

Learning and morphological change

Mary Hare^{a,*}, Jeffrey L. Elman^b

^a*Center for Research in Language, University of California, San Diego, 9500 Gilman Drive,
La Jolla, CA 92093–0526, USA*

^b*Department of Cognitive Science, University of California, San Diego,
La Jolla, CA 92093–0515, USA*

Received May 18, 1994, final version accepted October 17, 1994

Abstract

An account is offered of change over time in English verb morphology, based on a connectionist approach to how morphological knowledge is acquired and used. A technique is first described that was developed for modeling historical change in connectionist networks, and that technique is applied to model English verb inflection as it developed from the highly complex past tense system of Old English towards that of the modern language, with one predominant “regular” inflection and a small number of irregular forms. The model relies on the fact that certain input–output mappings are easier than others to learn in a connectionist network. Highly frequent patterns, or those that share phonological regularities with a number of others, are learned more quickly and with lower error than low-frequency, highly irregular patterns. A network is taught a data set representative of the verb classes of Old English, but learning is stopped before reaching asymptote, and the output of this network is used as the teacher of a new net. As a result, the errors in the first network were passed on to become part of the data set of the second. Those patterns that are hardest to learn led to the most errors, and over time are “regularized” to fit a more dominant pattern. The results of the networks simulations were highly consistent with the major historical developments. These results are predicted from well-understood aspects of network dynamics, which therefore provide a rationale for the shape of the attested changes.

1. Introduction

The forms that languages take are many and varied, and one of the goals of linguistics is to try to understand the sources of such diversity. One way

* Corresponding author; e-mail: hare@crl.ucsd.edu.

to understand why languages assume the forms they do is to look at historical change. Snapshots of language behavior at any point in time may be less revealing than tracing the path of that behavior as it evolves over time. In a paper called “How do languages get crazy rules?” Bach and Harms (1972) make exactly this point. They show, with a number of examples, how synchronically complex and apparently arbitrary rules may result from an earlier set of simpler rules that are well motivated from a phonological viewpoint; the effect of historical change can be to render the environments for rules opaque and bring about reanalysis, leading to more complex – sometimes “crazy” – rule systems.

While this does not imply that synchronic grammars recapitulate change in the strong sense sometimes assumed by early works in generative grammar (e.g., Chomsky & Halle, 1968), the forces which give rise to change must presumably be present synchronically. These forces may arise from a variety of reasons, which include learnability considerations (Kiparsky, 1968; Slobin, 1977) as well as sociolinguistic factors (Labov, 1980). Labov, for example, has argued that “the same mechanisms which operated to produce the large-scale changes of the past may be observed operating in the current changes taking place around us” (1973, p. 161). Plunkett and Marchman (1991, 1993) have used connectionist models to show that tensions arise when analogically based systems are required to learn and store multiple generalizations. We shall suggest that one consequence of such tensions is that the system changes over time to relieve the internal system stress. The focus of this article will be to explain the principles that underlie this pressure, and to illustrate their effect on the verbal morphology of English.

The pressure to eliminate imbalances in a morphological system is often referred to in the linguistics literature as “paradigm simplification”, or the tendency toward regularization. A clear example can be found in the history of English verb inflection. In Old English (ca. 870) there were a minimum of 10 forms of past tense marking on verbs. Four distinct subclasses of “weak” verbs took variants of the suffixes *-t* or *-d*, and at least six “strong” classes¹ marked the past through a stem vowel change, or *ablaut*, as does the modern verb *give–gave*. Over the past 1000 years the system has simplified dramatically as the suffixed past tense classes coalesced into one, which then spread through the ablaut classes. The result is the modern system in which the regular suffix /d/ applies to all but a handful of irregular verbs.

Regularization processes such as these are extremely common cross-linguistically, and raise a number of questions. First, the complex inflectional system of OE existed for hundreds of years – what permitted that stability in the face of an apparent drive toward simplification? What eventually disrupted that stability and caused the system to change? Can the direction

¹ A seventh class, the “reduplicating” verbs, is also treated as a strong class in some texts.

of change be predicted? It is also true that many irregular forms survive, despite the tendency to eliminate irregularity. What factors contribute to this immunity to regularization? Even more intriguingly, some weak forms become irregularized, adopting the patterning of the strong verbs. Why should this happen? What follows is an attempt to answer these questions, and to show the extent to which the answers fall naturally out of basic connectionist principles of learning and generalization.

In doing so, we also hope to show the relevance of historical change for an issue that has received a great deal of attention in the psycholinguistic and modeling literature. That is, the question of what sort of processing mechanism underlies the verbal morphology of English. On the one hand, it has been proposed that an analogically based network might handle all allomorphs, both regular and irregular (Daugherty & Seidenberg, 1992; MacWhinney & Leinbach, 1991; Plunkett & Marchman, 1991, 1993; Rumelhart & McClelland, 1986). On the other hand, it has been argued that at least two mechanisms are required: a network-based system for irregular forms, and a rule-based symbolic processor for the regulars (Kim, Pinker, Prince & Prasada, 1991; Pinker & Prince, 1988; Prasada & Pinker, 1993). Unfortunately, there is a considerable overlap in the predictions made by both models. The differences, though real, are subtle and may be easily obscured by methodological artifacts. However, it is also true that whichever model is adopted for synchronic language must also be consistent with the historical facts. One would even hope for more. Ideally, the correct model would also provide some principled explanation for the major historical developments.

This, then, provides the backdrop against which we undertook the current research. Our primary interest is to show that basic principles of connectionist learning offer insight into some of the forces underlying morphological change. Secondly, we see this as a way to test the viability of the hypothesis that both the regular and irregular verbal morphology in English are the product of a single-mechanism associative network.

The paper is organized as follows. In the next section we lay out, in some detail, the facts of the OE verb system and the changes it underwent over time. We then explain the general principles behind a network learning account of the historical facts, in order to motivate the connectionist models. The modeling section of the paper is divided into two parts. Since the major point of difference between the dual-mechanism and single-mechanism hypotheses rests on the treatment of the regular morphology, we will first focus on change in the regular or *weak* system. In the second half, we broaden the scope of the simulations to study the effect of allowing changes to occur in a single mechanism which is processing both regular and irregular verbs. As we shall see, the results of these simulations are highly consistent with the major historical developments. Furthermore, these results are readily understood in terms of network dynamics, and so provide a rationale for the shape of the attested changes.

2. Data and general issues

In Old English the major division in the verbal morphology was between the “weak” verbs, which formed the past tense with a suffix as modern regular verbs do, and the “strong” verbs, which inflected for tense and number by changing the vowel of the verb stem as does the modern verb *sing–sang*.² In Early Old English (ca. 870) there were two classes of weak verbs,² and six strong verb classes (Stark, 1982; Wright and Wright, 1923; Flom, 1930). The weak verb classes were in a state of transition during Old English. The class designations I and II referred to the relative frequency of these classes at earlier stages in the language (Stark, 1982), but by Early Old English weak Class I, once the largest, had splintered into three subclasses and scattered exceptions and Class II, now the largest and most productive, was the only one to exhibit a consistent paradigm. Over the next several hundred years the weak system simplified, as Class II adopted new members from Class I verbs and verb subclasses that were in any way irregular.

This tendency to eliminate irregularity was not limited to the weak verbs. It affected the strong verbs as well, particularly as the weak verbs grew in numbers. Originally, in ancestors of English, the ablaut series had been the dominant form of inflection. By Early Old English these strong verb classes were still intact and even somewhat productive, but they had diminished in size while the weak classes had grown. By OE only about 25% of English verbs took a form of strong inflection, while the remaining 75% were weak (Quirk & Wrenn, 1975). These classes continued to decrease in size over the Old and Middle English periods and into the current language, where fewer than 100 strong verbs remain, many of which are unstable.

Since in Modern English the majority of formerly strong verbs have become regular if they remain at all, it is easy to assume that the change was a straightforward process of the statistically larger weak class dominating and attracting members from statistically smaller strong classes. But the facts are more complicated than this. For one thing, many strong verbs moved to other strong classes rather than to the weaks, showing that at least some of the strong classes were semi-productive. For another, a number of weak verbs also “irregularized” into a strong class whose members they resembled, and various borrowed words did the same. Finally, a healthy number of strong verbs resisted change altogether, and remain irregular today.

The question of interest in this paper is to account for the details of this movement of verbs from one class to another. We will offer a network account of why the instability arose in the first place, and why the resulting changes took the direction they did. Since this account relies heavily on the

² Historically there was a third weak class, but by the period in question only four verbs (*live*, *have*, *think*, and *say*) remained in weak class III, and for the purposes of this paper we will refer only to weak classes I and II.

phonological structure of the various verb classes, we will begin with a detailed presentation of the relevant data.

2.1. Old English weak verbs

The weak verbs of Proto-Germanic are classified according to the derivative suffix they took between the stem and present tense suffixes. In Class I, the suffix was the high front glide *-j-*. This triggered various phonological changes in the verb stem, resulting in three distinct subclasses in Old English. The first subclass (Ia) was made up of verbs with heavy stems (i.e., with either a long vowel or a final consonant cluster). In these heavy-stem verbs, the *-j-* deleted. The *-j-* also deleted in most light-stem verbs, triggering gemination (or doubling) of the stem-final consonant. This resulted in a second subclass (Ib) with geminate stems in the infinitive and most present tense forms. The third subclass (Ic) was made up of light-stem verbs ending in *r*. In these verbs the *-j-* suffix did not delete, and there was no gemination.

In Class II the verbs took the derivative suffix *-i-*. Since this class had never splintered along phonological lines it offered no formal criteria for class membership: there was no requirement that the stem end in a specific consonant or contain a long vowel, nor was the presence or absence of gemination predictive of the choice of inflection. As a result there were no phonological constraints on membership in this class; or, in the terminology of Bybee (1994), the phonological features defining the class “schema” were relatively open.

Typical EOE West Saxon paradigms for the four weak subclasses are shown in Table 1.³ The inflectional distinctions to note are these. In Class I, the light-stem verbs (subclasses Ib and Ic) take the suffix vowel *e* both in the

Table 1
Weak verb inflection in Early Old English, ca. 870

| Class: | Ia | Ib | Ic | II |
|----------------|-------------------|-----------------|------------------|------------------|
| Example: | de:man 'judge' | fremman 'do' | nerjan 'save' | lufian 'love' |
| <i>Present</i> | | | | |
| 1st sing. | de:m + e | fremm + e | nerj + e | lufij + e |
| 2nd sing. | de:m + st | frem + est | ner + st | luf + ast |
| Plural | de:m + ath | fremm + ath | nerj + ath | lufij + ath |
| <i>Past</i> | | | | |
| 1st sing. | de:m + de | frem + e + de | ner + e + de | luf + o + de |
| Plural | de:m + don | frem + e + don | ner + e + don | luf + o + don |
| P. part. | de:m + e + d | frem + e + d | ner + e + d | luf + o + d |

³ Data are from Stark (1982). The dialect cited is West Saxon, since this was the dominant literary dialect of the period and that referred to in most scholarly texts.

present singular and as the “medial vowel” between the stem and suffix in the past tense. The heavy-stem verbs (subclass Ia) take this *e* in the first person present, but not otherwise. Verbs of Class II, by contrast, take the suffix vowel *a* in the present, and the medial vowel *o* in the past. Furthermore, the high front vowel *i* remains in the Class II paradigm in all the forms where the corresponding glide *-j-* remains in Ic.

Classes II and Ia were large classes of verbs. Ib was smaller, although it was still a moderately good-sized class. Ic, since it contained only the *r*-final light-stemmed verbs of Class I, was quite small.⁴

2.2. OE strong verbs

The six strong classes were differentiated by their *ablaut series*, the series of vowels each took in the present (and infinitive), the preterit singular, preterit plural, and past participle. These are the equivalent of the three “principal parts” of a modern strong verb like *sing/sang/sung*, the only difference being that the OE verb took four “parts” rather than three. Each class showed what are called a class *characteristic*, the feature that served as the basis for class categorization. For most classes this characteristic was the vowel of the present tense stem; in some cases there were also constraints on the consonants allowed in the stem coda. The characteristic for each class is shown in Table 2.⁵

Patterns of change show that these characteristics were more than a descriptive tool for historical linguists – speakers relied upon them when categorizing the verbs. As examples, Class I adopted the borrowed verbs *strive* (from Old French) and *thrive* (from Norse) which fit the *i*: +*fricative* pattern; the weak verb *werian* ‘wear’ was adopted into strong Class IV since it exhibited the appropriate *e* + *sonorant* coda, and the common *a* + *k* coda structure of Class VI drew in the weak verb *cwacian* ‘quake’ and the borrowed Old Norse verb *take*. Note that the vowel *e* does not uniquely identify any class, since it is the present stem vowel for Classes III, IV, and V. This overlap also led to movement, as a number of original Class V verbs shifted to Class IV.

⁴ While we have presented the handbook account of the divisions in the first weak class, other analyses of these data are possible (Dresher, 1981). Kiparsky and O’Neil (1976), for example, offer an account of the distinctions that is based on the phonological differences between the heavy and light-stemmed verbs. These authors acknowledge that over time the system they propose may well have undergone reanalysis leading to less abstract underlying representations and rules that were at least partially morphologically conditioned. They further point out that “the tendency for all weak verbs to move into the second class might indicate morphologization of the system” (p. 546).

By either analysis, however, the Class I verbs divide along precisely the same lines, and whether the distinctions are determined morphologically or phonologically is not crucial to our account. In either case what must be explained is why the light-stemmed Class I verbs changed, and why these changes took the direction they did.

⁵ Data are from Flom (1930) and Wright and Wright (1923).

Table 2
OE strong class characteristics

| Class | Present tense/ infinitive stem vowel | Coda consonant | Examples |
|------------------|--|--|----------------------------------|
| I | i: | Originally, any one consonant (often a fricative) | dri:fan 'drive' ri:san 'rise' |
| II | e:o | Any one consonant | fre:osan 'freeze' |
| III ^a | e | Sonorant + C | singan 'sing' helpan 'help' |
| IV | e | One sonorant | beran 'bear' |
| V | e | One C (usually stop or spirant) | etan 'eat' |
| VI | a | One consonant (often /k/) | scacan 'shake' |

^a Although this class originally took the stem vowel *e*, by the period in question it took a number of different vowels.

2.3. *The basis of morphological change*

The Old English verb system offered a learning task of great complexity, since the learner needed to assign each new verb to one of the 10 or more inflectional subclasses. If this assignment had been entirely arbitrary the task would have been extremely difficult, particularly in light of the fact that weak Class II was much larger than other verb classes, consistent in its paradigm, and highly productive. All things being equal this more regular class could have dominated the system, interfering with the learning of the smaller classes. There were two factors that conspired to prevent this.

The first was relative token frequency. The weak past tenses had by far the higher class size, and were by that measure the more regular form of inflection. But the individual weak verbs tended to be infrequent. The opposite pattern was found with strong verbs; class size tended to be small, but the frequency of occurrence of many individual strong verbs was often quite high (Quirk & Wrenn, 1975). This high token frequency contributed to the ease of learning the members of the relatively small classes.

Second, as the data in section 2.2 show, verbs were not assigned arbitrarily to past tense classes. Instead, the majority of classes showed strong phonological cues to class membership. Since the strong verb classes were organized around a particular stem form, for example, a learner could abstract and operate on a small number of generalizations of the type "long *i* in the infinitive → Class I vowels in the past". Similar form-based generalizations were available for certain weak subclasses. These generalizations could be pivotal in the learning of classes with low type and token frequency, and the data suggest that speakers did indeed rely on them. As was discussed in the previous section, verbs that historically belonged to one class, but more closely fit the form-based criteria of another, often changed their inflections to enter the class whose members they more closely resembled.

This tendency to base membership in the smaller inflectional classes on extra-morphological features like phonological form arguably made the classes easier to learn, but the drawback of such a system is that if those features are altered, the class structure may also be lost. The data in the next section show that morphological change soon followed when phonological changes obscured the key generalizations that identified members of the smaller classes.

2.4. *Developments in the weak verb system*

Two phonological changes affecting the language as a whole had interesting consequences for the weak verbs (Stark, 1982). First, throughout this period there was a strong tendency toward glide vocalization. Both the *-j-* of the weak Ic (*nerjan*) verbs and that of the first singular present in Class II reduced to the corresponding vowel, *i*. This left the medial vowel as the only formal difference between the verbs of Class II and those of Ic. At the same time, and we believe as a result of this increased similarity, the two classes collapsed into one. The verbs of the small Ic class adopted the medial vowel *-o-* of Class II, becoming indistinguishable from original members of that class.

Second, during this period English also began to simplify its geminate consonants. Recall that one major distinction of the Ib (*fremman*) subclass was its alternation between geminate and non-geminate stems. This distinction disappeared by late OE. Interestingly, most of the verbs of the *fremman* subclass then adopted the Class II paradigm as well.⁶ This shift involved two changes: these verbs both adopted the Class II inflections and altered their stems to include the stem-final high vowel.

The heavy-stem verbs of Class Ia were not affected by these phonological changes, and Ia continued as an independent and even somewhat productive inflection class. By late OE (approximately AD 1250) there remained essentially two weak verb classes, II and Ia, which differed only in their medial vowels and in the fact that the Class II verb stem ended in *-i-*. Over the course of Middle English the picture simplified further. The stem-final *-i-* disappeared from most Class II verbs. Vowel reduction in unstressed syllables, followed by deletion of the unstressed medial vowel, eliminated most remaining distinctions between the two classes (see Fig. 1).

The result of these changes is the regular past tense inflection of Modern English. Thus from a historical standpoint the regular past is the natural outcome of an analogically based series of developments, although this fact is obscured when one looks at the synchronic data.

⁶ Although most Ib verbs moved into Class II, a small number also drifted in to Class Ia. The significance of this will be discussed in the modeling section.

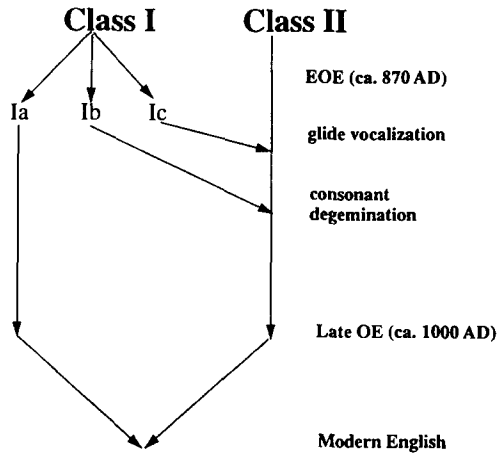


Fig. 1. Coalescence of weak verb classes over time. General phonological changes affect the language as a whole, and members of the smaller weak verb classes, having lost their cues to class membership, respond by shifting to the productive Class II.

2.5. Change in the strong verb system

Similar patterns of change affected the strong verbs. Since the class characteristics served as the basis for categorization, a phonological change that affected them could be expected to lead to morphological change as well. This was indeed the case, as shown when five of the six major classes developed variant subclasses.

One such group was the subset of strong Class I verbs referred to as the *contract verbs*. In these verbs the consonant /h/ was deleted intervocallically, and the two vowels that were left adjacent as a result developed into the diphthong *e:o* in the present tense in place of the class characteristic *i:*. These verbs then moved into Class II, since the diphthong was the characteristic of that class. In a similar development, a set of Class II verbs (the *aurist presents*) took the vowel *u* in the present stem instead of the characteristic *e:o*. These were irregular by the standards of their own class, and did not fit any other ablaut class either. Over time, these items were either lost from the language or took on weak past tenses.

A more complicated development involved Class III. Originally this class took *e* in the present tense, followed by a sonorant + consonant cluster, but a series of phonological changes splintered it into a number of subclasses that varied in size and phonological coherence. One such phonological process changed the *e* to *i* when it preceded a nasal (as in the verb *sing*); a second changed the *e* to *ie* following a palatal consonant (resulting in an earlier form of the modern *yield*). These two processes had different consequences. The first affected a large number of verbs, creating a new class with the characteristic *i + NC*. Existing verbs of the same form then

moved to join this class: *sink*, *cling*, which had been exceptional members of strong Class I, *ring* from the weak verbs, and so on.

The second change, on the other hand, affected only a few items, and there were no grounds for differentiating these from other verbs with *ie* in the present tense. These verbs were soon regularized.

Token frequency also influenced the effect that the loss of a phonological generalization had on particular verbs. As an example, Class IV contained three items that developed variations on the characteristic stem vowel *e*: *shieran* ‘shear’, *niman* ‘take’, and *cuman* ‘come’. Over time *shieran* has been regularized while *niman* has dropped out of the language. *Cuman*, on the other hand, was (as now) a very high-frequency item, and has survived as a strong verb. The same was the case in Class VI, where the frequent verb *standan* ‘stand’ remained strong despite being irregular by the standards of its class.

To summarize, the smaller inflectional classes tended to organize themselves on the basis of phonological similarity, with clear phonological regularities providing templates for class members to match. This allowed the system to remain stable until phonological changes eroded these identifying characteristics. But once the phonological regularity was lost, the task of learning the correct verb class assignment became more difficult, and it was at this point that the system changed to re-establish a more stable configuration.⁷ Change took a number of forms: verbs were classed together under a new basis for generalization (as in the *sing* subclass of Class III), were “regularized” into a more dominant class (as were the contract verbs of Class I or the aorist presents of Class II), or, if their token frequency was sufficiently high, resisted change altogether.

3. Motivation for the modeling account

This description of historical change leaves many questions of specific interest that must be addressed if it is to avoid the charge of vagueness that characterizes many post hoc accounts of historical change. After the fact, the explanation makes sense. but one wonders what other outcomes might have been possible, and what conditions were responsible for this particular outcome. How robust are the phenomena? Can one meaningfully quantify notions such as “complexity” of the learning task, “stable configuration”, “phonological similarity”, and so on? Can one predict just how much disruption will be tolerated before the system moves into another configuration? In general, we would like a more precise account which lets us make predictions about which classes or specific items will change. In developing that account we will suggest that the linguistic facts show the effects of

⁷ In languages that rely less heavily on phonological distinctions, morphological classes may be characterized by syntactic or semantic properties (Wurzel, 1989). In either case, if these extramorphological characteristics are lost the morphological system changes as well.

frequency and regularity in the learning of complex inflectional systems. This is an interaction that falls out quite naturally from particular connectionist learning procedures, as Seidenberg and McClelland (1989) discuss in detail. How does this come about?

In networks using the backpropagation algorithm (Rumelhart, Hinton & Williams, 1986), learning occurs in the following manner. The output response to an input stimulus is compared to a “teacher” (or the expected output) and the discrepancy between the two, referred to as the error, is calculated. The learning algorithm then adjusts the connection weights of the network in a way that reduces the overall error. For frequent patterns, a reduction in error on one presentation of the pattern entails error reduction on all other presentations. Similarly, a weight change that leads to better performance on a highly regular mapping will improve performance on all other patterns that share in that regularity. Thus both regular patterns and frequent patterns make large contributions to error reduction. Since the backpropagation algorithm changes weights in proportion to the effect that the change will have in reducing *total* error, greater weight changes will be associated both with frequent and with regular patterns. As a result, these items are mastered more quickly, while items that are irregular and infrequent are learned with more difficulty.

In the remainder of this paper, we will show that this aspect of network behavior can account for the historical regularization patterns of the English verbs. It has been commonly assumed that productivity in connectionist models is a simple function of frequency (see, for example, Prasada & Pinker, 1993); we intend to make more precise what role frequency actually plays in productive behavior, and how the effects of frequency are modulated by the phonological characteristics of patterns.

Because of the current interest in the issue of whether the English regular verbs are produced in a qualitatively different fashion than the irregulars, we will begin with an account of the weak verb system. The goal in this section will be to show that the process of analogical attraction; which is known to affect the strong verbs of English, motivated movement among the weak verbs as well. If this is true, then the fact that the same process affected both categories of verbs suggests that both are products of the same mechanism. This argument assumes, naturally, that all classes of weak verbs were “regular” in the relevant sense. If this were not the case, a dual mechanism account could easily explain analogical movement among the weak verbs by assuming that the verbs which shifted category were irregular.⁸

⁸ We are aware, of course, that such an account still leaves open the possibility that although *historically* the weak classes were the product of an analogical mechanism such as a network, the resulting changes created the conditions necessary for the weak verbs to be handled *currently* by a rule mechanism. We see no possible way to falsify such a counter-argument, but we do point out that it leaves unexplained the mechanism by which the putative shift from network to rule system is supposed to occur.

The clearest argument for the regular status of the weak verbs comes from Pinker and Prince (1988), which states that denominal verbs cannot be irregular:

... irregularity is a property of verb roots. Nouns and adjectives by their very nature do not classify as irregular (or regular) with respect to the past tense... [denominal and de-adjectival] verbs can receive no special treatment and are inflected in accord with the regular system, regardless of any phonetic resemblance to strong roots. (p. 111)

The great majority of the weak verbs were derived from nouns, adjectives, and other verbs. Class II was almost entirely composed of denominal verbs (Wright & Wright, 1923, p. 64; Flom, 1930) making it “regular” by the criteria give above. Yet despite this regularity, Class II served as an analogical attractor, particularly for the verbs of Class Ic. Class I contained verbs of all three types, denominal, deverbal, and de-adjectival (Moore & Knott, 1955, pp. 71–72). It is true that many verbs of Class I were causatives (Lass, 1992), and these might, by virtue of their verb roots, be treated as irregulars instead. But as the discussion in section 2.4 made clear, the divisions in Class I were based on phonological form, not on derivational history. Weak subclasses Ib and Ic behaved “irregularly” while Ia behaved in more “regular” fashion, yet all three contained denominal and de-adjectival verbs. Table 3 gives examples of denominal verbs in the three subclasses of Class I.

Crucially, the Ic verb *ferian* moved into Class II over the course of the OE period, as did its denominal classmate *answerian*, ‘answer’ (Mossé, 1968).⁹

The weak verb model will demonstrate the role of phonological similarity and type frequency in the regularization process. In the second half of the paper we will expand the data set to include both the strong and the weak verbs, in order to consider the question of regularization and immunity to regularization in more detail. In the second model, as in the first, the hypothesis will be that change results from cumulative errors which occur as language is learned, imperfectly, and transmitted across generations. If this is the case, then verbs with high token frequency, and those that retain their phonological cues to class membership, should resist change more successfully. This claim is consistent with the English data. For example, Bybee

Table 3
Examples of denominal verbs in weak Class I

| Subclass | Verb | |
|----------|-----------------|----------------------|
| Ia | de:man ‘judge’ | cf. do:m, ‘judgment’ |
| Ib | wemman ‘defile’ | cf. wamm, ‘stain’ |
| Ic | ferian ‘carry’ | cf. fær, ‘journey’ |

⁹ The Ib verb *wemman* appears to have dropped out of the language; no Middle English form was found in the cited texts.

(1985) examines three classes of OE ablaut verbs, and shows a systematic difference such that low-frequency class members have regularized while high-frequency members tend to remain irregular in Modern English. Bybee and colleagues (1982, 1983) have also shown that modern strong verbs tend to cluster into sets according to a phonological prototype structure. The implication is that shared aspects of form helped these verbs resist regularization; low-frequency irregular verbs lacking such class characteristics have not survived. Plunkett and Marchman (1991) have shown parallel effects in connectionist models. In an analysis of the conditions under which a single network can learn competing stem to inflectional mappings, they show that the learning of exceptional forms is highly dependent on type and token frequency and the availability of phonological cues. In the second model we will show that these factors are crucial both for morphological productivity and for immunity to regularization.

We will begin with the weak verb model.

4. Network account of the weak verb change

As pointed out earlier, in Early Old English the smaller weak verb classes exhibited a large degree of formal coherence. This allowed the learner to exploit information about the phonological characteristics of each class. As general phonological change eroded these characteristics, membership shifted. The drift of certain verbs from Class I to Class II is consistent with our assumption that the morphological change resulted from difficulties in learning inflected forms. The model that follows demonstrates why this is so.

4.1. Architecture

The problem was modeled with a feedforward network implementing the backpropagation learning algorithm. The network had 518 input units, 18 hidden units divided into two layers, and 21 output units (see Fig. 2).

The first 512 units of the input bank were localist representations of individual verbs. These connected to a first layer of 10 hidden units which converted the localist representation into a more compact distributed form in order to reduce the computation time required by the model. The last six input units stood for individual tense/person/number inflections. At each training iteration one “verb” unit and one “inflection” unit were activated simultaneously, and the task of the network was to produce a representation of the fully inflected verb over the output units.

The output was designed to represent the formal features that distinguished the various classes of weak verbs. There were 21 output units, of which 12 units were dedicated to these phonological features. The other units encoded 9-bit random patterns that were assigned to each verb stem. This pattern marked each verb as unique, and also allowed the network to

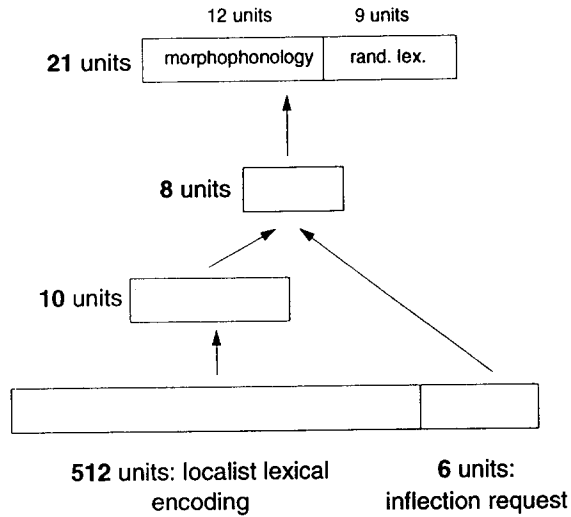


Fig. 2. Architecture of network used in weak verb model.

treat each set of six individual inflected forms as manifestations of the same verb.

The 12 “inflection” units on the output encoded the following information:

- presence of medial or inflectional vowel
- identity of this vowel
- presence of stem-final high segment
- identity of this segment (vowel or glide)
- presence of geminated stem consonant
- presence of a long vowel

4.2. Training stimuli

There were 480 verbs in the training set, divided into four subclasses whose numbers were roughly representative of their relative class size in OE (see Table 4). Each verb was learned in the six inflected forms shown in the paradigms in Table 1. These specific forms were chosen because in

Table 4
Training data for weak verb simulation

| Class | Number of exemplars |
|-------|---------------------|
| Ia | 128 |
| Ib | 64 |
| Ic | 32 |
| II | 256 |

combination they illustrate the significant distinctions among the four subclasses. Thus, a complete pass through the data set (what we refer to as a single epoch of learning) involves training on 2880 distinct forms (480 verbs \times 6 inflections).

4.3. Training

A first network was first taught the canonical OE weak verb system as described in section 2.1. This network was trained using backpropagation learning (Rumelhart *et al.*, 1986) for 30 passes through the data set. After training the three largest classes (II, Ia, and Ib) were all produced correctly. The verbs of the small Class Ic were also inflected correctly, but showed the effect of attraction to Class II by producing a vowel instead of a glide for the stem-final high segment.

At this point a training regime was adopted that was designed to show how errors in learning can lead to change over time. A new network was set up with random initial weights, and taught to produce the verb classes as they were formed after the processes of glide vocalization and degemination destroyed the identifying characteristics of Class Ib and Ic stems. This second network was trained for 10 sweeps through the data set. Errors, while more prevalent than in the first network, were still rare. Three Ic verbs took Class II inflections, and two verbs of subclass Ib altered their stems to match the stems of Classes Ia and II respectively. All other items took the inflections appropriate to their class.

The output of this network was then collected and used as the teacher signal for a new network. This new network was started with random initial weights, but had an architecture identical to the first, and was given the same input patterns as the first; the only difference between the two was the new teacher signal. Since the output of the one network was the teacher for the next, any errors in learning in the first network became part of the data set for the second. This process was repeated a further five times, with each new network receiving the output of its predecessor as teacher. The result was a total of seven networks after glide vocalization and degemination took effect, in addition to the “baseline” network which had learned the canonical, or unaltered, OE weak verb system. Each of these seven networks was trained for 10 epochs to reproduce the output of the previous net.

We will refer to the series metaphorically as “generations” of networks, since although this training regiment does not claim to exactly mimic the time-course of language change across generations of speakers, it does capture the gradual nature of such change and the causal role played by transmission of imperfectly repeated patterns. What we wish to demonstrate is that language can change through a process that occurs between individuals, from generation to generation. Although we are representing this with single networks, equivalent to individuals, we take these as idealiza-

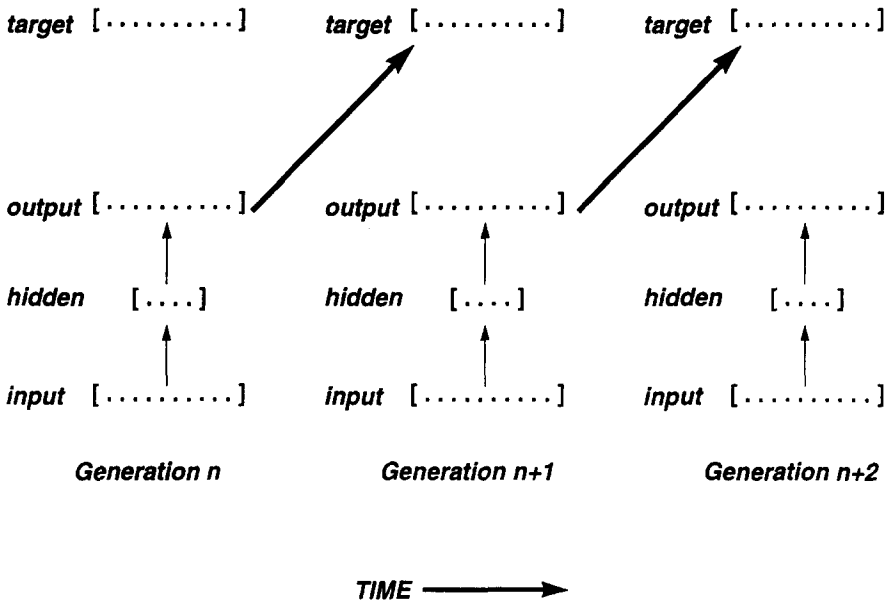


Fig. 3. Generational learning.

tions of this more general process and will speak of them that way in the discussions.

Fig. 3 gives a graphical representation of the generational learning regime.

4.4. Results

In each succeeding generation, the error increased for verbs belonging to the smaller Classes Ib and Ic; this occurred as a result of the network failing to learn an increasing number of these verbs correctly, and producing them on the model of Class II instead. In the first generation network after the phonological change, the two large and distinctive classes (Ia and II) were learned correctly. Error was also low in Ib (the *fremman* class), with only one verb showing a tendency to adopt the high stem-final vowel of Class II, and another showing an equally weak tendency to adopt the long vowel of Ib. The Ic (*nerjan*) verbs showed more interference. Three (or 9%) tend toward Class II inflection. Still, the great majority of verbs in this class firmly maintain their Class I inflections.

By the second generation after the phonological changes, the number of Ic (*nerjan*) verbs adopting Class II inflection increased to 14, nearly 50% of that class. In Ib, the *fremman* subclass, 15 (or 24%) of the verbs had altered their stems to include the stem-final *i* of Class II, and two of these adopted the Class II inflectional vowels as well. Fig. 4(a) and (b) illustrate this drift.

Fig. 4(a) gives the proportion of verbs of each class at the initial stage of the model (Generation 0); Fig. 4(b) reflects the decrease in the size of Classes Ib and Ic and corresponding increase in Class II by Generation 2.

After five generations, all members of the small Class Ic are conjugated as Class II. In addition, 39 of the Ib verbs now take the Class II *-i-*, and the number taking Class II inflectional vowels has increased by 20.

By the seventh generation, the process has significantly altered the classes that are learned. The major division in the data set is now between Classes II and Ia. All Ic verbs are identical to Class II. Fifty of the original 64 Ib verbs now include the Class II *-i-* in their stems, and 27 of these take the Class II inflections. Of the small number of Ib verbs that have not changed class completely, most show only marginal activation on the Ib inflectional nodes, indicating that they are also in the process of changing. Interestingly, although the majority of Ib verbs have merged with Class II, the two

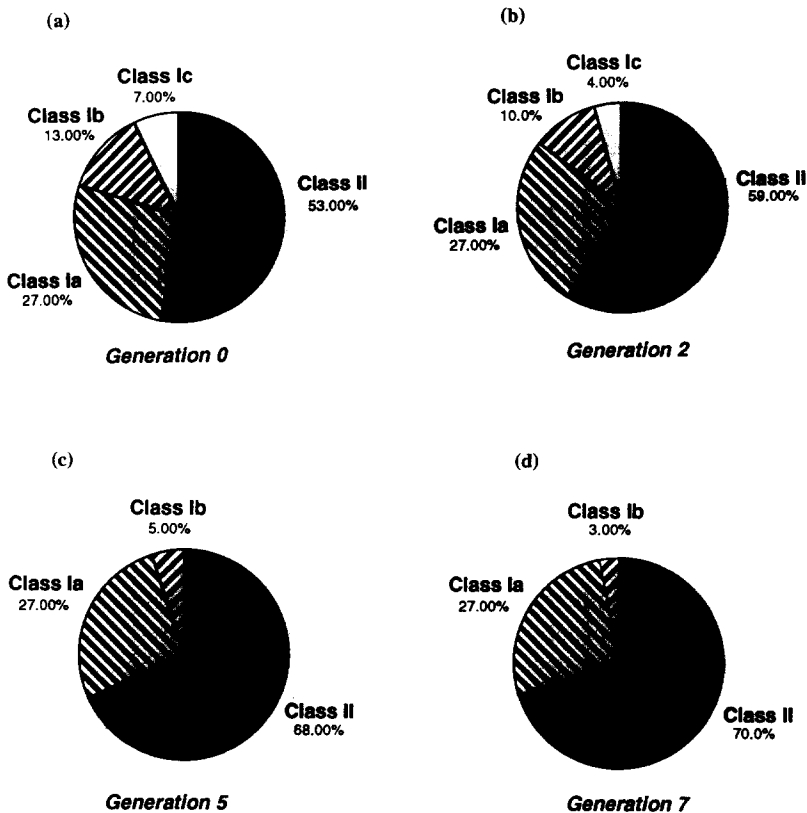


Fig. 4. Difference in proportion of verbs in data set conjugated according to each class over time. (a) Structure of the original data set, before the onset of generational learning. (b) Difference in numbers two “generations” after phonological change. (c) Classes at generation 5. (d) Final results, after seven generations of training.

long-vowel forms have taken on Ia inflection, since a long stem vowel is characteristic of Class Ia members.

Fig. 4(c) and (d) show the percentages of the data set in each weak class at the end of Generations 5 and 7. Although Class Ia has remained stable, Class II has grown considerably as it attracts increasing numbers of new members from the shrinking Classes Ib and Ic.

The set of simulations was repeated four times with different random initial weights, and gave consistent results.¹⁰

Why did this happen? The results of the simulation are not surprising. If we assume that language transmission involves noise in the form of variability in the input, then it is not difficult to see why the language should change in the way it does. At the onset, the classes differ in terms of their phonological coherence and their class size. Those sub-patterns that are initially less common or less well defined are the hardest to learn, and these tend to be lost over several generations of learning. This process snowballs as the dominant class gathers in new members and this combined class becomes an ever more powerful attractor.

5. Immunity to regularization

The model described in section 4 demonstrates that the loss of their phonological cues to class membership led to rapid assimilation of the smaller classes into a larger, more regular class. Our claim is that this regularization process resulted from the difficulty in learning of items that had neither high type frequency nor phonological class cohesion to support them. But since all items were equally low in token frequency, and the similarity structure of both small classes was lost, it was not possible to show that the inverse is also true: that frequent items, and those able to exploit class similarities, will resist change. In the next section we will enlarge the data set to include the strong verbs, and with this data look at the effects of token frequency and phonological similarity on immunity to change. The expanded set offers useful contrasts: certain of the strong verbs were extremely high in frequency, and others, though low frequency, belonged to classes that were internally highly consistent. Furthermore it will offer a more accurate picture, since the modern weak verbs did not evolve in isolation. Tracing the development of this more complex data set, we will show that the model offers a plausible explanation for the system's evolution.

¹⁰ The only variation shown was the occasional tendency for Class II verbs to adopt the common vowel of the past participle (*e*).

5.1. English strong verbs

In section 2 we saw that over time a large number of strong verbs were attracted to the more populous weak classes. Others, relying on the similarity structure of their ablaut class, survived more successfully. If phonological changes disrupted that structure, classes either restructured themselves around a new generalization (as in the *sing* example), or lost members to the weak classes or to strong classes whose characteristics they marched. Furthermore, high-frequency verbs were able to survive whether or not they had class similarity to support them. In the next section we will show that these patterns of change result naturally from the learning requirements of a connectionist network.

5.2. Modeling the weak and strong verb system

5.2.1. Description of the model

The model used a feedforward network trained with backpropagation, with 118 input units, 50 hidden units, and 165 output units (see Fig. 5). The 118 units of the input bank were divided into 100 units to represent each verb, and 18 to represent the tense/number combination with which each verb was to be inflected. The 100 “verb” units represented abstract lexical entries corresponding to the verb stem. These were created by generating 700 100-bit vectors, where each bit had a 0.1 probability of being activated, so that on average each vector had 10 of the 100 bits on. These vectors were then randomly assigned to the verbs of the training corpus as their input representations. The 18 “inflection” units were a series of patterns that encoded six tense/number/person combinations which were seen for each verb during training.

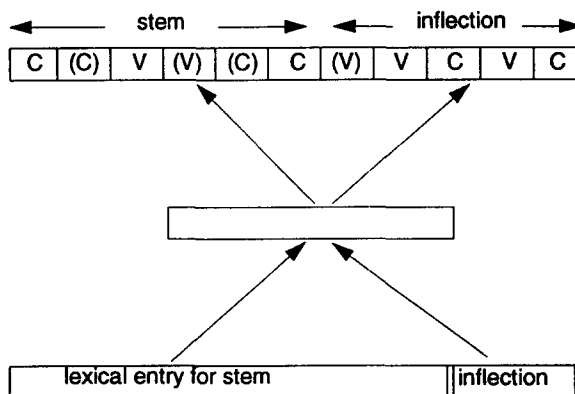


Fig. 5. Architecture of the network used in the combined model.

At each training iteration a verb pattern and an inflection pattern were activated simultaneously on the input, and the task of the network was to produce the phonemic form of the inflected verb over the output units.

The output layer represented an 11-phoneme template over which a fully inflected verb could be displayed. Each phoneme was made up of a 15-bit distinctive feature representation, for a total of 165 output units.

The first six slots in the output template were a CCVVCC string dedicated to the verb stem. This allowed the representation of a mono-syllabic verb with a C or CC onset, a single vowel (V) or diphthong (VV), and a C or CC coda. The final five slots were for inflectional affixes, in the following order:

- V for medial vowel used in the present tense of certain weak verbs
- VC for a weak past tense suffix
- VC for the suffix used to mark person in both the weak and the strong verbs

All empty slots, in both the stem and the suffix, were activated to 0.5 on all units.

5.2.2. *Stimuli*

At the beginning of the simulation, the data set consisted of 106 strong verbs and 327 weak verbs. This gave a starting total of 433 different stems, of which 32% were strong. The simulations used generational learning in a regime similar to the one employed in the first simulation. The strong verbs were taken from the six OE ablaut classes listed in Flom (1930) and Wright and Wright (1923); weak verbs were taken from these sources and also from Stark (1982).

The network was taught to produce each verb in six different inflected forms, three present (first person singular, second singular, and plural) and three past (first singular and plural preterit, and past participle). Again, these inflectional forms were chosen to illustrate the significant inflectional distinctions among the subclasses of verbs, both weak and strong.

To determine frequencies for the strong verbs we used the frequency of occurrence of each in two sources: the concordance of Old English (Di Paolo Healey & Venezky, 1980) and a Chaucer concordance (Tatlock & Kennedy, 1963) for frequency data of the Middle English period. Frequencies for individual words were significantly correlated across the two sources ($p < .001$). This indicates that the relative frequencies for our items remained stable for a considerable period of time, and consequently we did not change item frequencies over the course of generational learning. Verbs were presented to the network once for every 10 occurrences in the concordances; verbs that were not found in these sources were given a frequency of 1. All weak verbs were arbitrarily given a frequency of 1 to maintain an appropriate token-frequency ratio in the data set.

Table 5
Weak verb training data, Generation 0

| Class | Number of exemplars |
|-------|---------------------|
| Ia | 90 |
| Ib | 38 |
| Ic | 18 |
| II | 181 |

As in Simulation 1, the weak verbs were divided into four subclasses with a size roughly representative of relative class sizes in OE. In the first generation of learning, the classes were as given in Table 5. Because we know that historically there was an increase in the weak classes, particularly in weak Class II, as the language created new verbs, the number of weak verbs in Class II was increased by 10 at the beginning of each new generation.

In choosing the strong verbs we selected members from the major classes, as defined in section 2.2, and from the irregular subclasses that resulted from the phonological changes discussed in section 2.5. The structure of the strong verb data in the model is shown in Table 6. For each class, we indicate the number of verbs which conform to the overall class characteristics (shown as Number of consistent verbs) as well as the number of verbs which deviate in some way (shown as Number of inconsistent verbs). The numbers of each verb type in the data set reflect the relative sizes of these subclasses in the historical data.

As a result, the training set parallels the historical data in having a number of moderately large classes showing clear phonological cues to class membership, and a large number of scattered exceptions to those cues. In the discussion of the results, items of the first type will be referred to as our *consistent* classes, while the exceptions will be grouped together as the

Table 6
Number of consistent and inconsistent items in each class

| Verb class | Stem vowel | Examples | Number of consistent verbs | Variant stem vowel | Examples | Number of inconsistent verbs |
|------------|------------|-------------------|----------------------------|--------------------|---|------------------------------|
| I | i: | dri:fan 'drive' | 20 | e:o | the:on 'thrive' | 3 |
| II | e:o | fre:osan 'freeze' | 18 | u | lu:can 'lock' | 6 |
| III | i | singan 'sing' | 13 | e, ie, eo, u | helpan 'help' gielpan 'yield' feohtan 'fight' murnan 'mourn' | 15 |
| IV | e | beran 'bear' | 4 | ie, u | scieran 'shear' cuman 'come' | 4 |
| V | e | etan 'eat' | 7 | | | 0 |
| VI | a | scacan 'shake' | 11 | e, ie | hebban 'heave' hliehhan 'laugh' | 5 |

inconsistent classes. Frequency differences cut across the consistent and inconsistent classes.

5.2.3. Training

The model uses the generational learning scheme presented in section 4.3. The initial network was taught to produce the verbs as they were formed in Early Old English. This net was trained for 75 000 iterations (15 passes through the data set). At this point the total summed squared error was approximately 2.5. Rather than resulting from slightly imperfect learning across the board, this error reflects the fact that certain mappings were learned extremely well, while others were still incorrect. A second network was then set up with weights initialized to random values. The output from the initial network was used as the teacher for the second network, which was again trained for 75 000 iterations. The output of the second network was then used as the teacher to a third network. This procedure was followed for a total of five networks. As in the weak verb simulations, each new network was architecturally identical to the first, and the input representations were the same. Only the teacher signal changed from generation to generation.

As in the earlier model, the goal of this training regimen was to reproduce the effects of transmission in language change: what were errors in learning in the first generation became part of the data set of the second, and so were propagated through subsequent nets. Given the way learning takes place in the network, it is the low-frequency, inconsistent patterns that are hardest to learn, and consequently these lead to the most errors, and are the first to change or regularize over time.

5.2.4. Results

In analyzing the model's performance we first computed the closet target, for each phonemic segment on the output layer, among the phonemes in the model's vocabulary. Taking the closest target to be the model's response on each phoneme slot, we compared the response on the vowel and affix slots with templates corresponding to the correct response for each inflectional class. The model was judged to have inflected a verb according to a given class if the following two criteria were met: first, for each of the six inflected forms of the verb, the response of the model had to match the template of that class better than that of any other class (the *best match* criterion); second, the model's response had to be within a certain minimum distance of the template (the *proximity* criterion).

By these criteria, 94% of the model's responses matched one of the established inflectional classes. The remaining 6% did not fit any class on all six inflected forms. Instead, these responses appear to be blends of two or more potential categories.

In presenting the results we will attempt to answer the questions raised in the introduction. First, what factors lead to change, and which provide a

degree of immunity? Second, what is the direction of change – that is, what determines productivity? And finally, do these factors distinguish between the weak and strong verbs in a way that suggests the two result from qualitatively different mechanisms, or do both classes of verbs behave in a similar fashion?

(a) *Learning of correct past tense inflection.* High-frequency mappings were learned well in the model. Weak verbs with high type frequency, with exceptions to be discussed shortly, were learned with weak inflections, and even after five generations 80% of the high-frequency strong verbs were inflected correctly regardless of the phonological consistency of their inflectional class. By contrast, the performance of the low-frequency strong verbs (verbs with <5 presentations per epoch) varied as a function of their class. Importantly, the variation results from an interaction between frequency and phonological consistency of the class.

Fig. 6 gives the percentage of correctly inflected strong verbs of the consistent and inconsistent classes over the course of learning in the first generation network. As it shows, low-frequency verbs belonging to consistent groups do better than those from the inconsistent classes, reaching nearly 70% correct by the end of 70 000 iterations. The inconsistent low-frequency verbs, by contrast, never reach more than 27% correct.

This pattern continues over the five generations of training. All strong classes lose members over time, but there are clear differences in the rate at

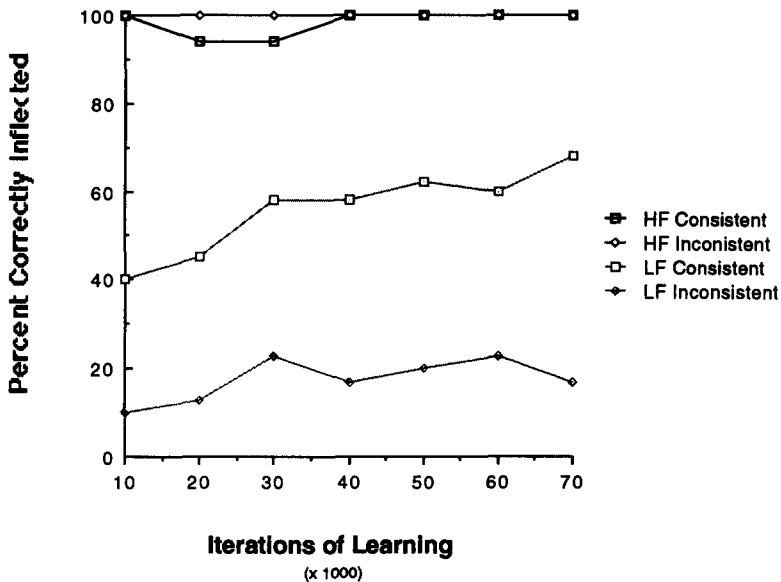


Fig. 6. Percentage of correctly inflected strong verbs over first generation of learning.

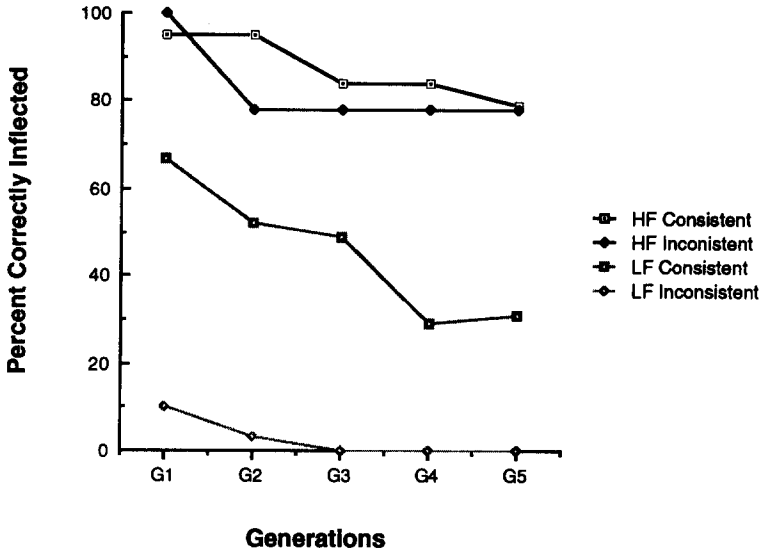


Fig. 7. Percentage of correctly inflected strong verbs over five generations of learning.

which items are lost. Small, inconsistent classes have trouble from the beginning, while the more consistent classes lose members more slowly. Fig. 7 shows that even after five generations, 35% of the low-frequency consistent verbs remain in their correct inflectional classes.

Even among these more consistent classes, however, there are differences in the rate of change. The *sing* subset of Class III exhibits the highest degree of class consistency of the larger classes, and not surprisingly it is also the most successful at the learning task. Other classes, with fewer members or less strict phonological constraints, do less well. However, even the small Class IV (the *tear* class) is able to retain low-frequency members through Generation 3, unlike the inconsistent classes. Fig. 8 shows the success rate for low-frequency verbs of the *freeze* (Class II), *sing* (Class III), and *tear* (Class IV) classes, and those of the inconsistent classes, over the five generations.

The same pattern of results was found among the weak verbs classes, although in this case type frequency (i.e., larger class size) played a larger role than token frequency. Weak Class II verbs, with exceptions to be discussed shortly, were learned with the correct inflections. Subclasses Ib and Ic were lost from the start as their members adopted the inflections of Class II and, more rarely, Class Ia. Class Ia had the advantage of higher type frequency and a phonological cue in the form of the heavy stem, and was correspondingly more successful in the learning task. In the early generations 75% of the Class Ia verbs were inflected correctly. By Genera-

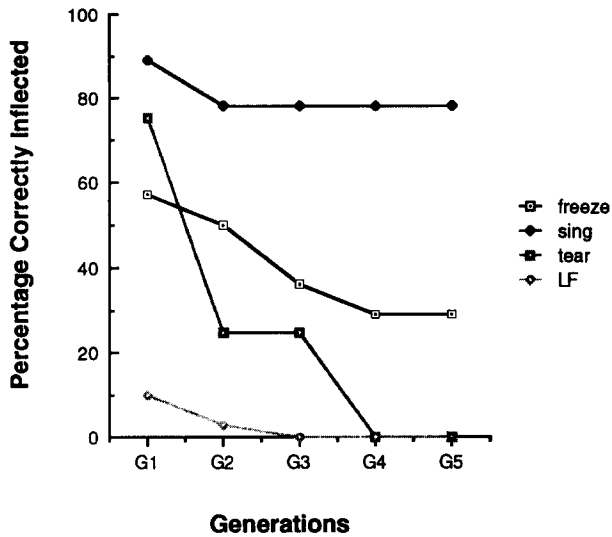


Fig. 8. Success rate of low-frequency members of selected strong classes over time.

tion 5 over half of these verbs had shifted to Class II, but Class Ia remained as a viable and productive inflection class.

In this simulation (unlike in simulation 1) many Class II verbs modified their form by dropping the stem-final *-i-*. This change occurred historically as well, and is interesting as an indication of the interactive nature of the regularization process: while Class II apparently dominated the weak verb system, causing a large number of Class I verbs to “regularize” by adopting Class II inflections, Class II verbs were in turn influenced to become more “regular” by dropping the stem-final vowel that had been distinctive of that class.

(b) *Productivity*. The verbs that changed inflection generally did so over a period of time, altering each inflected form individually rather than change all at the same point. As a result, there was a period of fluctuation for most verbs, and it varied in length. After this period most verbs settled into a new class and remained stable there for several generations. In analyzing the results, we are interested in learning which classes served as the stable resting points of the verbs that changed inflection.

i. Strong verbs. There were two competing pulls on vulnerable strong verbs. First, the weak classes had the advantage of higher type frequency, and as this class grew, it attracted an increasing number of the strong verbs. This was not an overwhelmingly strong attractor, since in the model the strong verbs still formed a large enough group to maintain a degree of balance, as indeed they did in the Old English period, where regularizations were relatively rare (Lass, 1992). Still, as Fig. 9 shows, over five generations

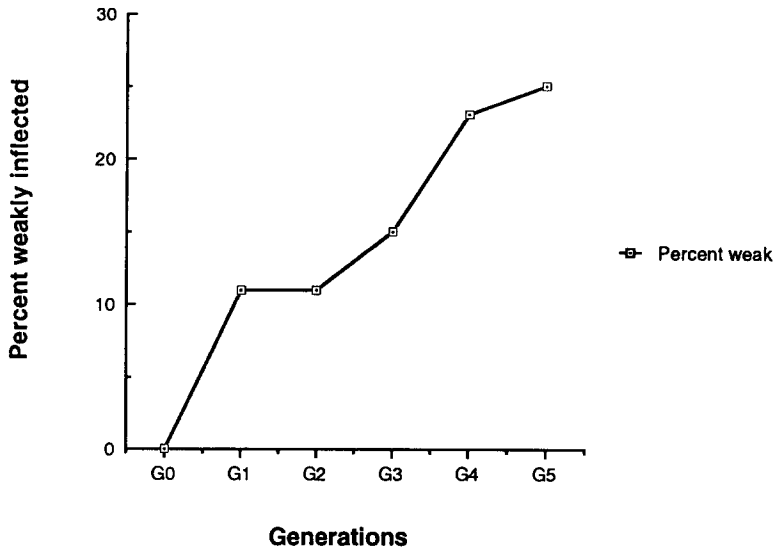


Fig. 9. Increase in percentage of initially strong verbs which regularize into the weak classes over time.

the number of originally strong verbs that regularized into the weak classes grew from 0 to 25% of the set.

The second pull came from other strong verbs, as the phonologically consistent strong classes attracted new members that fit their characteristics. In the model, as in the real language data, class membership was largely based on the form of the present/infinitive stem vowel, or the rime of the present/infinitive stem. If two or more classes shared a given characteristic, productivity was determined by their relative size and consistency.

If the strong classes competing for the same characteristic were unequal in type and token frequency, the larger was taken as dominant and other verbs with the same characteristic changed to join this dominant group. These results are summarized in Table 7. In the model, this competition resulted in productive behavior for Class II, as one of the *contract verb* set of Class I, and other verbs with *e:o* or *eo* in the stem moved to that class.

Table 7
Shifts to dominant class, based on class characteristic

| Verbs with stem vowel | Moved to | From | Example |
|-----------------------|--------------------------------|---|--------------------------------|
| e:o, eo | Class II | Class I contact verb (1) Class III (4) | le:on 'lend' sweorfan 'rub' |
| e | Class IV | Class III (3) | meltan 'melt' |
| u | Agree with <i>cuman</i> 'come' | Class II aorist present (1) | bru:can 'enjoy' |

The movement of *e*-stem verb to Class IV is also explained in these terms. Class IV and a small subset of Class III verbs both took *e* in the present tenses, and a liquid (*r* or *l*) in coda position. Since in our model these two classes were unequal in both type and token frequency, three of the four Class III variants moved to Class IV early in the training.¹¹

The third example, at first glance, is a less expected result. This was the case of a verb in present tense *u* that changed its past tense inflection to agree with that of the highly frequent but inconsistent verb *cuman*, ‘come’. This result is surprising, and despite the fact that it occurs only once in the model it does predict that items with high token frequency and extremely low type frequency might serve as analogical attractors. This finding is at odds with claims in the linguistic literature (Bybee, 1994) and with what might be expected in connectionist models (Plunkett & Marchmann, 1991). But despite the non-intuitive nature of this result, there is experimental evidence in its favor. Treiman and Zukowski (1988), for example, found that when asked to pronounce non-words, subjects occasionally produced responses that were analogies to high-frequency irregular words, such as pronouncing *vone* to rhyme with *gone* rather than the regular pronunciation *bone*.

If the classes sharing characteristics were equal in size, neither dominated the other. This pattern is summarized in Table 8. Classes IV and V, for example, both took present tense *e*, and in the model had similar type frequencies. Not surprisingly, these classes trade members over the five generations of learning.¹² Similarly, Class I and the *sing* class of III, whose present tense vowels were *i*: and *i* respectively, lost occasional members to each other. Historically there was also movement between Classes I and III (see section 2.5) affecting items that matched on relevant features, including the coda of the verb stem.

In the model, the present tense vowel was the strongest cue to class membership, but it was not the only one. Since the learning algorithm is

Table 8
Shifts between equal classes, based on class characteristics

| Verbs with stem vowel | Moved between classes |
|-----------------------|-----------------------|
| i, i: | I, III (4) |
| e | IV, V (2) |

¹¹ This discrepancy in frequencies was due to an error in coding the classes in the model; historically, these classes were much more evenly balanced, and maintained their identities into the Middle English period.

¹² Again, this is behavior that is seen in real-language data. It corresponds to what Wurzel (1989) calls “stability indifferent” classes, two inflectional classes of equal type and token frequency, based on the same characteristics, who lose members randomly back and forth. See the cited work for historical examples.

capable of extracting any generalization that is available in the data, the model operated on other class characteristics as well.

A number of changes were influenced by the structure of the verb's coda, as the examples above show. In a small number of other cases, verbs changed their present tense vowel to join classes whose past tense they agreed with. These cases are summarized in Table 9.

ii. Weak verbs. When weak verbs shifted class the great majority remained weak; the exceptions to this will be discussed under (c), below. Both Classes II and Ia attracted new members, but the conditions under which they did so were very different. Class II was both the largest inflection class in the data set and the most diffuse phonologically, and, as in the historical data, it was the most productive of the weak verb classes. This class exerted a general attraction on the other weak classes, one that was not dependent on phonological similarity alone. Verbs of Ic, which resembled the Class II verbs phonologically, were the quickest to assimilate and arguably did so on the basis of similarity. But members of the more distant Classes Ia and Ib moved into Class II as well, showing that phonological similarity was not the only basis for the shift.

Although Class II was dominant, it lost 21 (or 12%) of its members to Class Ia. Nineteen of these 21 fit the phonological criteria for Class Ia membership, showing that unlike the movement from Class Ia to Class II, phonological similarity *was* a decisive factor for shifts in the opposite direction.

A question raised by these data is why the verbs of weak Class Ia, which resisted assimilation altogether in simulation 1, should not be equally resistant to change in simulation 2. One crucial factor concerns the reliability of the phonological cue to Class Ia membership. Recall that the verbs of this class all had heavy stems. In the first simulation only the Ia verbs exhibited this trait, making a long vowel or a CC coda cluster an invariant sign of membership in that class. In the more historically accurate data set of the second simulation, on the other hand, verbs of other classes might display the same trait, so that the presence of a heavy stem no longer invariably predicts the form of inflection. As a result the verbs of Class Ia,

Table 9
Class shifts based on past tense vowels

| Verbs with past tense vowels | Moved to | Example of attractor (in past plural) | Moved from | Example (in past tense) |
|------------------------------|-----------------------|---------------------------------------|-------------------------------|---------------------------------|
| u | Class II | frus 'froze' | Class III (2) | guld 'yielded' slunk 'slunk' |
| u | Class III Class IV | sung 'sang' cum 'come' | Class II (3) Class III (1) | suc 'sucked' curf 'carved' |

while resisting assimilation better than those of Ib or Ic, were still subject to its pull.

(c) *Irregularization of regular verbs.* Weak verbs were also subject to “irregularization” into strong classes if they matched the characteristics of a productive strong class. Over the course of the simulation 12 of the weak verbs (approximately 3% of the weak set) dropped their affixes and adopted a strong past tense inflection. Again, weak verbs matching the strong characteristic in both the vowel and consonant structure were most vulnerable, but since the strong classes are largely defined by the stem vowel of the present tense, many weak verbs changed on that basis alone. The best attractors were again strong Classes I and III, but other classes were also productive. Table 10 gives the stem vowel or CVC structure of the weak verbs that irregularized in the model, along with characteristic members of the classes that they entered.

Note that while irregularization of weak verbs is possible in the model, it requires the attraction of phonological regularity to win out over the attraction of type frequency, and therefore is relatively rare.¹³ Overall, only 3% of the weak verbs move into a strong class, and matching the phonological characteristics of a strong verb class does not guarantee that a

Table 10
Cases of irregularization in strong verb model

| Irregularizing verb stem | Class entered | Present stem vowel of new class | Examples of existing class members |
|--|---------------|---------------------------------|--|
| <i>ri:an</i> ‘roar’; <i>hrishan</i> ‘shake’; <i>tilian</i> ‘labor’ | I | i: | <i>dri:fan</i> ‘drive’ <i>ri:san</i> ‘rise’ |
| <i>swinsian</i> ‘make music’; <i>shirtan</i> ‘shorten’; <i>wirsian</i> ‘get worse’; <i>kispan</i> ‘bind’ | III | i | <i>singan</i> ‘sing’ <i>swingan</i> ‘swing’ |
| <i>thakan</i> ‘stroke’ <i>sæ:gan</i> ‘lay low’ | VI | a | <i>scacan</i> ‘shake’ <i>wacan</i> ‘awake’ |
| <i>we:arn</i> ‘warn’ ^a | II | e:o | <i>fre:oze</i> ‘freeze’ |
| <i>werjan</i> from <i>ge-werjan</i> , ‘clothe’ <i>mæ:rsian</i> ‘make famous’ | V | e | <i>metan</i> ‘measure’ <i>etan</i> ‘eat’ |

^a The OE vowel was *a*. This was incorrectly learned as *e:a* in the model.

¹³ Irregularization is also relatively rare in children acquiring the synchronic language (Marcus, Pinker, Ullman, Hollander, Rosen & Xu, 1992), and Plunkett and Marchman (1993) have previously shown this is to be the case in connectionist models as well.

weak verb will irregularize. Many verbs that were equally good candidates for change remained regular, and in at least one case the data set contained two homophonous verbs (*werjan*: ‘clothe’ and ‘defend’) that behaved differently in this respect, one joining strong Class V, the other remaining weak. Parallel cases can be found historically; in fact *werjan* itself offers a historical example. The form (*ge-*) *werjan*, ‘clothe’, originally weak, has developed into the strong verb *wear/wore*. The verb meaning ‘defend’ remained weak until dropping out of the language (where it is now represented only by *beware* and related forms). *Ring* is a second case in point: the verb meaning “to sound resonantly” was originally weak, but developed the strong past tense *rang*, while the homophonous verb meaning “to encircle” takes the regular (weak) past *ringed*.¹⁴

5.2.5. Summary

Our claim has been that stability in the inflectional system results from the interaction of type frequencies (i.e., class size), exploitable regularities of classes (i.e., phonological consistency), and the token frequencies of individual verbs. Verbs which either belong to small classes, lack consistent defining characteristics, or are low in frequency should change most rapidly; change in other verbs will depend on the precise extent to which they possess the characteristics which make them resistant to assimilation.

These assumptions were borne out by the results. In the model, high-frequency mappings were learned more successfully than low-frequency ones, giving an advantage to the second weak class, with its high type frequency, and the high-frequency strong tokens. Members of classes showing a high degree of internal consistency resist change better than singletons and members of unstructured classes.

The interactions of frequency and phonological similarity account for productivity in the model as well. Classes displaying similarity structure not only held on to their own members more successfully, they also pulled in new members that show the same characteristics. The weak classes, with their high type frequency, were able to attract a large number of the strong verbs, particularly if those verbs did not match the characteristics of their own class.

Generally, if there is movement between weak and strong, it was the predominantly movement from strong to weak. There were, however, cases of weak verbs adopting strong inflection by analogy to existing members of strong classes. Thus both the strong and the weak verbs were susceptible to

¹⁴ It should also be noted that Pinker and Prince (1988) criticize the Rumelhart and McClelland (1986) model for failing to treat homophonous verbs distinctly, despite the fact that the English language does do so. The current data show that this failure is due only to the fact that the Rumelhart and McClelland model represented items only by their phonological form, and does not result from a more fundamental problem with connectionist model of inflection.

the same changes, although the degree of change was influenced by the discrepancy in class size.

6. General discussion

The goal of this project was to use basic connectionist principles of learning and generalization to explain, in a reasonably detailed manner, the changes that occurred in the English verb system over time, and the results are indeed consistent with this goal. The model produces the overall effect of regularization. More interestingly, it also captures less intuitive phenomena, such as the irregularization of irregular and regular verbs, and the immunity to change on the part of certain irregulars. Given the frequency and similarity of the items in the data set and the generational learning scheme, these aspects of historical change emerge naturally from the model.

One important issue, then, is how much of the model's success results from general principles of the learning mechanism, and how much was determined by the choice of parameters. One such parameter was the distribution of the data in the training sample: the patterns of frequency and similarity shown by the items of the training set were obviously crucial to the performance of the model. These patterns were determined by the historical data, however, rather than being a free choice for the modelers. It is also true that type frequency is considered independent of token frequency in simulation 1, in order to show the effects of high type frequency in the "attraction" of verbs from one class to another. While this was an artificial manipulation (for weak verbs varied in their token frequencies just as the strong verbs did) it does not change the overall performance of the network. Instead, it gives a clear picture of the independent effect of type frequency, which can then be distinguished from the effects of varying token frequency when this is introduced in simulation 2.

A second parameter was the amount of training time. The fact that training was *not* continued to convergence was crucial in the model's performance; this could lead to the suggestion that the results were overly sensitive to the particular number of training epochs used. A second look at Fig. 6 shows this suggestion to be incorrect: although the performance of the model resulted from the pattern of error in training, this pattern was consistent over training time. Thus while the decision to end training after 15 epochs was an arbitrary one, the relative differences among the verb conditions would have been the same had training ended at some other point.

A third parameter is the choice of representation. There are two points to consider here. The first has to do with the information that is made available to the network. In simulation 1, the output representation is purposefully constructed to contain all (and only) the specific phonological features considered necessary to distinguish among the different verb classes.

Although the possibility exists that this representation determined the results by building in the relevant aspects of the problem to be solved, the second simulation shows that this objection is unfounded. In simulation 2 the output representation was the phonological form of the complete inflected verb, and the network was left to determine which aspects of that representation were relevant. Despite the differences in their output representations, the two networks produced the same effects, showing that these effects were not due to the representations alone.

The two simulations also differ in their choice of input representations: in simulation 1 the input is localist, while in simulation 2 the input is a randomly chosen bit vector. In networks of this sort there is always the possibility that such randomly assigned vectors might introduce a chance bias that could affect the behavior of the model. To satisfy ourselves that this was not the case in the present model we performed a cluster analysis on the input patterns from simulation 2. The results of that analysis show that the random similarities found on the input level did not correlate with the pattern of results.

In sum, the behavior of the model results not from felicitous parameter settings, but rather from general principles of the learning mechanism itself. On this topic, though, it is important to clarify the extent to which the results rely on *connectionist* learning, rather than learning mechanisms in general, and the extent to which they rely on the “errorful” teacher of the generational learning scheme. In regard to the first point, the central claim of this paper is that the patterns of change it considers result from factors of class size, frequency, similarity, and the interactions that occur when a single mechanism is required to perform multiple input–output mappings. These are well-studied characteristics of connectionist networks, and our account is tightly linked to the realization of the learning within the connectionist framework (see Plaut, McClelland, Seidenberg, & Patterson, 1994; Plunkett & Marchman, 1991, 1993). In regard to the second point, the use of the generational learning scheme had a considerable effect on the model’s behavior. In tests with a single network, we found that error (and therefore change) stabilized over the course of learning, rather than spread throughout the data set. The process of change found in the model resulted directly from the use of multiple networks with changing teacher inputs. It was this “moving target” in the learning process that led the model to reanalyze the verb classes, and caused the spread of error through the data set.

At this point we should also dispel a possible misconception about this model, that it offers no possibility of reanalysis in the linguistic sense. This is incorrect, but it is easy to see how the confusion might arise. Reanalysis is a change in the underlying representations constructed by speakers, often based on surface phonological alternations that changed over time. Given the superficial parallelism between the input-output levels of the model, and the underlying-derived forms of linguistic theory one might take the input

representation to be the model's equivalent of an underlying form. And, since the input to the model never changes, this equivalence would mean that no reanalysis can take place.

This surface parallelism is misleading, however. An "underlying form" is the stored lexical knowledge from which the surface form can be computed. In the network, this stored knowledge exists not in the input, but in the learned weights, which do indeed change over the course of generational learning in ways that could accurately be described as a "reanalysis".

The story we offer – that much of morphological change is driven by analogy, and the direction of change is predictable from a consideration of class size (type frequency), token frequency, and the phonological coherence of a given class – is not new. Many linguistic accounts have assumed some version of these points (Bybee, 1985, 1994; Skousen, 1989; Wurzel, 1989). Importantly, however, we are able to relate the facts of morphological change to what is known about learning in a complex system. Analogical accounts have been criticized in the past for their lack of formality and their excessive explanatory power (Kiparsky, 1975). The current account attempts to define the factors on which the analogical drive is based, and to specify the way these factors must interact in the network. As such, it allows us to understand what seem to be idiosyncratic analogical effects as inevitable and predictable results of gradient descent learning.

This same reliance on learning in a complex system allows us to explain the co-occurrence of frequency and similarity effects in morphological change. Bybee (1985, 1994), in an account that agrees in crucial features with our own, offers compelling evidence that type frequency and the degree of "openness" of the defining features of the schema (the generalization that can be abstracted over the phonological similarities of members of an inflectional class) are equally important in determining productivity and resistance to change. Our model makes the further point that these are not independent factors, but are necessarily linked, since they are two manifestations of a single underlying cause, the gradient descent properties of the learning rule.

Thus we wish to emphasize that on our account analogical change in the network is *not* strictly based on type frequency, but instead on frequency as it interacts with phonological consistency. It is important to be clear on this point, since in the English data the dominant weak verbs had a massive advantage in type frequency, and this may give the misleading impression that frequency alone is sufficient to account for productivity. Examples from other languages may make it clearer that this is not the case. Consider one example from the history of German. At an early stage in that language neuter nouns of the *a*-declension took the plural suffix *-u*. Word-final *-u* later deleted for phonological reasons. This phonological change resulted in an unstable situation, in our terms, since there was now no difference between singular and plural for most neuter nouns even though for the language as a whole a singular/plural distinction was the norm. The system had to change

to re-establish a stable configuration, which in this case meant that the neuter *a*-declension nouns had to adopt the plural inflection of some other class. The neuter *er* declension, which matched the *a*-neuters both in gender and in its phonological characteristics, *did* distinguish between singular and plural, marking the plural with the suffix *-er* and stem vowel ümlaut (e.g., *lamm/læmmer*), and many of the *a*-neuters moved to this class. The crucial point here is that the *-er* class was much smaller, containing only about 15 verbs

This move to the *er*-class regardless of its lower type frequency is precisely what our model would predict – note that it is exactly parallel to the “irregularization” process that occurred in the model when weak forms were pulled into the much less populous strong classes by virtue of their phonological cohesion with verbs of the class.

As a further example, there are arguably cases where a particular inflection behaves like the default, applying to novel words, borrowings, and in other circumstances where the inflection is otherwise unavailable or unknown. These default inflections are largely unresponsive to phonological information, as in the modern English *-ed*, which can apply to any new word (but see Seidenburg & Bruck, 1990). Nor are they necessarily the most frequent, as has been argued in the case of the German plural suffix *-s*. If morphological productivity results from the interaction of frequency and phonological similarity, how can such default behavior arise?

On the current model, this behavior crucially depends not only on the characteristics of the default class itself, but on the generalizations that can be abstracted from the system as a whole. If a complex inflectional system largely consists of classes that are characterized non-morphologically, then they cannot adopt novel patterns that lack the appropriate characteristics. Under these circumstances a class without defining features, or with an “open schema” in the sense of Bybee (1994), can serve as the default since it is capable of accepting new members that do not fit elsewhere. In the current network this explains why many strong verbs that no longer fit the characteristics of their classes adopted weak inflections. In other work (Hare & Elman, 1992; Hare, Elman & Daugherty, submitted) we show that in a connectionist network a pattern may develop as the productive inflection even without the advantage of high type frequency, as long as it is the only one without defining features.

On a further issue, by assuming that particular patterns are difficult to learn, and these difficulties in learning affect the state of the language in the adult speaker, the current model also makes the claim that the factors of type and token frequency and phonological characteristics of a class will have repercussions for representation and access of morphologically complex words in the adult lexicon. Are these claims supported in the experimental literature?

Experiments by Prasada, Pinker, and Snyder (1990) and Seidenberg and Bruck (1990) offer one possible source of evidence. In the model, the

high-frequency irregular verbs are easier to learn than low-frequency items, and if we predict processing differences for items that differ in learnability, than these two should show such a difference. Prasada et al. and Seidenberg and Bruck ran naming experiments in which subjects were presented with a verb stem and asked to name the corresponding past tense form. In both experiments there was a reliable effect of past tense frequency for irregular verbs, with longer naming latencies for low-frequency than for high-frequency items. Furthermore, Daugherty and Seidenberg (in press) have shown that this pattern obtains in a connectionist model, for the same reasons as we have given here.

There is also experimental evidence that phonological consistency plays a role not only in how verbs are learned, but how they are then stored and accessed. In the naming experiment described above, Seidenberg and Bruck tested whether regular verbs with phonologically similar irregular neighbors (such as *bake*, which has the neighbors *take* and *shake*) are responded to differently than those whose neighbors are all regular. The assumption behind the experiment was that if phonological similarity plays a role in processing, then the inconsistent words, those with irregular neighbors, should be harder to generate. As the results show, this was indeed the case. The authors found a statistically reliable consistency effect: regular words whose neighbors were also regular were responded to more rapidly than those with irregular neighbors. Thus the evidence shows that not only is phonological consistency a factor, but it is a factor among *regular* verbs as well as irregulars, suggesting that there is no qualitative difference in how the two types of verbs are processed.

This finding relates to our final point, the implications of our model for the issue of whether two distinct mechanisms are required to process the English verb inflection system. One criterion for evaluating an account of a system is the extent to which that account predicts the ways in which the system will change. The current model, embedded in the single-mechanism account, not only offers a close fit to the historical data, but relates historical change to potential differences in the current state of the language as well. In particular, it should be noted that the morphological changes that took place did not distinguish the regular (weak) and irregular (strong) verbs in any qualitative manner. The weak verbs were affected by the factors of frequency and phonological consistency just as the strong verbs were. This is as expected in a network account, but more difficult to explain under the assumptions of the dual-mechanism explanation, where what is affected by the application of the rule is a symbol, and it is the essence of a symbol to have no phonological content to exploit.

In favor of the dual-mechanism account, of course, one could argue that the weak verbs of Old English only superficially resembled the regular verbs of the modern language, but that in fact many of these were irregular. And, as we stated in the Introduction, many weak verbs, particularly those of Class I, were causatives (Lass, 1992). Causatives, being derived from verbs,

clearly have verb roots, and could arguably have been irregular by the criteria discussed in the Introduction. In order for this argument to stand, however it would have had to be true that the Class I causatives showed different behavior than the Class I denominals and de-adjectival verbs, which by the same criteria were necessarily regular. This was not the case. As section 2.4 made clear, the divisions in Class I developed along phonological lines, with no correlation to derivational history.

Finally, even had this not been the case, a dual-mechanism account would need to account for the unexplained qualitative shift from a system with many verb classes of roughly equal status to one with a single rule and scattered exceptions, in order to explain how the current regular/irregular system developed.¹⁵ The evidence does not indicate that such a shift occurred. As the synchronic data mentioned earlier show, the factors we posited as leading to change in Old English are still viable in the current language.

Thus both the synchronic and diachronic evidence lead to the conclusion that the network approach is capable of offering a principled explanation of the morphological data, one that is more parsimonious than a dual-mechanism account, and able to give insightful account of facts that are otherwise unexplained. The current model joins a growing body of research investigating morphology from the point of view of connectionist theory, and raising interesting questions about language development (Plunkett & Marchman, 1991, 1993), lexical representation and access (Rumelhart & McClelland, 1986; Seidenberg, 1992; Daugherty & Seidenberg, 1992), productivity (Hare & Elman, 1992; Hare et al., submitted), the loss of ability under brain damage (Marchman, 1993), and so on. While in its current state of development the connectionist approach cannot offer a complete theory of morphology, these results reinforce the validity of the models in the study of human language ability, and point to the eventual success of the ongoing effort to develop a theory of morphology based on connectionist principles of learning and generalization.

Acknowledgements

We are grateful to members of the PDPNLP group at San Diego, and of the Centre for Speech and Language at Birkbeck, for valuable comments on earlier versions of this work. We would also like to thank Mark Seidenberg and two anonymous reviewers for helpful comments and suggestions.

This research was supported by National Science Foundation grant DBS

¹⁵ This would be parallel to the proposed shift from rote learning to a rule-based system in language acquisition. Work by Plunkett and Marchmann (1993) demonstrates that a single mechanism is also able to account for data that have been taken as supporting such a qualitative shift.

92-09432 and Office of Naval Research grant ONR N00014-93-0194, both to Jeffrey Elman. Much of this work was done while the first author was at Birkbeck College, University of London, supported by British Economic and Social Sciences Research Council grant R000232931. Additional support was provided by JCI grant SPG8931677 and a Medical Research Council programme grant, both to William Marslen-Wilson and Lorraine K. Tyler.

References

- Bach, E., & Harms, R. (1972). How do languages get crazy rules? In R.P. Stockwell & R.K.S. Macauley (Eds.), *Linguistic change and generative theory*. Bloomington, IN: Indiana University Press.
- Bybee, J. (1985). *Morphology*. Philadelphia: John Benjamins.
- Bybee, J. (1994). *Regular morphology and the lexicon*. MS, University of New Mexico.
- Bybee, J., & Modor, C. (1983). Morphological classes as natural categories. *Language*, 59, 251–270.
- Bybee, J.L., & Slobin, D. (1982). Rules and schemas in the development and use of the English past tense. *Language*, 58, 265–289.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.
- Daugherty, K., & Seidenberg, M. (1992). The past tense revisited. In *Proceedings of the 14th Annual Meeting of the Cognitive Science Society*. Princeton, NJ: Erlbaum.
- Daugherty, K.G., & Seidenberg, M.S. (in press). Beyond rules and exceptions: a connectionist modeling approach to inflectional morphology. In S.D. Lima, R.L. Corrigan, & G.K. Iverson (Eds.), *The reality of linguistic rules*. Philadelphia: John Benjamins.
- Di Paolo Healey, A., & Venezky, R.L. (1980). *Old English concordance*. Newark; University of Delaware.
- Dresher, B.E. (1981). *On the learnability of abstract phonology*. In C.L. Baker & J. McCarthy, (Eds.), *The logical problem of language acquisition*. Cambridge, MA: MIT Press.
- Flom, G. (1930). *Introductory Old English grammar and reader*. Boston: Heath.
- Hare, M., Elman, J.L., & Daugherty, K.G. (submitted). Default categorization in connectionist networks.
- Hare, M., & Elman, J.L. (1992) A connectionist account of English inflectional morphology: evidence from language change. In *Proceedings of the 14th Annual Meeting of the Cognitive Science Society*. Princeton, NJ: Erlbaum.
- Hogg, R. (1992). Phonology and morphology. In R. Hogg (Ed.), *The Cambridge history of the English language* (Vol. I) Cambridge, UK: Cambridge University Press.
- Kim, J., Pinker, S., Prince, A., & Prasata, S. (1991). Why no mere mortal has ever flown out to center field. *Cognitive Science*, 15, 173–218.
- Kiparsky, P. (1968). Linguistic universals and linguistic change. In E. Bach & R. Harms (Eds.), *Universals in linguistic theory* (pp. 170–202). New York: Hold, Rinehart & Winston.
- Kiparsky, P. (1975). What are phonological theories about? In D. Cohen & J.R. Wirth (Eds.), *Testing linguistic hypotheses* (pp. 187–209). New York: Wiley.
- Kiparsky, P., & O'Neil, W. (1976). The phonology of Old English inflections. *Linguistic Inquiry*, 7, 527–557.
- Labov, W. (1973). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (1980). *Locating language in time and space*. New York: Academic Press.
- Lass, R. (1992). In R.M. Hogg (Ed.), *The Cambridge History of the English Language. Volume I: The beginnings to 1066*. Cambridge, UK: Cambridge University Press.
- MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: revising the verb learning model. *Cognition*, 40, 121–157.

- Marchman, V. (1993). Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience*, 5, 215–234.
- Marcus, G.F., Pinker, S., Ullman, M., Hollander, M., Rosen, T.J. & Xu, F. (1992). *Overregularization in language acquisition*. Monograph of the Society for Research in Child Development, 57 (4, Serial No. 228).
- Moore, S., & Knott, T.A. (1955). *The elements of Old English*. Ann Arbor, MI: G. Wahr.
- Mossé, F. (1968). *A handbook of Middle English*. Translated by J.A. Walker. Baltimore: Johns Hopkins Press.
- Pinker, S., & Prince, A. (1988). On language and connectionism: analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28, 73–193.
- Plaut, D.C., McClelland, J.L., Seidenberg, M.S., & Patterson, K.E. (1994). *Understanding normal and impaired word reading: Computational principles in quasi-regular domains*. Technical Report PDP.CNS.94.5, Parallel Distributed Processing and Cognitive Neuroscience, Carnegie Mellon University, Pittsburgh, PA.
- Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: implications for child language acquisition. *Cognition*, 38, 43–102.
- Plunkett, K., & Marchman, V. (1993). From rote learning to system building. *Cognition*, 48, 21–69.
- Prasada, S., & Pinker, S. (1993). Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes*, 8, 1–56.
- Prasada, S., Pinker, S., & Snyder, W. (1990). *Some evidence that irregular forms are retrieved from memory but regular forms are run generated*. Paper presented at the thirty-first annual meeting of the Psychonomic Society (November, New Orleans).
- Quirk, R., & Wrenn, C.L. (1975). *An Old English grammar*. London: Methuen.
- Rumelhart, D.E., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In D.E. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing* (Vol. I). Cambridge, MA: MIT Press.
- Rumelhart, D., & McClelland, J. (1986). On learning the past tense of English verbs. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing* (Vol. II). Cambridge, MA: MIT Press.
- Seidenberg, M. (1992). Connectionism without tears. In S. Davis (Ed.), *Connectionism: Theory and practice*. Oxford.
- Seidenberg, M., & Bruck, M. (1990). *Consistency effects in the generation of past tense morphology*. Paper presented at the thirty-first meeting of the Psychonomic Society (November, New Orleans).
- Seidenberg, M., & McClelland, J. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 97, 447–452.
- Skousen, R. (1989). *Analogical modeling of language*. Dordrecht: Kluwer.
- Slobin, D. (1977). Language change in childhood and in history. In J. Macnamara (Ed.), *Language learning and thought*. New York: Academic Press.
- Stark, D. (1982). *Old English weak verbs*. Turbingen: Niemeyer.
- Tatlock, J., & Kennedy, A. (1963). *A concordance to the complete works of Geoffrey Chaucer*. Gloucester, MA: Peter Smith.
- Treiman, R., & Zukowski, A. (1988). Units in reading and spelling. *Journal of Memory and Language*, 27, 466–477.
- Wright, J., & Wright, E. (1923). *Old English grammar*. Oxford: Oxford University Press.
- Wurzel, W.U. (1989). *Inflectional morphology and naturalness*. Dordrecht: Kluwer.