Title
From Weared to Wore: A Connectionist Account of Language Change

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 15(0)

Authors
Hare, Mary
Elman, Jeffrey L .

Publication Date
1993

Peer reviewed

From weared to wore: A connectionist account of language change

Mary Hare
Department of Psychology
Birkbeck College, University of London
hare@crl.ucsd.edu

Jeffrey L. Elman
Department of Cognitive Science
University of California, San Diego
elman@crl.ucsd.edu

Abstract

This paper describes a technique developed for modeling historical change in connectionist networks, and briefly reviews previous work applying that technique to the problem of the historical development of the regular verb system from early to late Old English. We then broaden the scope of the simulations and ask what the effect would be of having such changes occur in a single mechanism which is processing both the regular and irregular verbs, and extended over a longer time period. As we shall see, the results are highly consistent with the major historical developments. Furthermore, the results are readily understood in terms of network dynamics, and provide a rationale for the shape of the attested historical changes.

Introduction

The forms which languages take are many and varied, and one of the goals of linguistic science is to try to understand the sources of such diversity. One way to understand why languages might assume the forms they do is to look to 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.

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 are presumably present in synchronic grammars. The focus of this paper will be on those forces which may arise from learnability considerations (Kiparsky, 1971; Slobin, 1977). Plunkett & Marchman (1991; in press), for example, have shown that tensions and stresses arise when analogically based connectionist 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 in such a way as to relieve the internal system stress.

The relevance of historical change to understanding contemporary language is particularly relevant for an issue which has received a great deal of attention in the psycholinguistic and modelling literature. That is the question of what sort of processing mechanism underlies the verbal morphology of English, and specifically, the formation of the past tense allomorphy. There has been considerable controversy regarding the English past tense. On the one hand, it has been proposed that an analogically based network might handle all morphological relationships, both regular and irregular (Daugherty & Seidenberg, 1992; MacWhinney & Leinbach, 1991; Plunkett & Marchman, 1991, in press; Rumelhart & McClelland, 1986). On the other hand, it has been argued that at least two mechanism are required: A network-based system for irregular forms, and a rule-based symbolic processor for regular forms (Kim et al., 1991; Pinker & Prince, 1988; Prasada & Pinker, in press). Unfortunately, there is 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. First, we are interested in discovering whether an analogically based mechanism such as a connectionist network might be used to simulate language change and give insight into some of the forces which underlie change. Secondly, we saw this as a way to test the hypothesis that both the regular and irregular verbal morphology in English are carried out by a single-mechanism associative network.

The paper is organized as follows. We begin with a review of the historical developments. We then describe a technique we developed for modeling historical change in connectionist networks, and briefly review previous work applying that technique to the problem of the historical development of the regular verb system from early Old English to late Old English. We then broaden the scope of the simulations and ask what the effect would be of having such changes occur in a single mechanism which is processing both the regular and irregular verbs. As we shall see, the results are highly consistent with the major historical developments. Furthermore, the results are readily understood in terms of network dynamics, and provide a rationale for the shape of the attested historical changes.

Historical data

The major division in Old English (OE) verbal morphology was between the so-called "weak" and "strong" verbs. The strong verbs were the descendents of Indo-European verb classes that inflected for tense, person, and number by changing the vowel of the stem (as in the modern *sing-*

song). The weak verbs formed the past tense by adding an affix, as the regular verbs do today.

In Early OE (*ca.* 870) there were approximately five subclasses of weak verbs which varied in their size and consistency (Stark 1982). Over the next several hundred years the weak paradigms simplified, eliminating most forms or verb classes that were in any way irregular, so that by late OE there remained only two weak classes, distinguished by the presence or absence of a suffix vowel in the past tense. Over the course of Middle English the picture simplified further, as vowel reduction in unstressed syllables (and eventual deletion of the unstressed medial vowel) eliminated most remaining distinctions between the conjugation classes. The result of these changes is the regular past inflection of Modern English. This tendency to eliminate irregularity was not limited to the affixing verbs. It affected the seven classes of "strong" or ablaut verbs as well, especially as the weak paradigm grew more dominant. In Early OE there had been something of an equilibrium between the weak and strong verb classes. Approximately 75% of the verbs were weak, but this advantage in numbers was counterbalanced by the fact that weak verbs tended to be infrequent items, while the individual strong verbs tended to be highly frequent (Quirk and Wrenn, 1957). This led to a balance between the weak and strong verbs at an early stage, but the equilibrium was disrupted for the following reason. The weak past was generally treated as the default inflection and was applied to new verbs derived from nouns or adjectives or borrowed from other languages, so that the weak class grew in size as new verbs entered the language. As a measure of the effect this had on relative class size, consider that in Modern English approximately 60% of all verbs are borrowed from French or Latin, and another 20% come from nouns (Prasada and Pinker, in press). These exact numbers cannot be applied to Old English, of course, but they illustrate trends that were as true of the language then as they are in Modern English. Not surprisingly, then, the weak class increased rapidly, and the balance between weak and strong was increasingly disrupted. As early as the thirteenth century many of the strong verbs had taken on affixed past tenses in addition to their strong form, and went through a prolonged period with two possible past tenses. By the end of this period a majority of the alternating verbs had been firmly regularized into the weak class.

Still, the facts are more complicated than a simple one-way shift towards regularization would suggest. For one thing, some alternating verbs lost the weak alternative and remained strong. In the 13th and 14th century, for example, the verbs *sink* and *take* both appeared variably with strong and weak past forms, yet both are strong in Modern English. Other verbs shifted from one strong class to another, rather than into the weak class. As an example, most members of OE strong Class V regularized early in the OE period, but several members of this class shifted into strong Class IV instead (Wright, 1925). These verbs, which include *break*, *speak*, *weave*, and *tread*, have never regularized.

A further complication is that in addition to the regularization of the strong verbs, a number of weak verbs were "irregularized" into strong classes whose members they resembled. In this way the weak verb *dive* adopted a strong past tense by analogy to strong *drive* or *rise*, or the weak verb *wear* took on strong inflection by analogy with *bear* and *tear*.

Finally, although the weak inflection generally behaved like the default form of the past tense, a number of borrowings and derived verbs adopted strong pasts instead. This was the case with *strive*, borrowed from Old French or the denominal *string*, among others.

Given the complexity of the historical facts, what determined which strong verbs regularized into the weak class and which were immune to regularization? In the remainder of this paper we present a simulation model in an attempt to this question.

Model

Previous work

The model presented here is an extension of previous work exploring how the interaction of frequency and phonological coherence accounts for the regularization process internal to the OE weak verb classes (Hare and Elman 1992). The earlier work showed, using artificial data, that when a number of inflection types competed for network resources, classes with extremely low type and token frequency could be learned successfully only if they displayed strong formal criteria of class membership. When these formal criteria were allowed to erode, the learning task became more difficult, and members of smaller, less regular classes were increasingly regularized into a larger and more consistent class. Interestingly, the pattern of regularization closely followed that seen in the historical data. In the current paper we will broaden the scope of this model to include both the strong and the weak systems, to show that similar principles can account for the complex set of facts outlined above. We also drop the simplifying assumption of artificial data, and develop a representation based on the actual forms of Old English verbs. This will allow us to explore the model's predictions in more detail.

The learning procedure used in the original model involved training a series of networks, metaphorically seen as "generations" of learners. The target for each new generation was the output, after learning, of the previous generation net. The motivation for that training regimen was to try to capture in a schematic way both the gradual nature of historical change and the causal role played by inaccurate transmission, without making strong claims about the precise relationship between the time course of training and real time. For this reason the generational learning technique will also be applied in the current model.

Motivation

All things being equal, if a number of patterns are learned in the same network, those with high type frequency will dominate those with lower type frequency. Thus if a limited number of strong verbs are stored in the same network as a large number of weaks, the strong verbs will be harder to learn and will tend to be absorbed into a weak class. However, the picture is more complicated than this implies. Under certain circumstances, strong verbs may remain strong and may even attract new members. There are two factors that determine which verbs or verb classes these are.

The first factor is token frequency. Plunkett and Marchman (1991, in press) have shown that high token frequency (i.e., occurrence of individual exemplars) improves network learning. This is consistent with the real-language data: many of the verbs that are still strong in Modern English are very frequent, and the highly irregular (although not strong) verbs like *be* and *go* are among the most frequent in the language.

Following the results of our previous model, we will argue that the second crucial factor is the degree of phonological definition in a verb class. There is ample evidence in the real language data for this claim. Bybee and colleagues (1982, 1983) have pointed out that modern strong verbs tend to cluster into sets according to a prototype or "family-resemblance" structure. The existence of such a prototype structure can aid lexical memory, they point out, allowing these items to be retained while others are regularized. Consider the following examples of modern strong verbs:

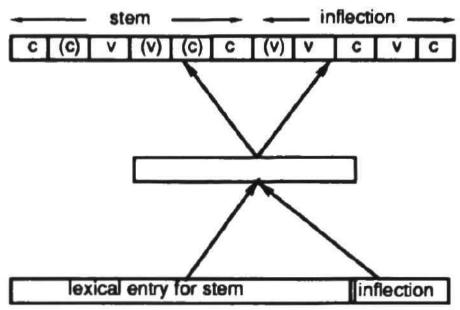
Class	Alternation	Examples
I	ay ~ ou	*dive, strive, drive, ride, rise, arise*
III	I ~ ^	*string, fling, cling, wring*
IV	E ~ O	*wear, tear, bear*
VI	ei ~ U	*take, shake, forsake*

For each alternation, the participating verbs share aspects of their phonological form. These formal cues to class membership promote learning in a network (Plunkett and Marchman 1991, Hare and Elman 1992). Thus even a low frequency item like *forsake* can be more easily than would be expected, given its token frequency, because of its similarity to higher-frequency classmates *take* and *shake*.

Together these factors predict that high frequency items will be relatively immune to regularization errors, while low frequency members of clearly defined classes will also benefit from their similarity to high frequency classmates. Over the course of generational learning these factors combine to predict which verb classes will lose their members to regularization, and which will maintain their form and even attracted new members.

Architecture and stimuli

The network architecture involved a feedforward network trained with backpropagation (Rumelhart, Hinton, & Williams, 1986). There were118 input units, divided into 100 "verb" units and 18 "inflection" units. The verb units represented abstract (and in this case, random) lexical entries; each verb root was assigned a unique pattern. The verb representations were developed by creating 700 100-bit vectors, where each bit had a 0.1 probability of being active so that on average each vector had 10 of the 100 bits on. These vectors were randomly assigned to the verbs of the training corpus as their input representation. There were also six person/number/tense combinations, each of which was represented as a three-bit pattern over the 18 inflection units. At each training iteration, a verb pattern and an inflection pattern were activated simultaneously. The task of the network was to produce the phonetic form of the inflected verb over the output units.



The output consisted of 165 units divided into 11 phoneme slots, each composed of an 15-bit distinctive feature representation. This created a 11-phoneme template over which a fully inflected verb could be represented. The first 6 phonemes were stem slots, allowing for a monosyllabic verb stem 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: the medial vowel used in the present tense of some weak verbs, a VC past tense affix, and a VC personal ending. Empty slots were activated to 0.5 on all units. Examples of a weak and a strong verb in two different inflections are given below:

Type	Inf.	Pret. pl.	Gloss
weak	-beot-i--an	-beot--odon	'boast'
strong	-h-elp---an	-h-olp---on	'help'

There were 106 strong verbs and 327 weak verbs in the data set, producing a total of 433 different stems. The subclasses of strong verbs will be described shortly. Strong verbs were taken from the first six OE strong classes in Flom (1930) and Wright (1945). As a rough measure of frequency we used the frequency of occurrence for each strong verb in a Chaucer concordance (Tatlock and Kennedy, 1963). Although this method relies on data from only one author writing late in the period under consideration, it gives a non-arbitrary estimate of frequencies for individual verbs that agrees in general both with claims about OE classes (Wright, 1945) and frequency in the current language. Verbs were presented to the net once for every 20 occurrences in the concordance.

101 of the 106 strong verbs were divided into 12 subclasses, following divisions given in the sources cited above. For ease of presentation we will rank these subclasses according their relative strength on the parameters of size, token frequency, and phonological cohesion, and refer to them as classes 1-12 according to their place on the list. While the list is not intended to reflect an exact gradient in immunity to regularization, Class 1 verbs are expected to be least vulnerable to regularization, and classes in the top half of the list are predicted to be spared relative to those in the bottom half. The table below gives the size, broken down into high- and low-frequency members, for each class, and indicates the dependability of phonological cues to class membership.

Table 1:
(numbers indicate the total number of verbs in each class, divided into high and low frequency members)

Class (actual OE verb subclass given in parent.)	Hi	Lo	Tot	Phonological cues apply to:
Most immune:				
1	4	9	13	all members
2	5	15	20	most
3	4	14	18	most
4	4	7	11	most
5	2	5	7	some
6	2	4	6	some
Most vulnerable:				
7	3	5	8	few members
8	0	3	3	none
9	0	6	6	none
10	0	4	4	none
11	1	1	1	none
12	2	1	1	none

Although the class numbering is our own, each of these was an inflectional verb class in Old and Middle English, and four of the first six classes are those given in the modern examples on the previous page. Our Class 1 is the historical class III, with members like *string*, *drink*, or *find*. All members of the class end in a stem-final NC cluster, and most begin with a CC cluster as well. Class 2 is made up of verbs like *drive*, *strive*, *ride*, with a long i in the present tense, often preceded by r (or a CC cluster with r), and often followed by a fricative. Class 3 contains the only strong verbs to consistently take a diphthong in all forms of the present tense. Class 4 is the *take* class, whose members have *a* in the present tenses, and often a stem-final -k or -g. At the bottom end of the list are small classes made up of infrequent members with no identifying characteristics. The other classes fall in between.

In addition to the 12 strong classes listed above, there were 5 idiosyncratic strong verbs, each of which will be treated as a class with only one member. Of the five single-word classes, *stand*, with a frequency of 300, and *come* (100), will be contrasted with the low-frequency singletons *mourn* (11), *shear* (6), and *laugh* (1).

The 327 weak verbs were taken from the same two authors and from Stark (1982). At the beginning of the training regimen, all weak verbs had a frequency of one.

Each verb was learned in six inflected forms: 1st singular present, 2nd singular present, present plural, 1st signaler preterit, preterit plural, and past participle. These specific forms were chosen because in combination they illustrate the significant distinctions among the subclasses of verbs.

Training

The training regimen was as follows. An initial network was taught to produce the verbs as they were formed in Early OE. The net was trained for 15 passes through the data set. At this point the patterns were learned well, but not perfectly. We collected the output produced by this net and used it as the teacher to train a second network. This second network was again trained for 15 epochs, at which point its output was collected and used as the teacher for a third net. The goal of this method of training was to reproduce the effects that can occur when a language learner attempts her task given variable or imperfectly learned target patterns. The errors in learning in the first network are propagated through subsequent nets. Our expectation was that patterns which are hardest to learn would lead to the most errors, and would be regularized to a more common pattern.

We repeated this training regimen for a total of 10 "generations". This training period was divided into three stages of learning, with a different data configuration at each. Generations 1-4 formed Stage I. Here each of the 327 weak verbs was presented once, while the 106 strong verbs varied in frequency as discussed above. While there are only 32% as many strong as weak verb types, weak verb tokens outnumber the weaks 3 to 2. This configuration was intended to mimic the effects of the relatively balanced strong/weak relationship in the early stages of OE.

Stage II began at generation 5, at which point we doubled the token frequency of the weak verbs. This was done to simulate the effect of the historical increase in the weak verb class. Weak verb tokens now outnumbered strong tokens 4 to 3. This stage continued until generation 8. At generation 9 we began Stage III by adding an additional token to each weak verb.

Results

For the output of each generation we measured the Euclidean distance between each output phoneme and all possible targets; if the closest target was the correct response this was considered a hit. We then computed the percentage of correct phoneme-by-phoneme hits per word and averaged this across class members to determine the correct percentage of hits per class. As would be expected, high frequency verbs consistently outperformed their low frequency counterparts. Table I gives the hit rate for the single-item classes at the first generation, at the end of Stage I (generation 4) and again at the end of Stage II (generation 8). Note that the high frequency verbs do well despite the influx of weak verbs after generation 4, while the low frequency singletons do poorly from the beginning, and lose ground rapidly with increase in weak verbs.

Table 2:
(numbers in Gen columns indicate percentage of hits/target)

Verb	Freq	Gen 1	Gen 4	Gen 8
come	100	97	97	71
stand	300	100	100	97
mourn	11	53	54	42
shear	6	70	52	26
laugh	1	70	53	34

The high frequency verbs are learned equally well in the multi-member classes. Over 9 generations only 3 (of 28) verbs with a presentation rate of two or more are regularized, two of these in the 9th generation. By contrast, the performance of the low frequency verbs is expected to vary across classes. To determine if this is the case we compared the actual number of correctly inflected low frequency verbs in each of the 12 multi-member subclasses.

These results are summarized in Table III. In the table Generation 0 refers to the original teacher (that is, to the number of potential low-frequency verbs in each class). The Table gives the number of low frequency verbs learned in each class at the beginning and end of Stage one, when strong and weak verbs were relatively balanced, at the beginning and end of Stage II, as the weak verbs increased their attraction, and after the second increase in weak verbs at Stage III.

As the table shows, all the strong verb classes lost members over time, and more so from generation 5 onwards as the weak classes became increasingly dominant.

Table 3:
(numbers indicate the number of corrects inflected low-frequency verbs in each class

Cl.	G 0	G 1	G 4	G 5	G 8	G 9
1	9	8	7	7	6	6
2	15	11	7	4	1	2
3	14	7	8	4	4	4
4	7	4	3	3	1	2
5	5	2	2	1	2	1
6	4	3	2	1	1	-
7	5	-	1	4	-	-
8	3	-	-	-	-	-
9	6	1	-	-	-	-
10	4	-	-	-	-	-
11	1	-	-	-	-	-
12	1	-	-	-	-	-

There is a clear difference, however, in the rate of loss. Classes 7-12 lose their low frequency members almost immediately. As was expected, the small classes' size and lack of phonological cues to class membership left these verbs vulnerable to regularization.

Classes 1-6 lose members more slowly, and even among these classes there are differences in the number of low-frequency items that are spared. Class 1, the *sing* class, is the most consistent phonologically and also the most successful in the learning task. Only 3 of its 9 low frequency verbs are regularized, even at a point when the weak verbs dominate the data set. Classes 2-5 also retain low frequency items into generation 9. Looking at actual items that are spared, we see that for example the class 5 survivor is *tear*, which shares the form of its high frequency classmate *bear*, or in class 6 *wef*, analogous to *swef*, is the last to change. This is consistent with the claim that in the larger, more consistent classes low frequency verbs that resemble their high frequency classmates share in their resistance to change. These classes attract new members from the weak verbs and other strong classes as well. Over the first 5 generations a total of 11 weak verbs dropped their affixes and were irregularized into one of the strong classes. All of these went into one of our first five strong classes, with six of the eleven going into Class 1. Verbs were vulnerable to this attraction if their phonological form fit that of the target group. For example, the weak verb *ri:* was adopted by Class 2 whose members included *drive*, *ride*; weak *driep* moved into Class 3, the diphthong class, and weak *thak* was attracted to Class 4 on the pattern of *wake* or *take*.

Discussion

The notion that variation and imperfect learning may lead to linguistic change over time is not new with us; Labor (1973) and many others have provided compelling examples of this phenomenon. In a similar way, it has long been noted that the particular patterns of change—which words are susceptible to change and which are resistant—can often be understood as the result of analogy. One limitation of such accounts, however, has been the apparently idiosyncratic and unpredictable nature of the effects. *Post hoc* explanations abound; principled predictions are rare.

One of the values of computational modeling is that it may permit a more precise understanding of mechanisms which are hypothesized to underlie behavior. Our work here is offered in that spirit. We have attempted to reproduce some of the essential characteristics of the Old English verbal system and to suggest a mechanism which might explain in a reasonably detailed manner the nature of some of the changes which occurred over time. We have seen that the model produces the overall effect of levelling, as many strong verbs assimilated to the weak category. Perhaps more interestingly, we have seen that the model also captures the less intuitively expected phenomena of strong verbs assimilating to other strong classes, and weak verbs and loan words becoming strong. Indeed, the specific verb classes and forms which underwent these changes in our model turn out to be precisely those for which this was true in the real language data.

This latter result also seems to us to bear on an issue which is currently of great interest: Is the English verbal system the product of one single mechanism (which might be modeled by a network) or of two (a network to handle irregulars and a rule system to handle regulars)? The latter hypothesis might predict that irregulars might move to regulars, but it does not obviously accommodate the opposite finding: That regulars and loan words may also, under certain conditions, be treated as irregulars. On the other hand, the single-mechanism account we have presented here not only accommodates the finding, but lets us predict exactly when it will occur.

References

Bybee, J.L., and Moder, C.L. 1983. Morphological Classes as Natural Categories, *Language* 59:251-270.

Bybee, J.L., and Slobin, D., 1982. Rules and Schemas in the Development and Use of the English Past Tense, *Language*, 58: 265-289.

Chomsky, N., and Halle, M., 1968. *The Sound Pattern of English.* NY: Harper and Row.

Daugherty, K., and Seidenberg, M., 1992. The Past Tense Revisited, in *The Proceedings of the 14th Annual Meeting of the Cognitive Science Society*, Princeton, N.J.: Erlbaum.

Flom, G. T., 1930. *Introductory Old English grammar and reader.* Boston: D.C. Heath.

Hare, M., and Elman, J.L., 1992. A Connectionist Account of English Inflectional Morphology: Evidence from Language Change, in *The Proceedings of the 14th Annual Meeting of the Cognitive Science Society*, Princeton, N.J.: Erlbaum.

Kim, J., et. al. 1991. Why No Mere Mortal Has Ever Flown Out to Center Field, *Cognitive Science* 15: 173-218.

Kiparsky, P., 1971. Linguistic Universals and Linguistic Change.

Labov, W., 1973. Sociolinguistic Patterns, Philadelphia: University of Pennsylvania Press.

MacWhinney, B., and Leinbach,J., 1991. Implementations are Not Conceptualizations: Revising the Verb Learning Model, *Cognition* 40:121-157.

Pinker, S., and Prince, A., 1988. On Language and Connectionism: Analysis of a Parallel Distributed Processing Model of Language Acquisition, *Cognition* 28: 73-193.

Plunkett, K., and Marchman, V., (in press) From Rote Learning to System Building, *Cognition.*

Plunkett, K., and Marchman, V., 1991. U-shaped learning and frequency effects in a multilayered perceptron: Implications for child language acquisition, *Cognition* 38: 73-193

Prasada, S., and Prince, S., (in press) Generalization of Regular and Irregular Morphological Patterns, *Language and Cognitive Processes.*

Quirk, R., and Wrenn, C.L., 1975. *An Old English grammar*, London: Methuen.

Rumelhart, D.E., Hinton, G., and Williams, R., 1986. Learning Internal Representations by Error Propagation, in In *Parallel Distributed Processing, Vol. I* (Rumelhart, D.E., and McClelland, J., Eds.), Cambridge, MA: MIT Press.

Rumelhart., D.E., and McClelland, J., 1986. On Learning the Past Tense of English Verbs. In *Parallel Distributed Processing, Vol. II* (Rumelhart, D.E., and McClelland, J., Eds.), Cambridge, MA: MIT Press.

Slobin, D., 1977. Language change in Childhood and in History, in *Language Learning and Thought*, (J. Macnamara, Ed.), New York: Academic Press.

Stark, D., 1982. *Old English Weak Verbs*, Turbingen: Niemeyer.

Tatlock, J., and Kennedy, A,. (1963) *A Concordance to the Complete Works of Geoffrey Chaucer*, Gloucester, Mass.:Peter Smith.

Wright, J. and Wright, E., 1925. Old English Grammar, Oxford: Oxford U. Press.