To appear in Trends in Cognitive Sciences

Explaining Derivational Morphology As The Convergence of Codes

Mark S. Seidenberg

University of Southern California

Laura M. Gonnerman

Carnegie Mellon University

Running Head: Convergent Morphology

Address for correspondence:
Mark S. Seidenberg
Neuroscience Program
University of Southern California
Los Angeles, CA 90089-2520

marks@gizmo.usc.edu
213-740-9175
fax: 213-740-5687

Abstract

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, interlevel 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 GOVERNALMENT could not be. Language users' 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 that are recombined by rules. In this article we briefly review some of the limitations of this approach and offer an alternative, inspired by connectionist theories of knowledge representation and learning, in which graded, nondiscrete morphological structures emerge in the course of learning relations among the spellings, sounds, and meanings of words.

Our goal in this Opinion piece 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 in order. The article emphasizes phenomena concerning derivational morphology in English; issues concerning other types of morphology (Box 1) and cross-linguistic 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 we think 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[1] 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. We will call this approach *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, HITTER, 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 into question by Aronoff[2] 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 DISCOVER's pronunciation 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 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[3].

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[4] 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.[5]

Morphological Structure In Word Recognition

In principle words could be recognized without using morphological structure at all because spelling and sound 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[6], 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 (not incorporating a stem morpheme). Subjects found it more difficult to decide that the pseudowords containing morphological stems were not words. The findings from this and many subsequent studies suggested that recognition involves decomposing words into component morphemes.[7]

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[8]. This effect is standardly 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 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[9-11]. Effects of root rather than word frequency have been reported in studies involving several languages, (e.g., French[10], English[12], Italian[13]), also suggesting that recognition involves the use of morphological units.

A third approach utilizes the phenomenon of lexical priming, in which the processing of a word is affected by its similarity to a preceding word. Murrell and Morten[14] 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[15-18].

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 wholistically and morphologically complex ones such as GOVERNMENT through decomposition into morphemic primitives. In Caramazza and colleagues' version[19], the two mechanisms operate in parallel with a race between them (see also [20]). Recently researchers have proposed models patterned on McClelland and Rumelhart's interactive activation model,[21] 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.[22]

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[6], 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, nonaffixed 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[4] produced word frequency effects without word units. 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[23]. Effects attributed to morphology may 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 units are represented.[16,19,22] However, the units may 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 due to nonmorphological factors to be partialled out. For example, Napps[24] 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[25]. Thus RIBBED—RIB produced greater priming than RIBBON—RIB. Note, however, that these stimuli differ in semantic overlap. Napps concluded that because there was priming with formal overlap controlled in one study and with semantic overlap controlled in another, the effects must have been due to morphology per se (see[16] for similar reasoning). However, the data may 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., [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 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 we have noted, both systems are quasiregular; there are cases that can be described as rule-governed but also exceptions that deviate from the rules in 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.[27,28] Seidenberg and McClelland's model[4] showed how both the rule-governed cases and exceptions could be processed within a single network employing distributed representations. 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, HAVE's 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).[4]

Morphology can be construed in a similar manner. The basic idea is that it is a graded, interlevel 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 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 utilized 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 simple words the correspondences between form and meaning are largely arbitrary. Thus there is nothing about the meaning of CAT that dictates that it take this 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—i.e., 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. In 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 to RELOAD). However, RETURN is not semantically unrelated to TURN, either. DELIVER, in contrast, is structurally similar to RETURN but semantically unrelated to LIVER. 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.[29] 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. 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. 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 interlevel 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 the Box 2c model represent statistical regularities that hold across orthographic, phonological and semantic information. These regularities are not limited to the ones 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 interlevel 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*, our approach 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 here is driven by properties of the input, which contains families of related words, whose orthographic, phonological, and semantic codes converge, creating graded interlevel 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[30].

Recent Research

A growing body of research is consistent with this approach. Bybee[31-32] provided a wealth of evidence concerning graded aspects of morphological structure. Rueckl et al.[33] also proposed that morphology might represent a hidden unit level of representation. Other studies motivated by traditional approaches have yielded data consistent with the convergence theory. For example, Schreuder and Baayen[7] 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). They took the data as evidence against lexical decomposition by heuristics such as prefix-stripping[6] 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[34] analyses of prefixes in Italian yielded similar results. Interestingly, Laudanna et al.[35] showed that statistics about how reliably a letter string functions as a prefix predicted lexical decision latencies for nonwords in Italian. 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 gets 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[4]. Similar results have been reported for English prefixes.[36] Kostiç[37] 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.

Gonnerman[38] presented 4 experiments showing that "morphological" priming effects derive from semantic and phonological factors. The stimuli in one study were pairs such as BAKER-BAKE and BACKER-BACK that varied in terms of semantic relatedness but were similar in terms of formal overlap (see Figure 1). Priming effects (latencies for related pairs compared to 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 to 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[38].

Finally, this approach is consistent with the existence of other types of nonarbitrary form-meaning correspondences in language. Above we repeated the common observation that the correspondences between form (e.g., CAT) and meaning (small domestic feline) are arbitrary; this is sometimes taken as an important universal property of languages[28]. However, this generalization is itself only quasiregular because languages violate it in different ways and 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[39]. More subtle cases concern correlations between phonology and grammatical category, whose roles in language acquisition and processing are the focus of intensive study[40] Cassidy et al.[41] examined phonological differences between male and female names; 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 accurately classify names as male or female using phonological cues. All of these phenomena illustrate the nonarbitrary, 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

Our approach is new and it raises a number of unresolved 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 may 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 due to morphological structure per se.

This issue is likely to be the focus of future research. Here we briefly note that deciding the issue will require developing 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.[42] Finally, the statistics of words may be transiently affected by experimental manipulations, such as repetitions of patterns across many stimulus items. The studies by Rapp[43] provide an illustration. 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 repetition of syllables and morphemes multiple times across items (e.g., the stimuli in one condition included WILMA, SIGMA, DOGMA, VODKA, and MAGMA; another included UNABLE, UNEASY, UNKIND, UNTIDY, UNSEEN, UNTIED, etc.), which alters the statistics for these words. 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 may change. Thus, BAKE and BAKED are both verbs, but BAKER is a noun, and BAKEABLE is an adjective. Morever, the components of words bear particular relationships to each other that 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. These phenomena raise important issues about syntactic representation that clearly extend well beyond the scope of this article. Aside from acknowledging their importance, we would point out that progress might be made by embedding a lexical processing system such as the one we have described within a device that tracks the distributions of words in sentences. Nouns and verbs, for example, differ systematically in meaning, a fact we would expect the Box 2c network to pick up; however, nouns also differ from verbs in terms of the contexts in which they occur, a fact that may be picked up by a network encoding word sequences. 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 to languages such as Serbo-Croatian, Italian, or Hebrew. It therefore cannot be assumed that what holds for English will necessarily apply in other languages. On the other hand, one of the principal claims of modern linguistic theory is that languages are similar insofar as they exhibit universal properties, with the degree of variation across languages quite limited. 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 other languages 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. To the contrary, what is striking is the extent to which the results have been similar (compare, e.g., [16,17]). 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 mrophology of Hebrew. The fact that most Hebrew words consist of a root (usually 3 consonants, sometimes 2 or 4) 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[44]. However, Hebrew exhibits the same pattern of central tendencies-plus-deviations-in-different-degrees seen in English. For example, verbs occur in 7 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 comparing English and Hebrew[45].

Comprehension vs. 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 comprehend 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[9] summarized models of word production that contrast with the approach proposed here in several respects. 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 vs. 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

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 employing 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 cross-linguistic 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.

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, Basil Blackwell

4 Seidenberg, M.S. and McClelland, J.L. (1989) A distributed, developmental model of word recognition and naming. PSYCHOLOGICAL REVIEW 96, 447-452

5 Taft, M. (1998) Lexical representation of gender: A quasiregular domain. JOURNAL OF PSYCHOLINGUISTIC RESEARCH 27, 23-45

6 Taft, M. and Forster, K. (1975) Lexical storage and retrieval of prefixed words. JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 14, 638-647

7 Schreuder, R., and Baayen, R. (1994) Prefix-stripping re-revisited. JOURNAL OF MEMORY AND LANGUAGE 33, 357-375

8 Forster, K. and Chambers, S. (1973) Lexical access and naming time. JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 12, 627-635

9 Levelt, W.J.M. (1999) Models of word production. TRENDS IN COGNITIVE SCIENCES 3, 223-232

10 Meunier, F. and Segui, J. (1999) Frequency effects in auditory word recognition: The case of suffixed words. JOURNAL OF MEMORY AND LANGUAGE 41 327-344

11 Burani, C. et al. (1984) Morphological structure and lexical access. VISIBLE LANGUAGE 18, 419-427

12 Taft, M. (1979) Lexical access via an orthographic code: The basic orthographic syllable structure (BOSS). JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 18, 21-40

13 Burani, C. and Caramazza, A. (1987) Representation and processing of derived words. LANGUAGE AND COGNITIVE PROCESSES 2, 217-227

14 Murrell, G.A. and Morton, J. (1974) Word recognition and morphemic structure. JOURNAL OF EXPERIMENTAL PSYCHOLOGY 102, 963-968

15 Feldman, L.B. (1994) Beyond orthography and phonology: Differences between inflections and derivations. JOURNAL OF MEMORY AND LANGUAGE 33, 442-470

16 Marslen-Wilson, W. et al. (1994) Morphology and meaning in the English mental lexicon. PSYCHOLOGICAL REVIEW 101, 3-33

17 Frost, R. et al. (1997) What can we learn from the morphology of Hebrew. J. EXPERIMENTAL PSYCHOLOGY: LEARNING, MEMORY AND COGNITION 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. JOURNAL OF MEMORY AND LANGUAGE 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. PSYCHOLOGICAL REVIEW 88, 375-407

22 Taft, M. (1995) Interactive-activaiton as a framework for understanding morphological processing. LANGUAGE AND COGNITIVE PROCESSES 9, 271-194

23 Seidenberg, M.S. (1987) Sublexical structures in visual word recognition: access units or orthographic redundancy? In ATTENTION AND PERFORMANCE 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? MEMORY AND COGNITION 17, 729-739

25 Napps, S.E. and Fowler, C.A. (1987) Formal relationships among words and the organization of the mental lexicon. JOURNAL OF PSYCHOLINGUISTIC RESEARCH 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. PSYCHOLOGICAL REVIEW 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 Co

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.L. (1995) Regular morphology and the lexicon. LANGUAGE AND COGNITIVE PROCESSES 10, 425-455

33 Rueckl, J. et al. (1997) Morphological priming, fragment completion, and connectionist networks. JOURNAL OF MEMORY AND LANGUAGE 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. LANGUAGE AND COGNITIVE 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 Kostiç, A. 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. (1999) Morphology and the lexicon: Exploring the semantics-phonology interface. Doctoral dissertation, Department of Linguistics, University of Southern California

39 Gonnerman, L.M. et al. (1999) Graded semantic and phonological effects in lexical processing: Implications for morphology. ABSTRACTS OF THE PSYCHONOMIC SOCIETY 4, 44.

40 Kelly, M.H. (1992) Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. PSYCHOLOGICAL REVIEW 99, 349-364

41 Cassidy, K.W. et al. (1999) Inferring gender from name phonology. JOURNAL OF EXPERIMENTAL PSYCHOLOGY: GENERAL 128, 362-381

42 Bishop, Christopher M. (1995) NEURAL NETWORKS FOR PATTERN RECOGNITION, Clarendon Press

43 Rapp, B. (1992) The nature of sublexical orthographic organization: The bigram trough hypothesis examined. JOURNAL OF MEMORY AND LANGUAGE 31, 33-53

44 McCarthy, J. (1981) A prosodic theory of nonconcatenative morphology. LINGUISTIC INQUIRY 12, 373-418

45 Plaut, D.C. and Gonnerman, L.M. Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing. LANGUAGE AND COGNITIVE PROCESSES (in press)

Acknowledgements. We thank Elaine Andersen for her substantial contribution to the development of these ideas, and our colleagues Jay McClelland, David Plaut, and Maryellen MacDonald for helpful suggestions.

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[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[b] . However, other theorists see derivation and inflection as differing only in degree[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 may 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 the two types of morphology, consistent with our approach in which the same principles apply to both.

References

a Spencer, A. (1991) MORPHOLOGICAL THEORY, Basil 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

Box 2: Main Theoretical Approaches to Morphological Processing

----

Figure 2 here

HYBRID MODELS. Some early models proposed that all words are recognized through decomposition[a], whereas others proposed that recognition is always wholistic[b]. Most current models are hybrids that include both processes. Models differ as to which forms are decomposed and which are not; some decompose semantically transparent but not opaque forms[c], some suffixed but not prefixed[d], and some inflected but not derived[e]. The models also vary in whether this decomposition applies to lexical access, to storage of complex forms, or to both[f]. Some models, such as Caramazza and colleagues[g] Augmented Addressed Morphology model, assume variable processing speeds for the two mechanisms in lexical access, creating a kind of race model (see also[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. The figure drawn here is modified from Taft and Zhu[i]. In their discussion of an interactive activation approach to morphology, they point out that there are several choices which 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[j].

DISTRIBUTED CONNECTIONIST MODELS. In this figure, 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. JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 14, 638-647

b Manelis, L. and Tharp, D.A. (1977) The processing of affixed words. MEMORY AND COGNITION 5, 690-695

c Marslen-Wilson, W. et al. (1994) Morphology and meaning in the English mental lexicon. PSYCHOLOGICAL REVIEW 101, 3-33

d Colé, P. et al. (1989) On the representation and processing of prefixed and suffixed words: A differential frequency effect. JOURNAL OF MEMORY AND LANGUAGE 28, 1-13

e Stanners, R.F. et al. (1979) Memory representation for morphologically related words. JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 18, 399-412

f Andrews, S. (1986) Morphological influences on lexical access: Lexical or nonlexical effects. JOURNAL OF MEMORY AND LANGUAGE 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.B., ed.), pp. 293-316, Erlbaum

j Plaut, D.C. and McClelland, J.L. Stipulating versus discovering representations. BEHAVIORAL AND BRAIN SCIENCES (in press)

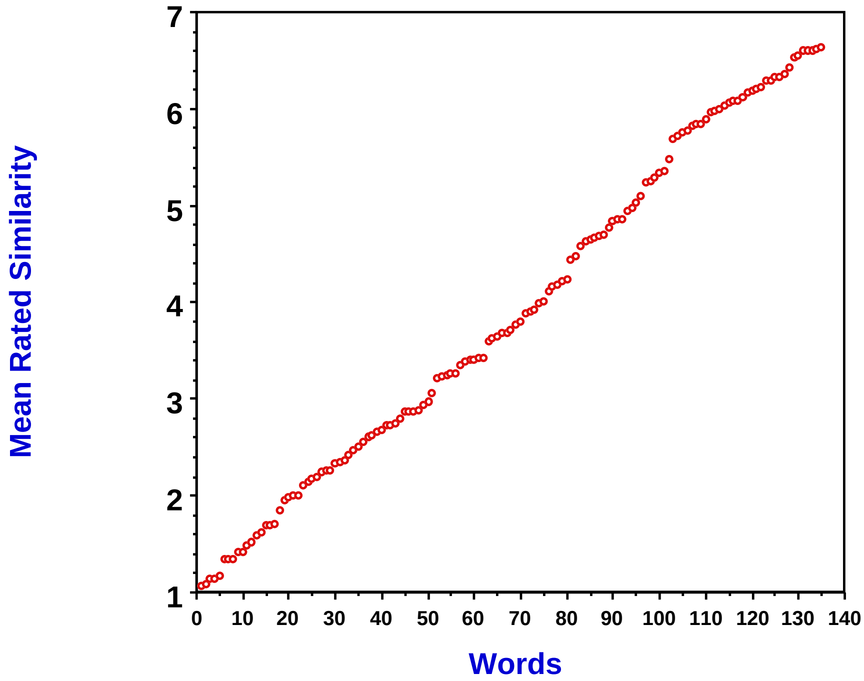

Figure 1: Ratings of the similarity of 137 pairs such as TEACHER-TEACH, BACKER-BACK, and CORNER-CORN (data from Gonnerman[38]). Each data point is the mean rating derived from 30-48 subjects. Suffixed words with classical morphemic structure, such as TEACHER, are rated as highly related to their roots (e.g., TEACH); pseudosuffixed words such as CORNER are rated as unrelated to their "roots" (e.g., CORN). However, these examples represent endpoints on a continuum; there are intermediate cases such as BACKER-BACK. Traditional theories involve finding criteria for imposing a boundary between simple and complex words but fail to capture the graded nature of the data.

Outstanding Questions

• Will implemented connectionist models trained on realistic corpora develop appropriate interlevel 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 cross-linguistically, accounting for data from typologically distinct languages that exhibit morphological regularities not present in English (e.g., reduplication)?