

5. Implications for mechanisms of language representation, language processing, and language change

As stated at the beginning of our article, we view language structure as emerging from forces that operate during language use. Here we discuss the implications of the findings we have reviewed for mechanistic models of the representations and processes arising in the minds of language users as they acquire and use language and gradually change its structure through the accumulation of small changes occurring with each use.

Perhaps the central point is that the findings we have reviewed pose significant challenges for all combinatorial models, which treat language as consisting of a small and definite set of primitive elements combined into larger units that in turn combine into larger units still. Consider, for example, the idea that the languages of the world each select a set of phonetic segments

from a universal inventory, grouping subsets of these phones into phonemes, then composing syllables from sequences of phonemes, morphemes from sequences of syllables, and words from sequences of morphemes. This view is inconsistent with the facts of phonology and morphology we have reviewed above. Synchronically, it is apparent that existing forms exhibit conformity to specifiable strings of units to varying degrees, and diachronically, these forms gradually change with use, so that over time, a form goes from being closer to one compositional form to being closer to a different one.

While the example just given targets the traditional taxonomy of phoneme, syllable and morpheme, our point applies equally to other taxonomies of linguistic units. For example, Archangeli and Pulleyblank (1994), building on a large body of earlier work, argue that there are significant generalizations about phonological structure that are better captured using structural descriptions in which the featural information traditionally associated with phonemes is spread across several representational tiers, none of which corresponds exactly to the traditional notion of phoneme or phonetic segment. While there may be advantages to this scheme, it is still the case that a given form is composed of a definite structure, comprising a specified set of representational slots each filled with a discrete categorical feature (or left blank). Such a system still faces the fact that actual forms supposedly sharing identical structure and content actually differ in detail and gradually change over time so that the best-matching structural description can eventually be quite different than it was at first. Adequately capturing either the synchronic state or the diachronic progress of a language seems impossible in this or any other representational system in which the word is a structure made from a closed inventory of discrete parts, no matter what the parts are or how the parts are assembled.

Our point applies also to all levels of linguistic analysis, and not only to the phonological and morphological levels that have been our main focus here. The notion of 'word' runs into all the problems already alluded to. To be sure, words can be combined in novel ways, but they very often combine in familiar ways, forming phrases such as *look up, find out, cut it out, take a break* and so on. Such phrases have specific meanings related to but not predictable from the meanings of the words that comprise them. With frequency of use, such phrases can become reduced and fused just as morphemically complex words reduce and fuse. For instance, phrases that have entered into the process of grammaticization, such as *going to, want to, supposed to* undergo both phonological reduction and semantic/pragmatic change, transforming over time; so that what started out as analyzable as a sequence of words no longer has any recognizable internal structure (*gonna, wanna, s'posta*). Note, again, that we are not suggesting that the notion of word be replaced by another notion like collocation or construction, but rather, we are suggesting a much more radical claim: *there is no analysis into units at any level or set of levels that will*

ever successfully and completely capture the realities of synchronic structure or provide a framework in which to capture language change.

We do not wish to suggest that linguistic science could ever proceed without a notational system that treats larger forms as composed of smaller, recurring units. Indeed, we find ourselves continually relying on such units to refer to aspects of linguistic forms in our own writings, and we see no prospect for abandoning this practice. What we are saying, however, is that it is crucial to avoid taking the descriptive utility of treating language as though it were composed of units to indicate that it is actually composed of such units in fact.

We also wish to note that the descriptive adequacy of treating language as though it were composed from a discrete inventory of units is a matter of degree that varies from case to case. The fact that this approach seems to work quite well for some cases may contribute to the apparent tendency to suppose that what is essential about language is its combinatoriality. Certainly it must be acknowledged that it is crucial to language that it allows the generation and comprehension of novel forms. Our claim here is that there are other approaches, including those we will review below, that address the generation and comprehension of novel forms without assuming that the underlying mechanisms for language processing, language learning, and language change actually treat these novel forms as arising from a combinatorial mechanism.

A second set of implications of the facts we have reviewed here relates to the processes at work in language use and language change. These processes must have two essential characteristics:

(5) a. They must be sensitive to both general and specific information, in a way that allows individual items to reflect a combination of both. Synchronically they must allow forms representing continuous variation in the degree of compositionality vs. integration.
b. They must allow for both spreading of influences among forms and the gradual reduction and integration with repetition.

5.1. Candidate mechanisms

We will consider three broad approaches to the candidate mechanisms that have been proposed to capture the facts we have reviewed about language structure and language change.

The first approach is based on the traditional idea that language knowledge consists of a table of listed atomic forms and a system of rules for combining them, and that language use involves accessing the atomic forms and assembling them according to general, structure-but-not-content sensitive rules (Fodor and Pylyshyn 1988; Pinker 2001). This approach, perhaps the most fa-

miliar, nevertheless suffers from several problems. Fundamental, of course, as discussed throughout this article, is the problem that linguistic forms often lie at intermediate points between purely compositional and purely unitary. Related to this is the problem that the processing of particular items must simultaneously exploit both general and specific information, but most discussions of how the two forms of knowledge would be put into use suggest that they are mutually exclusive. Thus, according to Pinker (1991, 2001), to form the past tense of a word one checks to see if an exceptional form is listed for it; if not, the regular rule is used. Such an approach leaves all items that are not strictly regular to be looked up in the lexicon where there may be some similarity-based generalization, but importantly, the regular rule is not available to constrain the resulting form that is created. The fact that many irregulars share in the pattern typical of regular forms (as in *keep-kept*, *say-said*, etc.) thus lacks any motivation or explanation.

A second approach arises from what are known as *exemplar models*. According to such models, each instance of a language form a speaker hears or produces is stored in memory in a way that preserves articulatory, acoustic, contextual, semantic, and pragmatic details associated with the particular instance. When it comes time to produce a form, say the past tense of a word with a particular stem phonology, all of the forms in memory with similar stem features are activated, to an extent depending on the degree of similarity. Active forms then vote for the properties of the resulting to-be-produced form, with the strength of the vote cast depending on the degree of activation. In this way highly specific knowledge (knowledge of the past tense associated with the stem itself) as well as knowledge distributed across other forms both simultaneously play a role in shaping the resultant production.

Models of the kind just described have been very successful in a wide range of applications in cognitive psychology, and have recently become very popular in addressing word-specific variation in fine phonetic detail, context-sensitivity of such details, and gradual language change (Johnson 1997; Pierrehumbert 2001). For example, to address the fact that frequent words are compressed in duration relative to infrequent words, Pierrehumbert suggests that each production of a given word is based on a sampling from the stored examples; she further suggests that the sampled form is then reduced just a little as it is produced, and that the result is stored in memory, in both the speaker and the hearer. The result of such a process would be to gradually reduce the average length of the stored examples, so that when next the word is to be produced, it will on average be shorter than it was the time before. Because the reduction occurs with each repetition, the frequent forms change faster than less frequent forms (as mentioned above). An implication of this account of reductive sound change is that the average length of stored forms for frequent items will be less than the average length for infrequent items (Gregory et al. 1999). Spread-

ing of reductive effects through the language can then occur because of partial activation of the stored traces of a frequent word when other, similar words are encountered. For example, the sound change in Old English that raised [a] to [o] before nasals, started out in high frequency words and later spread to all words (Phillips 1980). Closer to home, the deletion of American English word-final [t] and [d] has started in high frequency words, such as *just, went*, and *and* and appears to be gradually spreading to words of lower frequency. Often, the spread of a change does not affect words in isolation, but rather in the contexts in which they appear the most. Thus the form *can* reduces to [kn] within the construction in which it acts as a modal and is followed by a verb. Reduction does not occur for other uses of this form, including both the modal *can* in other contexts (*Yes, I can*) and the homophonic noun. Models of this type seem quite well suited to capturing many of the findings we have considered in this article, and we look forward to their further development and elaboration.

A third approach that has been applied extensively in both linguistic and non-linguistic domains is the one provided by the connectionist or parallel-distributed processing framework. The innovation in this approach is that language knowledge is not stored in the form of items or rules, but in the form of changes to the strengths of connections among simple processing units. While early connectionist models (often called 'localist' connectionist models) assigned individual units to stand for cognitive and linguistic entities such as features, phonemes, and words (McClelland and Rumelhart 1981), it soon became apparent that it was possible to capture sensitivity to both general and specific information in connectionist models in which cognized items are represented as distributed patterns of activation over collections of processing units. We suggest that these distributed connectionist models, like exemplar models, show considerable promise for addressing the findings considered in earlier sections. In what follows we consider some of the key properties of existing models relevant to these findings, and we mention ongoing efforts extending this class of models to overcome some shortcomings of existing implementations.

Two early distributed connectionist models (Knapp and Anderson 1984; McClelland and Rumelhart 1985) exhibited sensitivity to general and specific information of the type demonstrated in experiments like the one with dot patterns by Posner and Keele (1968). In Knapp and Anderson, for example, units were assigned to overlapping regions of the rectangular plane on which the stimulus dots appeared, such that the presentation of a dot at a particular location activated a set of adjacent units. Knowledge of the dot patterns was stored in the strengths of the connections among the units that were active together within individual dot patterns used in training. The model exhibited sensitivity to the central tendency of the entire ensemble of patterns, as well as sensitivity to specific patterns presented to it. The model of McClelland and Rumelhart (1985) addressed details of sensitivity to general and specific information, in-

cluding the fact that sensitivity to specific patterns increases with frequency of presentation and decreases with overlap of the training patterns.

The past tense learning model of Rumelhart and McClelland (1986) drew on the same principles. One set of units was assigned to represent phonetic properties of the stems of words, and another was assigned to represent phonetic properties of their past tense forms. The network was trained by exposure to stem forms paired with their past tenses, and the strengths of connections from units for the stem to units for the past tense of each word were adjusted in small increments after each presentation. Like the models reviewed above, the model exhibited sensitivity to the general and the specific – in this case, the regular past tense rule and the properties of particular exceptions. Crucially, the knowledge of both general and specific information was embedded in the set of connection weights used to process all items, thereby allowing the network to exploit both general and specific information simultaneously in producing its output for a particular form. For example, in producing the past tense of *keep*, the network was able to exploit knowledge of the regular pattern (add /t/ after unvoiced consonants), knowledge of the /ɛ/ → /iː/ alternation typical of many verbs ending in *eep* (including *sleep*, *sweep*, *creep*, etc), and knowledge obtained from learning about the specific past tense form of the word *keep* itself.

These early models required the modeler to specify how each item was to be represented across the units in the network. Rumelhart and McClelland chose a particular scheme (assigning units to represent triples of phonetic features, one from each of three adjacent phonemes), and this particular choice, as well as the general problem of having to make a choice, has proved to be problematic. Critics assigned both the successes (Lachter and Bever 1988) and the failures (Pinker and Prince 1988) of the model to the choice of representations.

Thankfully, subsequent developments within connectionist theory made the specification of the properties of the internal representations of specific forms unnecessary. The key development was the discovery of methods (including the back-propagation learning algorithm, Rumelhart et al. 1986, and the contrastive Hebbian or Boltzmann machine learning algorithm, Ackley et al. 1986) that could be used to learn useful internal representations. As nicely demonstrated by Hinton (1986), a network trained with one of these algorithms can learn to assign system-internal patterns of activation to different inputs to capture patterns of co-occurrence among the different inputs. Hinton trained a network on simple propositions capturing familial relations among a number of individuals (Colin's Father is Harold, Luigi's Mother is Sophia). The set of propositions characterized relations among two separate families, but the network was given no information about family membership other than that implicit in the set of relations. The network learned to assign units in its internal representation that captured the implicit family membership, along

with other dimensions relevant to the relations encoded in the set of propositions.

It is true that the coding of inputs and outputs in some form is always necessary, but the form can be a relatively raw sensory representation as provided by peripheral sense organs such as the ear or a relatively raw motoric representation such as the time series of neural signals to the muscles controlling oral articulators. In principle, at least, all other forms of representation can be derived through the learning process itself, gradually adjusting the strengths of connections that determine how particular patterns of activation in the sense organs are recoded to allow the network to assign its own metric of similarity relevant to capturing the structure present in the set of items it is tasked to learn.

An illustration of the promise of this approach is provided by the *simple recurrent network* (*SRN*) used by Elman (1990). The simple recurrent network receives a sequence of inputs, and has the task of predicting the next element of the sequence from the current input, by way of a set of intermediate or 'hidden' units. Learning in the network occurs by adjusting the strengths of the connections from the input to the hidden units, and from the hidden units to the output units, to improve the match of the network's prediction to the actual next element. In addition, there is another set of connections from the hidden units back onto themselves so that the previous state of the hidden units can serve as a constraining influence on the next pattern of hidden unit activation. For illustrative purposes, Elman chose to assign a distinct input pattern for each distinct word in a small vocabulary of nouns and verbs; these words were then strung together to form sentences constrained by semantic and syntactic regularities, and then the sentences were strung together to make a long sequence. The patterns used to represent the different words on the input were random strings of 1's and 0's, but the network learned to assign patterns of activation at the hidden level, such that the similarities among the learned patterns captured similarities in the restrictions they placed on what words might follow them in the sequence. For example, some nouns were animate and others were not; some verbs were transitive and others intransitive; etc. As in the Hinton model, these features of the verbs were not coded explicitly, but instead were implicit in the sequential structure of the word stream. The network essentially discovered the relevant linguistic features by assigning similar representations to words with similar sequential constraints of their co-occurrences with other words in the sequence. Note also that these models exhibit sensitivity to sequential organization without explicit representation of superordinate units such as syllable, morpheme or word when the inputs correspond to letters or phonemes or of noun phrase, verb phrase, and sentence when the inputs correspond to individual words.

What is crucial for our purposes here is to note that the simple recurrent network is, like other connectionist networks, capable of sensitivity to general

as well as specific information. Thus, for example, if a particular noun tends to occur somewhat more than some other nouns as the subject of a particular verb, the network may become sensitive to this fact. Both learning and processing within such models is inherently gradient in nature and thus inherently compatible with the gradience and gradual change we see everywhere in language. The learned internal representation of an item can overlap with, and hence share in the knowledge the network has about, other related items to a varying degree. Furthermore, if distributional information in the input changes a little, the network can adjust its weights, so that there is a gradual change in representation corresponding to the distributional shift. These properties all contribute to our belief that something rather like a connectionist network must underlie human knowledge and use of language.

It is interesting to reflect on the fact that many linguists and linguistically-oriented psychologists working within the Chomskian paradigm have rejected connectionist models because of their failure to conform to their idealized conception of the combinatorial nature of language. Thus Pinker (1991) argues that a categorical rule is necessary to capture the regular English past tense because he claims that it applies equivalently to all forms, while connectionist models exhibit some sensitivity to item-specific information even within the regular forms. And Jackendoff (2002) urges that connectionist models can never be correct for other aspects of language such as syntax because it exhibits the principle of 'free combination' while connectionist models do not.

What we would say in reply is that these critiques are flawed both in their characterization of the properties of connectionist models and in their characterization of natural language. Recent connectionist models (e.g., Plaut et al. 1996) exhibit more systematicity and productivity than their predecessors (e.g., Seidenberg and McClelland 1989). While they are still not as strict in their systematicity as models constructed within the combinatorial paradigm, we argue that they ought not to be if they are to capture the actual degree of productivity and systematicity that is to be found in real languages. It is a strength and not a shortcoming of connectionist models that they are sensitive to item-specific as well as general information, even though this item-sensitivity makes them generalize non-uniformly in some cases, because this dual sensitivity is in fact a universal property of language, and because non-uniformity of generalization is the rule. Famous purported cases of uniform generalization, such as the German *-s* plural, do not show such uniformity in fact (Bybee 1995; Hahn and Nakisa 2000). Similarly, corpus-based studies of syntax show grammatical constructions to be highly specific in their application (Fillmore et al. 1994; Tottie 1991); natural language use is characterized by specific conventionalized expressions and phrases (Erman and Warren 2000; Wray 2002).

In our view the only hope for adequate mechanistic accounts of language processing, language learning and language change lies in the acceptance of

the fact that generalization and systematicity are not typically uniform and in the further exploration of mechanisms like those offered by exemplar models and connectionist models that intrinsically reflect this fundamental property.

5.2. Limitations and advancements

All of this is not to say that existing connectionist models are not without their limitations. Here we comment on a few of the limitations of Elman's simple recurrent networks (SRN), and note the state of progress in addressing them.

Firstly, Elman's SRN's treated language as simply a matter of sequential structure, and did not address the relation between spoken input and meaning. Language representations must be structured by both. Work by St. John and McClelland (1990) and by Miikkulainen and Dyer (1991) were initial efforts to address the relation between form and meaning. Subsequent work by Rohde (2002) combined sequential prediction with mapping from form to meaning, making some advances on St. John and McClelland, primarily in extending the meaning representations to accommodate the content of sentences with embedded clauses. These efforts have all been somewhat constrained by the difficulty of specifying what constitutes an appropriate representation of the meaning of an utterance. A degree of stipulation of the internal representation has been required here which cuts against the usual grain of the connectionist approach of allowing the internal representation to arise from a learning process. In theory, semantic representations arise from experience with the patterns of co-occurrence and interactions among objects in real events taking place in the world, but it is not yet possible to train a network to extract representations of these objects and their interactions from raw experience.

A second point is that none of the work mentioned above has broken free of the simplification of treating the input as a discrete sequence of items (letters, phonemes, or words). As such these models cannot really address the continuous nature of actual spoken inputs and the fully gradient nature of the similarity and compositional relations that hold between different items. An important next step is to move toward models that deal directly with the fine structure of the acoustic/perceptual input and articulatory output. There is now some relevant work actually directly applying connectionist methods to process continuous auditory and articulatory sequences (Keidel et al. 2003; Kello and Plaut in press), but more effort is needed in this direction.

Lastly, as yet we are not aware of any work with sequential networks that addresses the issue of use-dependent reduction and compression. Two ongoing projects are, however, of relevance, and suggest that progress on this front may be in the offing. First, Lupyan and McClelland (2004) have trained a pair of neural networks, one of which produces a pattern representing a word's pro-

nunciation from a pattern representing its meaning, and another that takes the pronunciation and uses it to recover the pattern representing the word's meaning. Each network is trained separately, one as a 'speaker' and one as a 'listener', and they are also trained together as a speaker-listener pair. In this latter situation, each is allowed to adjust its connection weights to (a) ensure adequacy of communication and (b) allow for reduction of the phonological pattern that the speaker produces for the listener to hear. The latter pressure is inversely proportional to the accuracy of comprehension – effectively, reduction is allowed only when comprehension is adequate. The result is that frequent forms gradually become compressed. When compression affects high-frequency past tenses, there tends to be some preservation of a trace of the regular past tense marking, resulting in forms like *said* or *had*. The limitation of this model is that the patterns used to represent the phonological content for word forms involve discrete units standing for separate phonemes. While units are allowed to take on intermediate degrees of activation, so that as compression continues, phonemes gradually disappear, we ultimately would prefer a model that uses a fully continuous representation of the output of articulation. The second project (Plaut, personal communication) still retains the essential notion of a sequence of phonemes, but produces graded outputs over time rather than a single output for each phoneme in a sequence. In this approach, the network is trained with the goal of first getting its outputs to match the first phoneme as quickly as possible, then achieving the same for the second, etc. Gradually with practice the network produces each phoneme in a sequence more and more quickly. Combining this approach with that of Lupyan and McClelland, so that the degree of speedup is not strictly constrained by matching target phoneme patterns, but is instead constrained by achieving adequate comprehensibility of the spoken form, would free the model from any explicit specification of the component structure of speech, while yet allowing the network to capture a more graded analog of such structure to the extent necessary to achieve adequate communication.

6. Conclusion

We have argued that the facts of phonology and morphology require an approach to language use and language representation in which there is no fixed set of stipulated units or rules for combining them. Instead there is knowledge derived from experience with language forms and their contexts of use that allows sensitivity to general and specific information to arise in a graded and malleable way from experience. This conclusion can be applied to other domains of language as well. As we mentioned earlier, discourse studies show that linguistic categories such as noun and verb are gradient and based on us-

age. Diachronic studies show that grammatical categories, constituents and constructions develop gradually over time and are highly impacted by frequency of use indicating that both units and categories can change gradually over time. Current approaches to language acquisition demonstrate that children first use new constructions in the context of very specific lexical material and only later learn to expand to use constructions productively. The pervasive presence of conventionalized collocations in natural speech and writing suggest strongly that specific instances are not thrown out when abstractions over such instances are formulated (Lieven et al. 1997; Tomasello 2000; Savage et al. 2003). Cognitive studies of word meaning and categorization in general show gradual changes responding to experience and usage (Geeraerts 1997; Traugott and Dasher 2002). Thus the evidence that linguistic knowledge is both specific and gradient, as well as heavily based on experience is strong at all of the traditional levels of analysis.

We now turn to the following question:

(6) If indeed language use is actually based on representations and processes that are gradient and malleable in nature, just where does it leave the enterprise that still lies at the heart of the field of linguistics, namely the formal characterization, not of what people actually say, but of the supposed underlying ideal object, the "language"?

What we suggest is that there may not in fact be any truly correct formal characterization, either of any given language or of the common elements of the set of all possible languages. Furthermore, to the extent that any given formal characterization does approximately capture the structure of a given language or set of languages, the properties of this formal structure are not themselves the underlying explanatory principles they have been taken to be within the Chomskian paradigm; instead they are facts about the structure that are to be explained as arising from the cumulative impact of the processes that shape each language, as it adapts through the process of language use.

University of New Mexico
Carnegie Mellon University

References

Ackley, David H., Geoffrey E. Hinton, and Terrence J. Sejnowski (1985). A learning algorithm for Boltzmann machines. *Cognitive Science* 9: 147–169.

Aske, Jon (1990). Disembodied rules vs. patterns in the lexicon: Testing the psychological reality of Spanish stress rules. *BLS* 7: 30–45.

Archangeli, Diana and Douglas Pulleyblank (1994). *Grounded Phonology*. Cambridge, MA: MIT Press.

Bailey, Todd M. and Ulrike Hahn (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods. *Journal of Memory and Language* 44: 568–591.
Blackburn, Julian M. (1936). Acquisition of skill: An analysis of learning curves. IHRB Technical Report, No. 73.
Bloch, Bernard (1947). English verb inflection. *Language* 23: 399–418.
Burzio, Luigi (2002). Missing players: Phonology and the past-tense debate. *Lingua* 112: 157–199.
Bybee, Joan L. (1985). *Morphology: A Study of the Relation between Meaning and Form.* Philadelphia: John Benjamins.
— (2000). The phonology of the lexicon: Evidence from lexical diffusion. In *Usage-Based Models of Language*, Michael Barlow and Suzanne Kemmer (eds.), 65–85. Stanford: CSLI.
— (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.
— (2003). Cognitive processes in grammaticalization. In *The New Psychology of Language*, Volume II, Michael Tomasello (ed.), 145–167. Mahwah: Erlbaum.
Bybee, Joan L. and Östen Dahl (1989). The creation of tense and aspect systems in the languages of the world. *Studies in Language* 13: 51–103.
Bybee, Joan L. and Paul Hopper (eds.) (2001). *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins.
Bybee, Joan L., Revere Perkins, and William Pagliuca (1994). *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World.* Chicago: University of Chicago Press.
Bybee, Joan L. and Dan I. Slobin (1982). Rules and schemas in the development and use of the English Past Tense. *Language* 58: 265–289.
Bybee, Joan L. and Sandra Thompson (1997). Three frequency effects in syntax. *BLS* 23: 378–388.
Chomsky, Noam (1957). *Syntactic Structures*. The Hague: Mouton.
— (1959). A review of B. F. Skinner's *Verbal Behavior*. *Language* 35: 26–58.
De Saussure, Ferdinand (1916). *Cours de Linguistique Générale*. Paris: Payot.
Diessel, Holger and Michael Tomasello (2000). The development of relative clauses in early child language. *Cognitive Linguistics* 11: 131–152.
DuBois, John (1987). The discourse basis of ergativity. *Language* 63: 805–855.
Eddington, David (1996). Diphthongization in Spanish derivational morphology: An empirical investigation. *Hispanic Linguistics* 8: 1–35.
Elman, Jeffrey L. (1990). Finding structure in time. *Cognitive Science* 14: 179–211.
Fillmore, Charles J., Paul Kay, and Mary Kay O'Connor (1988). Regularity and idiomaticity in grammatical constructions. *Language* 64: 501–538.
Fodor, Jerry and Zenon Pylyshyn (1988). Connectionism and cognitive architecture. *Cognition* 28: 3–71.
Geeraerts, Dirk (1997). *Diachronic Prototype Semantics: A Contribution to Historical Lexicology.* Oxford: Oxford University Press.
Givón, Talmy (1979). *On Understanding Grammar.* New York: Academic Press.
Goldberg, Adele (1995). *Constructions: A Construction Grammar Approach to Argument Structure.* Chicago: University of Chicago Press.
— (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Science* 7: 219–224.
Greenberg, Joseph (1978). How does a language acquire gender markers. In *Universals of Human Language, Vol. 3: Word Structure,* Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik (eds.), 47–82. Stanford: Stanford University Press.
Gregory, Michelle L., William D. Raymond, Alan Bell, Eric Fosler-Lussier, and Daniel Jurafsky (1999). The effects of collocational strength and contextual predictability in lexical production. *CLS* 99: 151–166.
Hahn, Ulrike and R. C. Nakisa (2000). German inflection: Single route or dual route? *Cognitive Psychology* 41: 313–360.
Hare, Mary and Jeffrey L. Elman (1995). Learning and morphological change. *Cognition* 56: 61–98.

Hay, Jennifer (2001). Lexical frequency in morphology. Is everything relative? *Linguistics* 39: 1041–1070.

— (2002). From speech perception to morphology: Affix ordering revisited. *Language* 78: 527–555.

Heine, Bernd, and Mechthild Reh (1984). *Grammaticalization and Reanalysis in African Languages*. Hamburg: Helmut Buske Verlag.

Hinton, Geoffrey E. (1986). Learning distributed representations of concepts. *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, 1–12.

Hopper, Paul and Sandra Thompson (1980). Transitivity in grammar and discourse. *Language* 56: 251–299.

— (1984). The discourse basis for lexical categories in universal grammar. *Language* 60: 703–752.

Hopper, Paul and Elizabeth Traugott (1993). *Grammaticalization.* Cambridge: Cambridge University Press.

Jackendoff, Ray (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. New York: Oxford University Press.

Johnson, Keith (1997). Speech perception without speaker normalization. In *Talker Variability in Speech Processing,* Keith Johnson and John W. Mullennix (eds.), 145–165. San Diego: Academic Press.

Jurafsky, Daniel, Alan Bell, Michelle Gregory, and William D. Raymond (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In *Frequency and the Emergence of Linguistic Structure*, Joan L. Bybee and Paul Hopper (eds.), 229–254. Amsterdam: John Benjamins.

Keidel, James L., Jason D. Zevin, Keith R. Klunder, and Mark S. Seidenberg (2003). Modeling the role of native language knowledge in perceiving nonnative speech contrasts. *Proceedings of the 15th International Congress of Phonetic Sciences*, 2221–2224.

Kello, Christopher T. and David C. Plaut (2004). A neural network model of the articulatory-acoustic forward mapping trained on recordings of articulatory parameters. *Journal of the Acoustical Society of America* 116: 2354–2364.

Kiparsky, Paul (1982). Lexical phonology and morphology. In *Linguistics in the Morning Calm*, I. S. Yang (ed.), 3–91. Seoul: Hanshin.

— (1985). Some consequences of lexical phonology. *Phonology Yearbook* 2: 85–138.

Knapp, Andrew G. and James A. Anderson (1984). Theory of categorization based on distributed memory storage. *Journal of Experimental Psychology: Learning Memory, and Cognition* 10: 616–637.

Knott, Andrea, R. Harald Baayen, and Robert Schreuder (2001). Analogy in morphology: Modeling the choice of linking morphemes in Dutch. *Linguistics* 39: 51–93.

Lachter, Joel and Thomas G. Bever (1988). The relation between linguistic structure and associative theories of language learning – a constructive critique of some connectionist learning models. *Cognition* 28: 195–247.

Lakoff, George (1987). *Women, Fire and Dangerous Things: What Categories Reveal about the Mind.* Chicago: University of Chicago Press.

Langacker, Ronald (1987). *Foundations of Cognitive Grammar, Vol. 1. Theoretical Prerequisites.* Stanford: Stanford University Press.

— (2000). A dynamic usage-based model. In *Usage-Based Models of Language*, Michael Barlow and Suzanne Kemmer (eds.), 1–63. Stanford: CSLI.

Lieven, Elena V. M., Jullian M. Pine, and Gillian Baldwin (1997). Lexically-based learning and early grammatical development. *Journal of Child Language* 24: 187–220.

Losiewicz, Beth L. (1992). The effect of frequency on linguistic morphology. Ph.D. dissertation. Austin: University of Texas.

Lupyan, Gary and James L. McClelland (2003). Did, made, had, said: Capturing quasi-regularity in exceptions. *Proceedings of the Annual Meeting of the Cognitive Science Society* 25: 740–745.

— (2004). Why irregulars make sense: Simulating the emergence of exceptions. Paper presented at The Fifth International Conference on the Evolution of Language: Leipzig, Germany.

McClelland, James L. and Karalyn Patterson (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences* 6: 465–472.

McClelland, James L. and David E. Rumelhart (1981). An interactive activation model of context effects in letter perception. Part I: An account of basic findings. *Psychological Review* 88: 375–407.

— (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General* 114: 159–188.

Menn, Lise and Brian MacWhinney (1984). The repeated morph constraint: Toward an explanation. *Language* 60: 519–541.

Medin, Douglas L. and M. M. Shaffer (1978). Context theory of classification learning. *Psychological Review* 85: 207–238.

Miikkulainen, Risto and Michael G. Dyer (1991). Natural language processing with modular PDP networks and distributed lexicon. *Cognitive Science* 15: 343–399.

Newell, Allen and Paul S. Rosenbloom (1981). Mechanisms of skill acquisition and the law of practice. In *Cognitive Skills and their Acquisition*, John R. Anderson (ed.), 1–55. Hillsdale, NJ: Erlbaum.

Pagliuca, William (1976). PRE-fixing. Ms. SUNY Buffalo.

Pierrehumbert, Janet (2001). Exemplar dynamics: Word frequency, lenition and contrast. In *Frequency and the Emergence of Linguistic Structure*, Joan L. Bybee and Paul Hopper (eds.), 137–157. Amsterdam: John Benjamins.

Pinker, Steven (2001). Four decades of rules and associations, or whatever happened to the past tense debate? In *Language, the Brain, and Cognitive Development: Papers in Honor of Jacques Mehler*, Emmanuel Dupoux (ed.), 157–179. Cambridge, MA: MIT Press.

— (1991). Rules of language. *Science* 253: 530–535.

Pinker, Steven and Alan Prince (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28: 73–193.

Plaut, David C., James L. McClelland, Mark S. Seidenberg, and Karalyn Patterson (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review* 103: 56–115.

Posner, Michael I. and Steven W. Keele (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology* 77: 353–363.

Rohde, Douglas (2002). A connectionist model of sentence comprehension and production. Ph.D. dissertation. Pittsburgh, PA: Carnegie Mellon University.

Rosch, Eleanor. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* 104: 192–233.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams (1986). Learning representations by back-propagating errors. *Nature* 323: 533–536.

Rumelhart, David E. and James L. McClelland (1986). On learning the past tenses of English verbs: Implicit rules or parallel distributed processing? In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, James L. McClelland, David E. Rumelhart, and the PDP Research Group (eds.), 216–271. Cambridge, MA: MIT Press.

Rumelhart, David E. and Donald A. Norman (1982). Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science* 6: 1–36.

St. John, Mark F. and James L. McClelland (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence* 46: 217–257.

Savage, Ceri., Elena V. M. Lieven, Anna Theakston, and Michael Tomasello (2003). Testing the abstractness of young children's linguistic representations: Lexical and structural priming of syntactic constructions? *Developmental Science* 6: 557–567.

Siegel, Dorothy (1979). *Topics in English Morphology.* New York: Garland.

Sweetser, Eve (1990). *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure.* Cambridge: Cambridge University Press.

Talmy, Leonard (1983). How language structures space. In *Spatial Orientation: Theory, Research and Application,* Herbert L. Pick, Jr and Linda P. Acredolo (eds.), 225–282. New York: Plenum Press.

Thompson, Sandra (1988). A discourse approach to the cross-linguistic category 'adjective'. In *Explaining Language Universals,* John Hawkins (ed.), 167–185. Oxford: Basil Blackwell.

Tomasello, Michael (2000). First steps in a usage-based theory of language acquisition. *Cognitive Linguistics* 11: 61–82.

Tomasello, Michael (ed.) (2001). *The New Psychology of Language, Vol. I.* Mahwah: Lawrence Erlbaum.

— (2003). *The New Psychology of Language, Vol. II.* Mahwah: Lawrence Erlbaum.

Tottie, Gunnel (1991). Lexical diffusion in syntactic change: Frequency as determinant of linguistic conservatism in the development of negation in English. In *Historical English Syntax,* Dieter Kastovsky (ed.), 439–467. Berlin: Mouton de Gruyter.

Traugott, Elizabeth and Richard Dasher (2002). *Regularity in Semantic Change*. Cambridge: Cambridge University Press.

Vitevitch, Michael S., Paul A. Luce, Jan Charles-Luce, and David Kemmerer (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech* 40: 47–62.

Whittlesea, Bruce W. A. (1983). Representation and generalization of concepts: The abstractive and episodic perspectives evaluated. Ph.D. dissertation. Hamilton, ON: MacMaster University.