

# Explaining derivational morphology as the convergence of codes

Mark S. Seidenberg and Laura M. Gonnerman

**Language users have a remarkable ability to create, produce and comprehend complex words. Words such as *undercut* and *bakery* appear to be composed of units, traditionally called morphemes, that recombine in rule-like ways to form other words, such as *underline* and *cannery*. However, morphological systems are quasiregular: they are systematic and productive but admit many seemingly irregular forms. Thus, *bakery* is related to *bake* and *cannery* to *can* but what is the *groce* in *grocery*? There is no bread in *sweetbreads*, liver in *deliver*, corn in *corner* or ginger in *gingerly*. Such words exhibit partial regularities concerning the correspondences between form and meaning, the treatment of which has important implications for linguistic and psycholinguistic theories. This article describes an approach to morphological phenomena called the convergence theory, in which morphology is a graded, inter-level representation that reflects correlations among orthography, phonology and semantics.**

Derivational morphology is the aspect of language concerned with the structure and formation of words. Words such as *government* and *fittingly* exhibit internal structure: components such as *govern*, *-ment*, *fit*, *-ing* and *-ly* appear in other words in which they perform similar functions. Moreover, there are constraints on the forms that words can take that apparently operate over such subword units; thus *governmental* is a word in English but *governalmment* could not be. Our ability to create and understand new words (*geneticize* and *impactful* are two we heard recently) also implicates this level of lexical structure. How this information is acquired, represented and used are central issues in the study of language. Most theories assume that complex words consist of discrete units (morphemes) that are recombined by morphological rules. In this article, we briefly review some of the limitations of this approach, and offer an alternative approach inspired by connectionist theories of knowledge representation and learning, in which graded, non-discrete morphological structures emerge in the course of learning relations among the spellings, sounds and meanings of words.

The aim of this article is to articulate an approach to thinking about complex words rather than exhaustively review the literature or propose a specific model, and so some disclaimers are necessary. The article emphasizes phenomena concerning derivational morphology in English; issues concerning other types of morphology (Box 1) and crosslinguistic variation are discussed only briefly. Important questions about the role of modality in processing morphologically complex words and about the

impairments that occur as a consequence of brain injury are also beyond the scope of the article. Although the approach developed here is likely to prove valuable in understanding these issues – and indeed must if it is to be viable – the discussion is necessarily focused on a narrower range of phenomena.

## The classical approach

The classical treatment of word formation within structural linguistics<sup>1</sup> was that words consist of primitives, called morphemes, that are minimal meaning-bearing units. On this view, *bake*, *baker* and *bakers* consist of one, two and three morphemes, respectively, each of which contributes something to the meaning of these words and to other words in which they occur. Words are created by rules that combine these morphemic primitives, allowing the creation of new expressions such as *sun-baked* and *bake-off*. This approach will be referred to as ‘discrete morphology’, because morphemes are discrete units, like beads on a string.

The classical view works well for words such as *baker* or *recode*, whose meanings are a compositional function of their morphemic components. *Bake*, for example, makes similar contributions to the meanings of *baker*, *baking* and *bakery* and *-er* is the agentive suffix in words such as *runner*, *bitter* and *writer*. Hence *bake* and *-er* function as classical morphemes from which the meaning of *baker* derives. The classical theory works less well in other cases, however. For example, the assumption that morphemes are minimal meaning-bearing units was called

M.S. Seidenberg is in the Neuroscience Program and the Department of Psychology, University of Southern California, Los Angeles, CA 90089-2520, USA.

tel: +1 213 740 9175  
fax: +1 213 740 5687  
e-mail: marks@gizmo.usc.edu

Laura M. Gonnerman is at the Department of Psychology and the Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213-2683, USA.

### Box 1. Main types of morphological structure

Inflectional morphology is concerned with variations in the forms of words related to the syntactic structures in which they occur. The main types of inflectional morphology in English are tense on verbs (*bake–baked–baking*) and number on nouns (*boy–boys*). Other languages encode other types of information, such as gender and case, inflectionally.

Derivational morphology concerns word formation processes that create related words through devices such as prefixing (e.g. *type–retype*), suffixing (e.g. *govern–government–governmental*) and infixing (insertion of elements within a base form, as in Arabic and Hebrew). There is also circumfixing (adding an element both before and after a morpheme), which is analyzed as either a distinct type of morphological structure or simultaneous prefixing and suffixing. Finally, reduplication (repeating all or part of a morpheme, as in Tagalog) is relatively rare across the worlds' languages and differs from the previously mentioned processes in that the element that is affixed derives from the stem itself.

Compounding is the process by which existing words are combined to form new ones. In English, most compounds consist of two nouns (e.g. *boathouse, houseboat*) or an adjective and noun (e.g. *blackbird, redhead*), although there are occasional other cases (e.g. verb–noun, *swearword, cutthroat*).

This list is not exhaustive; see Spencer (Ref. a) for an excellent overview. An important unresolved issue is whether the same principles govern all types of morphological structure. Linguistic theories have traditionally treated inflection and

derivation as systems governed by different principles (Ref. b). However, other theorists see derivation and inflection as differing only in degree (Ref. c). In English, inflection, derivation and compounding are similar insofar as all are quasiregular and so the same issues about the treatment of partial regularities arise. For example, past tense formation, a type of inflectional morphology, appears to be rule governed but there are numerous exceptions. As in the cases discussed in the text, many of the exceptions are not arbitrary; for example, the irregular forms *slept* and *kept* overlap phonologically with rule-governed forms such as *stepped*. Similarly, 'derivational' processes such as affixation are also used for inflection (e.g. the regular English plural morpheme *-s* is a suffix), but with somewhat different effects: whereas an inflected form tends to have the same grammatical category as its stem, a derived form might have a different grammatical class (e.g. *bake* is a verb, *baker* a noun), although not always (*heat* and *preheat* are both verbs). These observations suggest that there may not be a discrete boundary between inflectional and derivational morphology, consistent with the convergence theory, in which the same principles apply to both.

#### References

- a Spencer, A. (1991) *Morphological Theory*, Blackwell
- b Anderson, S.R. (1992) *A-morphous Morphology*, Cambridge University Press
- c Bybee, J. (1985) *Morphology: A Study of the Relation Between Meaning and Form*, John Benjamins

into question by Aronoff<sup>2</sup> who noted that there are subword patterns that seem to function as morphemes insofar as they enter productively into word formation, but have little or no meaning. Aronoff's examples included *-mit* in *permit, submit* and *remit*, *-duce* in *reduce, induce* and *deduce*, and *cran-* in *cranberry*, but there are many others. To appreciate the scope of the problem, consider a few additional cases. *Grocer* appears to pattern with *baker* and *grocery* with *bakery*. Yet unlike *bake, groce* does not appear to have a specific meaning. In other cases, words consist of morpheme-like units that contribute to their meanings but in a less transparent way. For example, the meaning of *dislike* is a compositional function of the meanings of the morphemes *dis-* and *like*, and similarly for *disagree* and *disinter*. *Discover* also seems to consist of two morphemes and, unlike *grocer*, each morpheme has a meaning and participates in many words. However, although *cover* is not unrelated to the meaning of *discover*, its contribution is less direct than in a word such as *uncover*. Note also that the pronunciation of *discover* differs from what it would be if it were truly compositional. This pattern, in which semantic and phonological properties are correlated, is quite common. There are also some forms such as *sweetbreads, bootleg* and *slapstick* in which the meanings of the component morphemes are unrelated to the meaning of the word. For traditional theories the central question is whether such words are to be treated as morphologically complex or listed in the mental lexicon. It is widely assumed that there should be a principled basis for deciding the issue, but there has been little agreement about where to draw the line<sup>3</sup>.

To summarize, English words appear to consist of subunits that combine in a productive, rule-like manner to form new words. However, there are many words that deviate in different ways and degrees from the simple and transparent ideal. Seidenberg and McClelland<sup>4</sup> coined the term 'quasiregular' to describe systems with this character. The term was initially introduced with reference to the spelling–sound correspondences of English. The pronunciations of many letter strings appear to be rule governed (e.g. *save, pave* and *gave* all rhyme), and knowledge of these regularities allows people to pronounce novel letter strings such as *mave*. However, the language also admits many exceptions (such as *have*) that deviate from the rules in different degrees. Word formation also exhibits this characteristic<sup>5</sup>.

#### Morphological structure in word recognition

In principle, words could be recognized without using morphological structure at all because the spelling and sound of a word usually provide sufficient information for this purpose. Psycholinguistic research has therefore focused on demonstrating that there is a level of morphological representation in lexical memory and that this information is used in processing. Like most of the studies in the literature, our discussion focuses on word reading, but similar issues arise in connection with other uses of language.

In an influential study by Taft and Forster<sup>6</sup>, subjects performed a lexical decision task, deciding if letter strings were words or not. The critical comparison was between pseudowords such *dejuvenate* (analyzed as containing the

stem morpheme that occurs in *rejuvenate*) and ones such as *depertoire* (which do not incorporate a stem morpheme). Subjects found it more difficult to decide that the pseudo-words containing morphological stems were not words. The findings from this and many subsequent studies suggested that recognition involves decomposing words into their component morphemes<sup>7</sup>.

A second approach involves using frequency effects to diagnose the use of morphemes. The frequency with which a word is used affects how hard it is to process<sup>8</sup>. This effect is usually taken as evidence that frequency information is stored as part of word-level representations in the mental lexicon. The same logic has been used to examine morphological units. The main prediction is that the processing of a word such as *talker* should be affected not only by its own frequency but also by how often *talk* occurs across words (e.g. *talk*, *talking*, *talked*, etc.). Units such as *talk* which underlie neighborhoods of related words are variously called roots, lemmas or stems<sup>9–11</sup>. Effects of root rather than word frequency have been reported in studies involving several languages (e.g. French<sup>10</sup>, English<sup>12</sup> and Italian<sup>13</sup>), also suggesting that recognition involves the use of morphological units.

A third approach uses the phenomenon of lexical priming, in which the processing of a word is affected by its similarity to a preceding word. Murrell and Morten<sup>14</sup> examined the priming of words that are morphologically related (e.g. *cars–car*) and ones that are only related in form (*card–car*) with respect to an unrelated baseline condition (e.g. *book–car*). Because they found priming for pairs such as *cars–car*, but not for pairs such as *card–car*, they concluded that the effect was due to morphological overlap rather than formal (orthographic or phonological) overlap. This priming methodology has also been used in many subsequent studies<sup>15–18</sup>.

The kinds of models that have been proposed to account for these effects are summarized in Box 2. Most models assume a dual-mechanism approach in which morphologically simple words such as *kangaroo* are recognized holistically and morphologically complex ones such as *government* through decomposition into morphemic primitives. In Caramazza and colleagues' version, the two mechanisms operate in parallel with a race between them<sup>19</sup> (see also Ref. 20). Recently, researchers have proposed models patterned on McClelland and Rumelhart's interactive activation model<sup>21</sup>, in which nodes corresponding to various units (e.g. syllables, morphemes, words) are connected in a hierarchical fashion and activated in the course of processing<sup>22</sup>.

This research raises several methodological and theoretical issues. The criteria for classifying words as morphologically simple or complex have often been intuitive rather than explicit and varied across studies. Taft and Forster<sup>6</sup>, for example, assumed that *dejuvenate* and *depertoire* differ because *rejuvenate* is morphologically complex and *repertoire* is not. However, the morphological status of words such as *rejuvenate* is itself unclear. Stimuli such as *reduce* or *grocer* were treated as affixed in some studies, non-affixed in others. These inconsistencies reflect the uncertainty about the nature of morphological units within

linguistic theory; they greatly complicate the task of interpreting empirical results.

The studies of frequency effects are problematic because the assumption that effects of root or word frequency show that these units are discretely represented in memory is too strong. This interpretation is a natural one in models using localist (e.g. word or morpheme) units; however, models using distributed representations account for the same effects without such levels. For example, the Seidenberg and McClelland model produced word frequency effects without word units<sup>4</sup>. In general, frequency effects for a given sized unit do not demand a corresponding discrete level of representation in distributed networks.

Perhaps the most important concern derives from a basic fact about human languages: morphology is correlated with other types of lexical information, including spelling, sound and meaning<sup>23</sup>. Effects attributed to morphology might instead be due to these correlated factors. Consider a morphologically complex word such as *government*. It can be treated as consisting of two morphemes, each of which participates in other words (e.g. *governor*, *amusement*). Some studies have been taken as providing evidence for a distinct level in lexical memory at which these morphemes are represented<sup>16,19,22</sup>. However, these units might be perceptually salient not because there is a distinct morphological level but because they happen to make consistent orthographic, phonological and semantic contributions to different words.

A number of studies have attempted to show that there are effects of morphological structure above and beyond those that are attributable to these other factors. The logic of the studies involves using comparisons across conditions or experiments intended to allow effects as a result of non-morphological factors to be partialled out. For example, Napps<sup>24</sup> observed more priming for morphologically related words (e.g. *government–govern*) than for semantically related words (e.g. *bread–cake*), a pattern taken as evidence for an effect of morphological structure beyond that due to semantic overlap. Of course, *government–govern* exhibit more orthographic and phonological overlap than *bread–cake*; the comparison controls one confounding factor (semantic overlap) but introduces another (formal overlap). To rule out the latter factor, Napps cited a separate study in which formal properties of words were controlled<sup>25</sup>. Thus *ribbed–rib* produced greater priming than *ribbon–rib*. Note, however, that these stimulus pairs differ in the degree of semantic overlap. Napps concluded that because there was priming with formal overlap controlled in one study and priming with semantic overlap controlled in another, the effects must have been due to morphology *per se* (see Ref. 16 for similar reasoning). However, the data might merely indicate that there are nonadditive effects of formal and semantic overlap. For example, formal overlap might only have an effect for pairs that are also semantically related. Many other studies have attempted to deconfound morphology and other aspects of words (e.g. Ref. 26). However, the usual strategy of designing a factorial experiment that manipulates all of the relevant factors runs into difficulty when they are intrinsically highly correlated, as in morphology. Clearly the concerns about these data

## Box 2. Main theoretical approaches to morphological processing

### Hybrid models

Some early models proposed that all words were recognized through decomposition (Ref. a), whereas others proposed that recognition was always holistic (Ref. b). Most current models are hybrids that include both processes (Fig. 1a). Models differ over which forms are decomposed and which are not; some decompose semantically transparent but not opaque forms (Ref. c), some suffixed but not prefixed (Ref. d) and some inflected but not derived (Ref. e). The models also vary in whether this decomposition applies to lexical access, to storage of complex forms, or to both (Ref. f). Some models, such as Caramazza and colleagues' Augmented Addressed Morphology Model (Ref. g), assume variable processing speeds for the two mechanisms in lexical access, creating a kind of race model (see also Ref. h). An important unifying aspect of all these models is that they assume discrete morphemes and decomposition for at least some words.

### Interactive activation models

Figure 1b is modified from Taft and Zhu (Ref. i). In their discussion of an interactive activation approach to morphology, they point out that there are several choices that must be made about representational levels, such as whether to include a separate level for bound morphemes. These models differ crucially from distributed connectionist ones in that the representations are stipulated rather than discovered during the course of learning (Ref. j).

### Distributed connectionist models

In Fig. 1c, ovals represent banks of simple, neuron-like processing units and the lines represent connections between the groups of units. Representations are patterns of activation distributed across these units and knowledge is stored in the weights on connections between them. Morphemes are not

represented as discrete entities. Instead, as the network learns to map from one domain to another (e.g. sound to meaning) it picks up on regularities in the mappings. Morphology arises as a consequence of the correlations between codes. This approach predicts that effects of semantic, phonological and orthographic similarity on morphological processing should be graded, reflecting the degree of convergence between codes.

### References

- a Taft, M. and Forster, K. (1975) Lexical storage and retrieval of prefixed words. *J. Verbal Learn. Verbal Behav.* 14, 638–647
- b Manelis, L. and Tharp, D.A. (1977) The processing of affixed words. *Mem. Cognit.* 5, 690–695
- c Marslen-Wilson, W. et al. (1994) Morphology and meaning in the English mental lexicon. *Psychol. Rev.* 101, 3–33
- d Colé, P. et al. (1989) On the representation and processing of prefixed and suffixed words: a differential frequency effect. *J. Mem. Lang.* 28, 1–13
- e Stanners, R.F. et al. (1979) Memory representation for morphologically related words. *J. Verbal Learn. Verbal Behav.* 18, 399–412
- f Andrews, S. (1986) Morphological influences on lexical access: lexical or nonlexical effects. *J. Mem. Lang.* 25, 726–740
- g Caramazza, A. et al. (1988) Lexical access and inflectional morphology. *Cognition* 28, 297–332
- h Frauenfelder, U. and Schreuder, R. (1992) Constraining psycholinguistic models of morphological processing and representation: the role of productivity. In *Yearbook of Morphology 1993* (Booij, G.E. and van Marle, J., eds), pp. 165–183, Kluwer
- i Taft, M. and Zhu, X. (1995) The representation of bound morphemes in the lexicon: a Chinese study. In *Morphological Aspects of Language Processing* (Feldman, L., ed.), pp. 293–316, Erlbaum
- j Plaut, D.C. and McClelland, J.L. Stipulating versus discovering representations. *Behav. Brain Sci.* (in press)

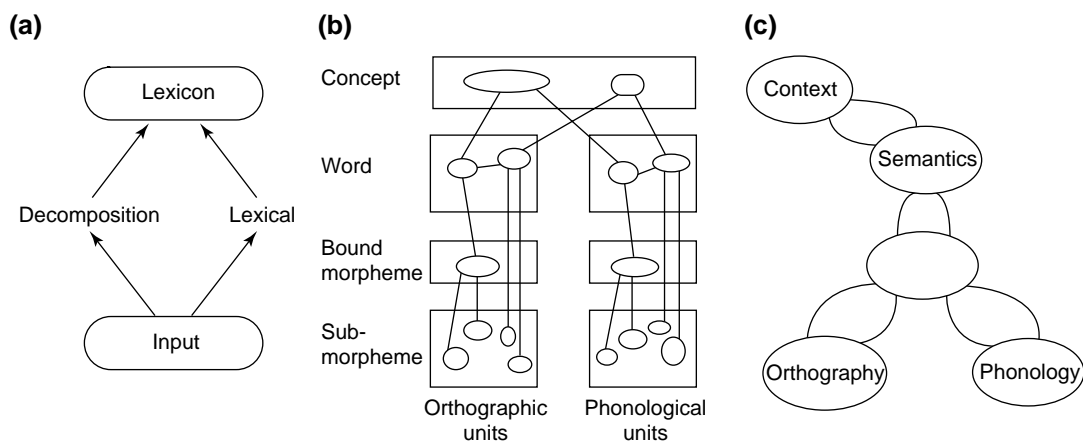


Fig. 1. Models of morphological processing. (a) Hybrid models, (b) interactive activation model and (c) distributed connectionist models (see text for details).

trends in Cognitive Sciences

also bear on the validity of theoretical models developed to explain them.

### The convergence theory

An alternative approach is provided by pursuing the parallels between derivational morphology and orthographic–phonological correspondences. As already noted, both systems are

quasiregular; there are cases that can be described as rule governed but also exceptions that deviate from the rules by differing degrees. Both are traditionally thought to require two mechanisms, a set of rules for the 'regular' cases and a word-based mechanism to handle the exceptions<sup>27,28</sup>. Seidenberg and McClelland's model showed how both the rule-governed cases and exceptions could be processed

within a single network employing distributed representations<sup>4</sup>. The model did not embody a categorical distinction between rule governed and rule violating; the weights used in pronouncing all words encoded different degrees of consistency in the spelling–sound mapping. The main reason to pursue this approach is because it captures an essential characteristic of quasiregular systems: the existence of partial regularities. A word such as *have* violates the pronunciation rules of English; however, its pronunciation is not arbitrary; it overlaps with ‘rule-governed’ forms such as *had*, *has* and *hive*. In dual-mechanism theories, the procedure by which *have* is pronounced is unrelated to the one used in pronouncing *had*, *has* and *hive*. In the connectionist approach, because there is a single set of weights, what is learned about one word carries over to partially overlapping words. This characteristic of the model allows it to explain empirical phenomena such as consistency effects: the fact that a word such as *wave* (which has an irregular neighbor, *have*) takes longer to name than *wade* (whose neighborhood is more consistent)<sup>4</sup>.

Morphology can be construed in a similar manner. The basic idea is that it is a graded, inter-level representation that develops in the course of acquiring lexical knowledge (Box 2, model c). The lexicon encodes information about the spellings, sounds and meanings of words; the regularities within these codes and between codes, plus additional information derived from the contexts in which words occur. The problem of lexical learning is framed in terms of a connectionist network employing distributed representations of these types of information. The lexical network supports computations from orthography to phonology, phonology to meaning, meaning to phonology, and so on, which are used in performing tasks such as production and comprehension.

A network trained to compute mappings between codes using standard connectionist learning techniques will pick up on the structure that is implicit in the training corpus (subject to limitations imposed by the quality of the input and output representations and other architectural constraints). In English, the codes are related in different ways. Previous research has focused on the strong though imperfect correlations between orthography and phonology in monosyllabic words, and their role in pronouncing letter strings aloud. Reading normally involves computing the meanings of words, however, and in monosyllabic words the correspondences between form and meaning are largely arbitrary. Thus, there is nothing about the meaning of *cat* that dictates its particular orthographic–phonological form. However, correspondences between form and meaning are not arbitrary in complex words; the units we think of as classical morphemes make similar semantic contributions to neighborhoods of related words (e.g. *drink*, *drinker*, *drinkable*). Note that *drink* also makes consistent orthographic and phonological contributions to these words. Importantly, the degree of consistency in the contributions of orthography, phonology and semantics to related words – that is, the degree to which these codes converge – varies across the lexicon. *Drink* makes highly similar contributions to different words and so behaves like a classical morpheme. By contrast, there is

less phonological overlap between *divine* and *divinity* because the stress rule called trisyllabic laxing has applied to the derived form; this difference is also reflected (weakly) in the orthography. Many other partial regularities between words have been created by diachronic changes in pronunciation and meaning. For example, *return* is no longer a prefixed word meaning ‘to turn again’, a fact that is also reflected in the destressing of the initial syllable *re-* (compare with *reload*). However, *return* is not semantically unrelated to *turn*, either. By contrast, *deliver* is structurally similar to *return* but semantically unrelated to *live*. In the *-mit* cases, there is a high degree of orthographic and phonological similarity across words but *-mit* makes only a weak contribution to their meanings. These and other classic morphological phenomena can be seen in terms of degrees of convergence among different types of lexical information across word neighborhoods. Typically, orthography, phonology and semantics are highly correlated, but the system admits many partial deviations for a variety of reasons, including diachronic change, lexical borrowing and historical accident. For example, *cranberry* and *grocery* are less puzzling if their etymologies are considered: *cran-* derives from the Low German *kraan*, meaning crane, apparently in reference to the beak-like shape of the plant’s stamens; *grocery* derives from the Old French *grossier*, a wholesale dealer<sup>29</sup>. However, these facts are buried in the history of the language and thus not relevant to everyday performance.

In traditional approaches, a word was either morphologically simple or complex, and the theoretical problem was how to distinguish between the two. The attempt to impose a discrete notion of morphemic compositionality misses the broader generalization about the quasiregular character of the lexicon. The word *turn* is very related to *turned*, somewhat less related to *return* and unrelated (any longer) to the name *Turner*. It would be desirable to have a knowledge representation that reflects these different degrees of relatedness. Thus, the newer approach views the issues in terms of acquiring the knowledge that allows a person to perform different linguistic tasks (e.g. comprehension or production). These tasks require learning the several codes that constitute knowledge of a word. We use an architecture that encodes this information in terms of mappings between codes. Because it is a connectionist network employing distributed representations, the same weights are used to encode the mappings for many different words. Hence the weights come to encode different degrees of similarity across words. If the architecture includes an inter-level of hidden units, it will come to represent convergences between different types of information across words. Thus, a ‘morphological’ level of representation will emerge in the course of learning to use language. This level captures the regularities attributed to units such as lemmas or affixes without representing them discretely.

This approach differs from traditional theories in several respects. First, there are no morphemes. The hidden units in Fig. 1c (Box 2) represent statistical regularities that hold across orthographic, phonological and semantic information. These regularities are not limited to those

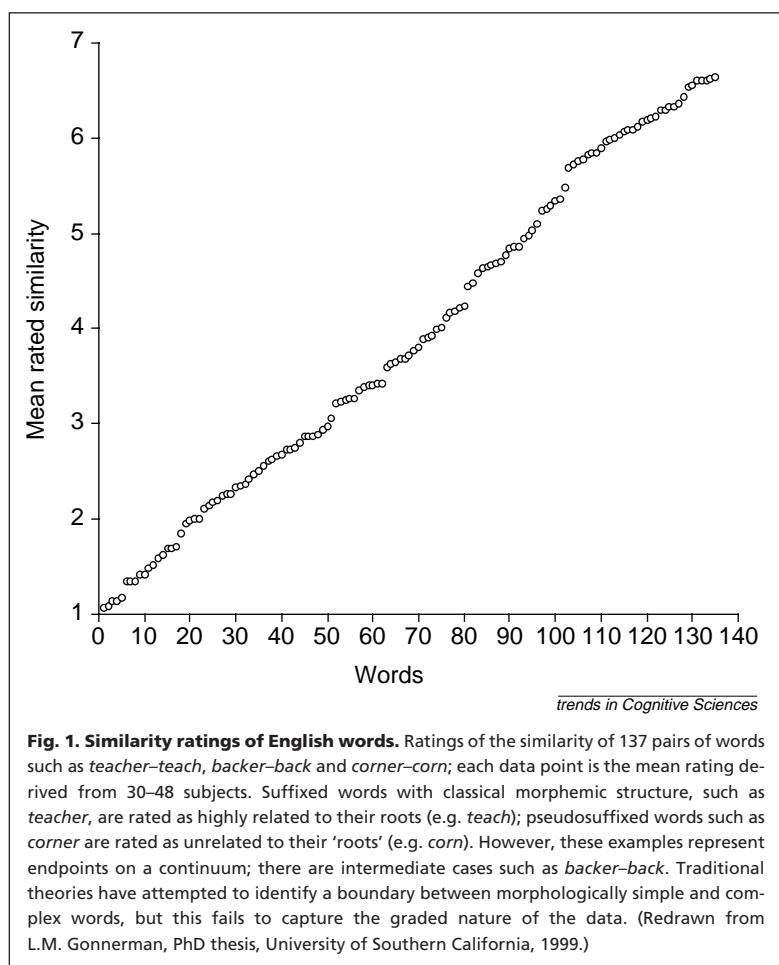
characteristic of classical morphemes. Second, morphological structure is governed by the same principles of knowledge representation, learning and processing as the other lexical codes. This contrasts with the idea of an independent morphological level or module, governed by its own domain-specific rules. Third, this inter-level representation admits different degrees of compositionality. It can therefore capture both the transparently compositional cases and the partial regularities exhibited by words such as *grocer* and *remit*. Fourth, whereas the goal of most studies was to establish morphological effects with the orthographic, phonological and semantic factors *controlled*, the approach of this article entails the opposite idea: that morphological structure is the *result* of the confluence of these factors and therefore morphological effects should be predictable from them. Finally, the theory is not derivational. In many theories, complex words are derived from a base or underlying form through the application of morphological rules, generating families of related words. The network approach proposed in this article is driven by properties of the input, which contains families of related words, whose orthographic, phonological and semantic codes converge, creating graded inter-level representations. These underlying representations are abstractions away from the surface forms of words but there is no sense in which related words are derived from a single underlying structure. There is good evidence that children are capable of learning the kind of complex

correlational structure presupposed here, and that, indeed, this kind of statistical learning provides the basis for solving problems such as discovering the boundaries between words, bootstrapping grammatical categories and learning verb argument structures<sup>30</sup>.

### Recent research

A growing body of research is consistent with this approach. Bybee provided a wealth of evidence concerning graded aspects of morphological structure<sup>31,32</sup>. Rueckl *et al.* also proposed that morphology might represent a hidden unit level of representation<sup>33</sup>. Other studies motivated by traditional approaches have yielded data consistent with the convergence theory. For example, Schreuder and Baayen conducted corpus-based analyses of morphology in Dutch, which showed that orthographic units differ in terms of how reliably they function as morphemes (e.g. prefixes)<sup>7</sup>. They took the data as evidence against lexical decomposition by heuristics such as prefix stripping<sup>6</sup> and proposed an interactive activation model in which prefix units could be activated in differing degrees. Our approach is similar except that there are no morpheme units because we claim that the regularities actually concern orthography, phonology and semantics. Laudanna and Burani's analyses of prefixes in Italian yielded similar results<sup>34</sup>. Interestingly, Laudanna *et al.* showed that statistics about how reliably a letter string functions as a prefix predicted lexical decision latencies for nonwords in Italian<sup>35</sup>. The more reliably a letter pattern occurs as a prefix, the harder it was to reject a nonword containing that pattern. They interpreted these results as indicating that frequency affects whether a prefix is established as a unit in lexical memory; as previously noted, however, the same sorts of effects are observed in connectionist models that do not include such units<sup>4</sup>. Similar results have been reported for English prefixes<sup>36</sup>. Kostić discusses findings that led him to abandon a lexical decomposition approach to Serbo-Croatian morphology in favor of an 'informational load' account in which probabilistic aspects of morphemes (e.g. the number of thematic roles associated with a given case) affect the extent to which they function as units<sup>37</sup>.

Gonnerman presented four experiments showing that 'morphological' priming effects are derived from semantic and phonological factors (L.M. Gonnerman, PhD thesis, University of Southern California, 1999). One study examined the effects of degree of semantic relatedness between a prime and a target. The stimuli were pairs such as *baker-bake* and *backer-back* that vary in terms of semantic relatedness but were similar in terms of formal overlap (see Fig. 1). Priming effects (latencies for related pairs compared with unrelated controls such as *radar-corn*) were larger for highly related pairs such as *baker-bake* (40 ms) than for moderately related pairs such as *backer-back* (19 ms). Semantically unrelated pairs such as *corner-corn* or *spinach-spin* did not yield significant priming. A second study examined effects of the degree of phonological overlap between prime and target for semantically related pairs. Priming was greater for pairs with only a consonant change, such as *deletion-delete* (65 ms), compared with



those with a vowel change *criminal–crime* (48 ms), which was in turn greater than priming for pairs with both a vowel and a consonant change, such as *introduction–introduce* (35 ms). The studies indicate that priming effects for ‘morphologically’ related words derive from the degree of semantic and phonological overlap between prime and target. It follows from this view that pairs such as *jubilant–jubilee* and *trifle–trivial*, which are morphologically and etymologically unrelated, should produce priming effects similar in magnitude to morphologically related pairs with the same degree of semantic and phonological similarity, a result that Gonnerman also reported (L.M. Gonnerman, PhD thesis).

Finally, this approach is consistent with the existence of other types of non-arbitrary form–meaning correspondences in languages. Above, the common observation that the correspondences between form (e.g. *cat*) and meaning (small domestic feline) are arbitrary is repeated; this is sometimes taken as an important universal property of languages<sup>28</sup>. However, this generalization is itself only quasiregular because languages violate it in different ways and to different degrees. For example, ‘sound symbolism’ is the familiar case in which for historical or accidental reasons, neighborhoods of words overlap in both form and meaning (e.g. *glisten–glimmer–gleam–glitter; sneer–sneeze–sniff–snout–snort*). These forms also produce significant priming effects related to degree of semantic and phonological overlap<sup>38</sup>. More subtle cases concern correlations between phonology and grammatical category, whose roles in language acquisition and processing are the focus of intensive study<sup>39</sup>. Cassidy *et al.* examined phonological differences between male and female names<sup>40</sup>; male names tend to have fewer syllables and end more often in stop consonants (compare *Eric* and *Erica*). They developed a connectionist model that learned to classify names accurately as male or female using phonological cues. All of these phenomena illustrate the non-arbitrary, probabilistic relationships between phonological and other lexical codes that languages exhibit. These regularities are not morphological in the classical sense, but can be subsumed by our approach, in which the lexicon encodes correlations over many different types of information simultaneously.

### Current issues

#### *Are statistics sufficient?*

In the convergence theory, morphology is given a statistical interpretation: it corresponds to statistical regularities in the mappings among spelling, sound and meaning. In traditional theories, morphemes are defined structurally rather than statistically; people might encode facts such as how often morphemes are encountered or how similar they are to each other, but the status of a unit as a morpheme does not depend on these factors. The same holds for other grammatical structures as well. Thus, the traditional view asserts that morphological structure has properties not predictable from mere statistics. It therefore naturally suggests that if the statistical properties of words were equated, there would be residual effects owing to morphological structure *per se*.

This issue is likely to be a major focus of future research. Here, we briefly note that deciding the issue will require a more explicit theory concerning which statistics are relevant. A further complication is that connectionist networks do not represent the statistics of the input exactly; rather, they perform a transformation of the input data<sup>41</sup>.

Finally, the statistics of words can be transiently affected by experimental manipulations, such as repetitions of patterns across many stimulus items. The studies by Rapp provide an illustration<sup>42</sup>. She carefully controlled the bigram characteristics of words in different conditions and found that words were decomposed into structurally defined units, whether the bigram statistics favored this division or not. However, the experimental design involved the repetition of syllables and morphemes multiple times across subjects (e.g. the stimuli in one condition included *wilma, sigma, dogma, vodka, and magma*; another included *unable, uneasy, unkind, untidy, unseen, and untied*), which alters the statistics for these patterns. All of these considerations suggest that taking into account the statistical properties of words complicates the task of performing critical experiments; however, it is equally clear that this information cannot be ignored.

#### *Lexical syntax*

Words consist of more than just spellings, sounds and meanings; they also have syntactic properties that need to be addressed. Words fall into grammatical categories such as noun and verb, which derivational but not inflectional rules might change. Thus, *bake* and *baked* are both verbs, but *baker* is a noun and *bakeable* is an adjective. Moreover, the components of words bear particular relationships to each other, which affect their interpretation. A compound such as *houseboat*, for example, consists of a head, *boat*, preceded by a modifier, *house*; whereas in *boathouse* the roles are reversed and thus the meaning different. This aspect of morphology raises important issues about the characterization of syntactic structure that extend beyond the scope of this article. We have emphasized the importance of the formal and semantic regularities that hold across words such as *baker, taker, and runner*. Syntax involves further regularities concerning the distributions of words in sentences. Thus *baker, taker, and runner* are also similar with respect to their distributional properties. Grammatical categories such as noun may reflect the conjunction of these lexical and distributional facts. Similarly, *boathouse* exhibits a modifier-head relationship that is also seen in adjectival phrases such as *new house*, which also exhibit systematic distributional properties. Thus, it may be fruitful to embed the lexical processor described here within a network that processes sequences of words. How far this approach will get in explaining syntactic phenomena is, of course, unknown.

#### *Cross-linguistic differences?*

Languages vary considerably in morphological richness. English morphology is usually considered to be impoverished compared with languages such as Serbo–Croatian, Italian or Hebrew. It therefore cannot be assumed that what holds for English will necessarily apply in other

languages. However, one of the principal claims of modern linguistic theory is that all languages are similar insofar as they exhibit universal properties, with the degree of variation across languages being quite constrained. Moreover, languages are used by people whose perceptual, learning and memory capacities are the same. These considerations suggest that there should be similarities across languages with respect to morphology despite the apparent differences between them.

Whether the convergence theory will extend to languages other than English is not known. At this point three observations seem relevant. First, experiments to date have not tended to yield radically different patterns of results across languages. On the contrary, what is striking is the extent to which the results have been similar (compare, for example, Refs 16 and 17). Second, the methodological issues that have arisen in studies of English also apply to studies in other languages. There are significant practical barriers here; the large corpora of English words that allow researchers to calculate different statistics and develop computational models are not as yet available for languages such as Hebrew. This makes it harder to do certain types of experiments and simulations in these languages. Third, although there are differences between languages in terms of morphological complexity and types of morphological mechanisms, they exhibit deeper commonalities. Consider the non-concatenative morphology of Hebrew. The fact that most Hebrew words consist of a root (usually three consonants, sometimes two or four) from which many related words can be derived and the fact that morphemes are intercalated within the triconsonantal roots (e.g. *gdl* + *\_o\_e\_* = *godel*; *gdl* + *mi\_a\_* = *migdal*) might seem to demand a different approach than for English<sup>43</sup>. However, Hebrew exhibits the same pattern of central tendencies-plus-deviations-in-different-degrees as seen in English. For example, verbs occur in seven basic morphological forms called binyans. The binyans are highly predictive of the semantics of the resulting forms; however, not all verbs appear in all binyans and some verbs that appear in a given binyan deviate from the expected meaning. There are also subforms of the binyans that introduce further partial regularities. In short, Hebrew morphology exhibits quasiregular characteristics, suggesting that it might be governed by the same principles as English, consistent with the results of recent simulation studies that have compared English and Hebrew<sup>44</sup>.

### Outstanding questions

- Will implemented connectionist models trained on realistic corpora develop appropriate inter-level representations?
- Can the same approach be applied to inflectional morphology and other aspects of morphological knowledge?
- Can morphological impairments following brain injury be explained in terms of damage to a network that does not have discrete morphological units?
- Can facts about the acquisition of morphological knowledge be explained in terms of the course of learning within such networks?
- Will the approach extend crosslinguistically, accounting for data from typologically distinct languages that exhibit morphological regularities not present in English (e.g. reduplication)?

### Comprehension versus production

A final issue is whether the same principles are at work in word recognition and production. Intuitively the tasks seem intimately related: learning a word typically involves learning both how to produce and understand it. The meaning and grammatical properties of a word do not change as a function of whether it is spoken, written or read. It would therefore be odd if completely different principles underlay these different aspects of performance. Studies of production and comprehension have nonetheless tended to be pursued independently, with different models, experimental methods and theoretical approaches. In a recent review, Levelt summarized models of word production that contrast with the approach proposed here in several respects<sup>9</sup>. Most research on word production assumes multiple discrete levels of linguistic representation, including morphology. This has proved useful in developing accounts of a range of facts about speech errors and the time course of word versus picture processing. The extent to which these representations of word structure might emerge in the course of learning to use language for both comprehension and production is unclear; very little has been said in the speech production literature about how the representations assumed by these models are acquired. It is also unclear whether any of the phenomena that have been explained to date demand the use of localist representations. We are doubtful that they do: distributed representations are always localist at some level, and it is likely that distributed models can mimic the behavior of localist ones to a very high degree of precision. Nonetheless, the issue is a controversial one and the subject of ongoing debate. At this point what can be said is that current theories of production have a different orientation than the one we have proposed for recognition; determining whether the same principles apply across the two domains is an important unresolved issue.

### Conclusions

In this article, we have put forward a non-discrete approach to derivational morphology as an alternative to classical approaches. The important point to emphasize at this early stage in the development of the convergence theory is the quasiregular character of derivational morphology. The main reason for pursuing the approach we have described is not because there is something intrinsically correct about multilayer networks that employ distributed representations or incorrect about other approaches; rather, it is because the quasiregular character of morphology seems to demand the kinds of graded representations that the convergence theory entails. This approach is a radical departure from linguistic tradition and it remains to be seen whether it can account for the mass of crosslinguistic data that have accumulated. The practical and conceptual obstacles to developing explanatory models of nontrivial scope are also considerable. The issues raised here are therefore likely to play out for some time to come.

### Acknowledgements

We thank our colleagues Elaine Andersen, Jay McClelland, David Plaut and Maryellen MacDonald for sharing insights about these issues.

## References

- 1 Hockett, C.F. (1958) *A Course in Modern Linguistics*, Macmillan
- 2 Aronoff, M. (1976) *Word Formation in Generative Grammar*, MIT Press
- 3 Spencer, A. (1991) *Morphological Theory*, Blackwell
- 4 Seidenberg, M.S. and McClelland, J.L. (1989) A distributed, developmental model of word recognition and naming. *Psychol. Rev.* 96, 447–452
- 5 Taft, M. (1998) Lexical representation of gender: a quasiregular domain. *J. Psycholinguist. Res.* 27, 23–45
- 6 Taft, M. and Forster, K. (1975) Lexical storage and retrieval of prefixed words. *J. Verbal Learn. Verbal Behav.* 14, 638–647
- 7 Schreuder, R. and Baayen, R. (1994) Prefix-stripping re-revisited. *J. Mem. Lang.* 33, 357–375
- 8 Forster, K. and Chambers, S. (1973) Lexical access and naming time. *J. Verbal Learn. Verbal Behav.* 12, 627–635
- 9 Levelt, W.J.M. (1999) Models of word production. *Trends Cognit. Sci.* 3, 223–232
- 10 Meunier, F. and Segui, J. (1999) Frequency effects in auditory word recognition: the case of suffixed words. *J. Mem. Lang.* 41, 327–344
- 11 Burani, C. et al. (1984) Morphological structure and lexical access. *Visible Lang.* 18, 419–427
- 12 Taft, M. (1979) Lexical access via an orthographic code: the basic orthographic syllable structure (BOSS). *J. Verbal Learn. Verbal Behav.* 18, 21–40
- 13 Burani, C. and Caramazza, A. (1987) Representation and processing of derived words. *Lang. Cognit. Processes* 2, 217–227
- 14 Murrell, G.A. and Morton, J. (1974) Word recognition and morphemic structure. *J. Exp. Psychol.* 102, 963–968
- 15 Feldman, L.B. (1994) Beyond orthography and phonology: differences between inflections and derivations. *J. Mem. Lang.* 33, 442–470
- 16 Marslen-Wilson, W. et al. (1994) Morphology and meaning in the English mental lexicon. *Psychol. Rev.* 101, 3–33
- 17 Frost, R. et al. (1997) What can we learn from the morphology of Hebrew. *J. Exp. Psychol. Learn. Mem. Cognit.* 23, 829–856
- 18 Raveh, M. and Rueckl, J. (2000) Equivalent effects of inflected and derived primes: long-term morphological priming in fragment completion and lexical decision. *J. Mem. Lang.* 42, 103–119
- 19 Caramazza, A. et al. (1988) Lexical access and inflectional morphology. *Cognition* 28, 297–332
- 20 Frauenfelder, U.H. and Schreuder, R. (1992) Constraining psychological models of morphological processing and representation: the role of productivity. In *Yearbook of Morphology 1991* (Booij, G.E. and van Marle, J., eds), pp. 165–183, Kluwer
- 21 McClelland, J.L. and Rumelhart, D.E. (1981) An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychol. Rev.* 88, 375–407
- 22 Taft, M. (1995) Interactive-activation as a framework for understanding morphological processing. *Lang. Cognit. Processes* 9, 271–294
- 23 Seidenberg, M.S. (1987) Sublexical structures in visual word recognition: access units or orthographic redundancy? In *Attention and Performance Vol. XII: Reading* (Coltheart, M., ed.), pp. 245–263, Erlbaum
- 24 Napps, S.E. (1989) Morphemic relationships in the lexicon: are they distinct from semantic and formal relationships? *Mem. Cognit.* 17, 729–739
- 25 Napps, S.E. and Fowler, C.A. (1987) Formal relationships among words and the organization of the mental lexicon. *J. Psycholinguist. Res.* 16, 257–272
- 26 Feldman, L.B., ed. (1995) *Morphological Aspects of Language Processing*, Erlbaum
- 27 Coltheart, M. et al. (1993) Models of reading aloud: dual-route and parallel distributed processing approaches. *Psychol. Rev.* 100, 589–608
- 28 Pinker, S. (1991) Rules of language. *Science* 253, 530–535
- 29 Morris, W., ed. (1981) *American Heritage Dictionary of the English Language*, Houghton-Mifflin
- 30 Seidenberg, M.S. (1997) Language acquisition and use: learning and applying probabilistic constraints. *Science* 275, 1599–1604
- 31 Bybee, J.L. (1985) *Morphology: A Study of the Relation Between Meaning and Form*, John Benjamins
- 32 Bybee, J. (1995) Regular morphology and the lexicon. *Lang. Cognit. Processes* 10, 425–455
- 33 Rueckl, J. et al. (1997) Morphological priming, fragment completion, and connectionist networks. *J. Mem. Lang.* 36, 382–405
- 34 Laudanna, A. and Burani, C. (1995) Distributional properties of derivational affixes: implications for processing. In *Morphological Aspects of Language Processing* (Feldman, L.B., ed.), pp. 345–364, Erlbaum
- 35 Laudanna, A. et al. (1997) Morpho-lexical representations in naming. *Lang. Cognit. Processes* 12, 49–66
- 36 Stolz, J. and Feldman, L. (1995) The role of orthographic and semantic transparency of the base morpheme in morphological processing. In *Morphological Aspects of Language Processing* (Feldman, L.B., ed.), pp. 109–129, Erlbaum
- 37 Kostić, A. (1995) Informational load constraints on processing inflected morphology. In *Morphological Aspects of Language Processing* (Feldman, L.B., ed.), pp. 317–344, Erlbaum
- 38 Gonnerman, L.M. et al. (1999) Graded semantic and phonological effects in lexical processing: implications for morphology. *Psychonomic Soc. Abstr.* 4, 44
- 39 Kelly, M.H. (1992) Using sound to solve syntactic problems: the role of phonology in grammatical category assignments. *Psychol. Rev.* 99, 349–364
- 40 Cassidy, K.W. et al. (1999) Inferring gender from name phonology. *J. Exp. Psychol. Gen.* 128, 362–381
- 41 Bishop, C.M. (1995) *Neural Networks for Pattern Recognition*, Clarendon Press
- 42 Rapp, B. (1992) The nature of sublexical orthographic organization: the bigram trough hypothesis examined. *J. Mem. Lang.* 31, 33–53
- 43 McCarthy, J. (1981) A prosodic theory of nonconcatenative morphology. *Linguist. Inq.* 12, 373–418
- 44 Plaut, D.C. and Gonnerman, L.M. Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Lang. Cognit. Processes* (in press)

## Do you want to reproduce material from *Trends in Cognitive Sciences*?

This publication and the individual contributions within it are protected by the copyright of Elsevier Science Ltd. Except as outlined in the copyright statement (see p. V), no part of *Trends in Cognitive Sciences* may be reproduced, either in print or electronic form, without written permission from Elsevier Science Ltd. Please address any permission requests to:

Elsevier Science Global Rights Department, PO Box 800, Oxford, UK OX5 1DX