

David E. Rumelhart and
James L. McClelland

The Issue

Scholars of language and psycholinguistics have been among the first to stress the importance of rules in describing human behavior. The reason for this is obvious. Many aspects of language can be characterized by rules, and the speakers of natural languages speak the language correctly. Therefore, systems of rules are useful in characterizing what they will and will not say. Though we all make mistakes when we speak, we have a pretty good ear for what is right and what is wrong—and our judgments of correctness—or grammaticality—are generally even easier to characterize by rules than actual utterances.

On the evidence that what we will and won't say and what we will and won't accept can be characterized by rules, it has been argued that, in some sense, we "know" the rules of our language. The sense in which we know them is not the same as the sense in which we know such "rules" as "*i* before *e* except after *c*," however, since we need not necessarily be able to state the rules explicitly. We know them in a way that allows us to use them to make judgments of grammaticality, it is often said, or to speak and understand, but this knowledge is not in a form or location that permits it to be encoded into a communicable verbal statement. Because of this, this knowledge is said to be *implicit*.

So far there is considerable agreement. However, the exact characterization of implicit knowledge is a matter of great controversy. One view, which is perhaps extreme but is nevertheless quite clear, holds that the rules of language are stored in explicit form as propositions, and are used by language production, comprehension, and judgment mechanisms. These propositions cannot be described verbally only because they are sequestered in a specialized subsystem which is used in lan-

From D. Rumelhart and J. McClelland, On learning the past tenses of English verbs, in J. McClelland, D. Rumelhart, and the PDP Research Group, *Parallel distributed processing*, vol. 2 (1986). Cambridge, MA: MIT Press. Reprinted by permission.

guage processing, or because they are written in a special code that only the language processing system can understand. This view we will call the *explicit inaccessible rule view*.

On the explicit inaccessible rule view, language acquisition is thought of as the process of inducing rules. The language mechanisms are thought to include a subsystem—often called the *language acquisition device* (LAD)—whose business it is to discover the rules. A considerable amount of effort has been expended on the attempt to describe how the LAD might operate, and there are a number of different proposals which have been laid out. Generally, though, they share three assumptions:

- The mechanism hypothesizes explicit inaccessible rules.
- Hypotheses are rejected and replaced as they prove inadequate to account for the utterances the learner hears.
- The LAD is presumed to have *innate* knowledge of the possible range of human languages and, therefore, is presumed to consider only hypotheses within the constraints imposed by a set of *linguistic universals*.

The recent book by Pinker (1984) contains a state-of-the-art example of a model based on this approach.

We propose an alternative to explicit inaccessible rules. We suggest that lawful behavior and judgments may be produced by a mechanism in which there is no explicit representation of the rule. Instead, we suggest that the mechanisms that process language and make judgments of grammaticality are constructed in such a way that their performance is characterizable by rules, but that the rules themselves are not written in explicit form anywhere in the mechanism. An illustration of this view, which we owe to Bates (1979), is provided by the honeycomb. The regular structure of the honeycomb arises from the interaction of forces that wax balls exert on each other when compressed. The honeycomb can be described by a rule, but the mechanism which produces it does not contain any statement of this rule.

In our earlier work with the interactive activation model of word perception (McClelland and Rumelhart 1981, Rumelhart and McClelland 1981, 1982), we noted that lawful behavior emerged from the interactions of a set of word and letter units. Each word unit stood for a particular word and had connections to units for the letters of the word. There were no separate units for common letter clusters and no explicit provision for dealing differently with orthographically regular letter sequences—strings that accorded with the rules of English—as opposed to irregular sequences. Yet the model did behave differently with orthographically regular nonwords than it behaved with words. In fact,

the model simulated rather closely a number of results in the word perception literature relating to the finding that subjects perceive letters in orthographically regular letter strings more accurately than they perceive letters in irregular, random letter strings. Thus, the behavior of the model was lawful even though it contained no explicit rules.

It should be said that the pattern of perceptual facilitation shown by the model did not correspond exactly to any system of orthographic rules that we know of. The model produced as much facilitation, for example, for special nonwords like *SLNT*, which are clearly irregular, as it did for matched regular nonwords like *SLET*. Thus, it is not correct to say that the model exactly mimicked the behavior we would expect to emerge from a system which makes use of explicit orthographic rules. However, neither do human subjects. Just like the model, they showed equal facilitation for vowelless strings like *SLNT* as for regular nonwords like *SLET*. Thus, human perceptual performance seems, in this case at least, to be characterized only approximately by rules.

Some people have been tempted to argue that the behavior of the model shows that we can do without linguistic rules. We prefer, however, to put the matter in a slightly different light. There is no denying that rules still provide a fairly close characterization of the performance of our subjects. And we have no doubt that rules are even more useful in characterizations of sentence production, comprehension, and grammaticality judgments. We would only suggest that parallel distributed processing models may provide a mechanism sufficient to capture lawful behavior, without requiring the postulation of explicit but inaccessible rules. Put succinctly, our claim is that PDP [parallel distributed processing] models provide an alternative to the explicit but inaccessible rules account of implicit knowledge of rules.

We can anticipate two kinds of arguments against this kind of claim. The first kind would claim that although certain types of rule-guided behavior might emerge from PDP models, the models simply lack the computational power needed to carry out certain types of operations which can be easily handled by a system using explicit rules. We believe that this argument is simply mistaken. We discuss the issue of computational power of PDP models elsewhere in this work, with some applications to sentence processing. The second kind of argument would be that the details of language behavior, and, indeed, the details of the language acquisition process, would provide unequivocal evidence in favor of a system of explicit rules.

It is this latter kind of argument we wish to address in the present chapter. We have selected a phenomenon that is often thought of as demonstrating the acquisition of a linguistic rule. And we have developed a parallel distributed processing model that learns in a natural way to behave in accordance with the rule, mimicking the general trends seen in the acquisition data.

The Phenomenon

The phenomenon we wish to account for is actually a sequence of three stages in the acquisition of the use of past tense by children learning English as their native tongue. Descriptions of development of the use of the past tense may be found in Brown (1973), Ervin (1964), and Kuczaj (1977).

In Stage 1, children use only a small number of verbs in the past tense. Such verbs tend to be very high-frequency words, and the majority of these are irregular. At this stage, children tend to get the past tenses of these words correct if they use the past tense at all. For example, a child's lexicon of past-tense words at this stage might consist of *came, got, gave, looked, needed, took, and went*. Of these seven verbs, only two are regular—the other five are generally idiosyncratic examples of irregular verbs. In this stage, there is no evidence of the use of the rule—it appears that children simply know a small number of separate items.

In Stage 2, evidence of implicit knowledge of a linguistic rule emerges. At this stage, children use a much larger number of verbs in the past tense. These verbs include a few more irregular items, but it turns out that the majority of the words at this stage are examples of the *regular* past tense in English. Some examples are *wiped* and *pulled*.

The evidence that the Stage 2 child actually has a linguistic rule comes not from the mere fact that he or she knows a number of regular forms. There are two additional and crucial facts:

- The child can now generate a past tense for an invented word. For example, Berko (1958) has shown that if children can be convinced to use *rick* to describe an action, they will tend to say *ricked* when the occasion arises to use the word in the past tense.
- Children now *incorrectly* supply regular past-tense endings for words which they used correctly in Stage 1. These errors may involve either adding *ed* to the root as in *comed* /kɹɛnd/, or adding *ed* to the irregular past tense form as in *comed* /kɹɛnd/ (Ervin 1964; Kuczaj 1977).

Such findings have been taken as fairly strong support for the assertion that the child at this stage has acquired the past-tense "rule." To quote Berko (1958):

If a child knows that the plural of *witch* is *witches*, he may simply have memorized the plural form. If, however, he tells us that the plural of *gutch* is *gutches*, we have evidence that he actually knows, albeit unconsciously, one of those rules which the descriptive linguist, too, would set forth in his grammar. (p. 151)

In Stage 3, the regular and irregular forms coexist. That is, children have regained the use of the correct irregular forms of the past tense,

while they continue to apply the regular form to new words they learn. Regularizations persist into adulthood—in fact, there is a class of words for which either a regular or an irregular version are both considered acceptable—but for the commonest irregulars such as those the child acquired first, they tend to be rather rare. At this stage there are some clusters of exceptions to the basic, regular past-tense pattern of English. Each cluster includes a number of words that undergo identical changes from the present to the past tense. For example, there is *ing/lang* cluster, an *ing/ung* cluster, an *ee/it* cluster, etc. There is also a group of words ending in /d/ or /t/ for which the present and past are identical.

Table 24.1 summarizes the major characteristics of the three stages.

Variability and Gradualness

The characterization of past-tense acquisition as a sequence of three stages is somewhat misleading. It may suggest that the stages are clearly demarcated and that performance in each stage is sharply distinguished from performance in other stages.

In fact, the acquisition process is quite gradual. Little detailed data exists on the transition from Stage 1 to Stage 2, but the transition from Stage 2 to Stage 3 is quite protracted and extends over several years (Kuczaj 1977). Further, performance in Stage 2 is extremely variable. Correct use of irregular forms is never completely absent, and the same child may be observed to use the correct past of an irregular, the base+ed form, and the past+ed form, within the same conversation.

Other Facts About Past-Tense Acquisition

Beyond these points, there is now considerable data on the detailed types of errors children make throughout the acquisition process, both from Kuczaj (1977) and more recently from Bybee and Slobin (1982). We will consider aspects of these findings in more detail below. For now, we mention one intriguing fact: According to Kuczaj (1977), there is an interesting difference in the errors children make to irregular verbs at different points in Stage 2. Early on, regularizations are typically of the base+ed form, like *goed*; later on, there is a large increase in the frequency of past+ed errors, such as *wented*.

Table 24.1 Characteristics of the three stages of past-tense acquisition

Verb Type	Stage 1	Stage 2	Stage 3
Early verbs	Correct	Regularized	Correct
Regular	—	Correct	Correct
Other irregular	—	Regularized	Correct or regularized
Novel	—	Regularized	Regularized

The Model

The goal of our simulation of the acquisition of past tense was to simulate the three-stage performance summarized in Table 24.1, and to see whether we could capture other aspects of acquisition. In particular, we wanted to show that the kind of gradual change characteristic of normal acquisition was also a characteristic of our distributed model, and we wanted to see whether the model would capture detailed aspects of the phenomenon, such as the change in error type in later phases of development and the change in differences in error patterns observed for different types of words.

We were not prepared to produce a full-blown language processor that would learn the past tense from full sentences heard in everyday experience. Rather, we have explored a very simple past-tense learning environment designed to capture the essential characteristics necessary to produce the three stages of acquisition. In this environment, the model is presented, as learning experiences, with pairs of inputs—one capturing the phonological structure of the root form of a word and the other capturing the phonological structure of the correct past-tense version of the word. The behavior of the model can be tested by giving it just the root form of a word and examining what it generates as its “current guess” of the corresponding past-tense form.

Structure of the Model

The basic structure of the model is illustrated in Figure 24.1. The model consists of two basic parts: (a) a simple *pattern associator* network similar to those studied by Kohonen (1977, 1984) which learns the relationships

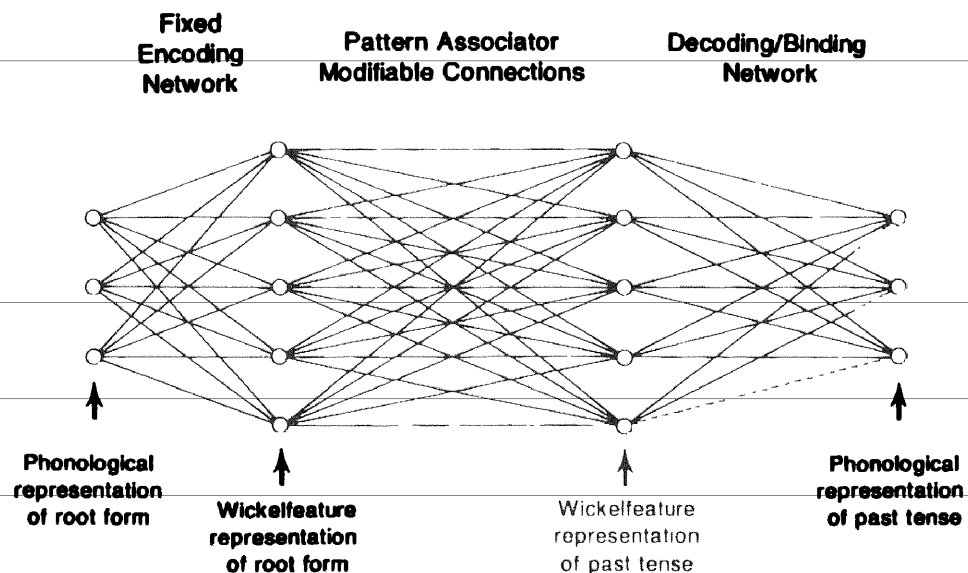


Figure 24.1 The basic structure of the model.

between the base form and the past-tense form, and (b) a decoding network that converts a featural representation of the past-tense form into a phonological representation. All learning occurs in the pattern associator; the decoding network is simply a mechanism for converting a featural representation which may be a near miss to any phonological pattern into a legitimate phonological representation. Our primary focus here is on the pattern associator.

Units The pattern associator contains two pools of units. One pool, called the input pool, is used to represent the input pattern corresponding to the root form of the verb to be learned. The other pool, called the output pool, is used to represent the output pattern generated by the model as its current guess as to the past tense corresponding to the root form represented in the inputs.

Each unit stands for a particular feature of the input or output string. The particular features we used are important to the behavior of the model, so they are described in a separate section below [omitted in this edition—Ed.].

Connections The pattern associator contains a modifiable connection linking each input unit to each output unit. Initially, these connections are all set to 0 so that there is no influence of the input units on the output units. Learning involves modification of the strengths of these interconnections, as described below.

Operation of the Model

On test trials, the simulation is given a phoneme string corresponding to the root of a word. It then performs the following actions. First, it encodes the root string as a pattern of activation over the input units. The encoding scheme used is described below. Node activations are discrete in this model, so the activation values of all the units that should be on to represent this word are set to 1, and all the others are set to 0. Then, for each output unit, the model computes the net input to it from all of the weighted connections from the input units. The net input is simply the sum over all input units of the input unit activation times the corresponding weight. Thus, algebraically, the net input to output unit i is

$$net_i = \sum_j a_j w_{ij}$$

where a_j represents the activation of input unit j , and w_{ij} represents the weight from unit j to unit i .

Each unit has a threshold, θ , which is adjusted by the learning procedure that we will describe in a moment. The probability that the unit is turned on depends on the amount the net input exceeds the threshold. The *logistic* probability function is used here as in the Boltzmann

machine and in harmony theory to determine whether the unit should be turned on. The probability is given by

$$p(a_i = 1) = \frac{1}{1 + e^{-(net_i - \theta_i)/T}} \quad (24.1)$$

where T represents the temperature of the system. The logistic function is shown in Figure 24.2. The use of this probabilistic response rule allows the system to produce different responses on different occasions with the same network. It also causes the system to learn more slowly so the effect of regular verbs on the irregulars continues over a much longer period of time. The temperature, T , can be manipulated so that at very high temperatures the response of the units is highly variable; with lower values of T , the units behave more like *linear threshold units*.

Since the pattern associator built into the model is a one-layer net with no feedback connections and no connections from one output unit to another or from one input unit to another, iterative computation is of no benefit. Therefore, the processing of an input pattern is a simple matter of first calculating the net input to each output unit and then setting its activation probabilistically on the basis of the logistic equation given above. The temperature T only enters in setting the variability of the output units; a fixed value of T was used throughout the simulations.

To determine how well the model did at producing the correct output, we simply compare the pattern of output Wickelphone activations to the pattern that the correct response would have generated. [Wickelphones are elements of a particular scheme of phonological representation.—Ed.] To do this, we first translate the correct response into a

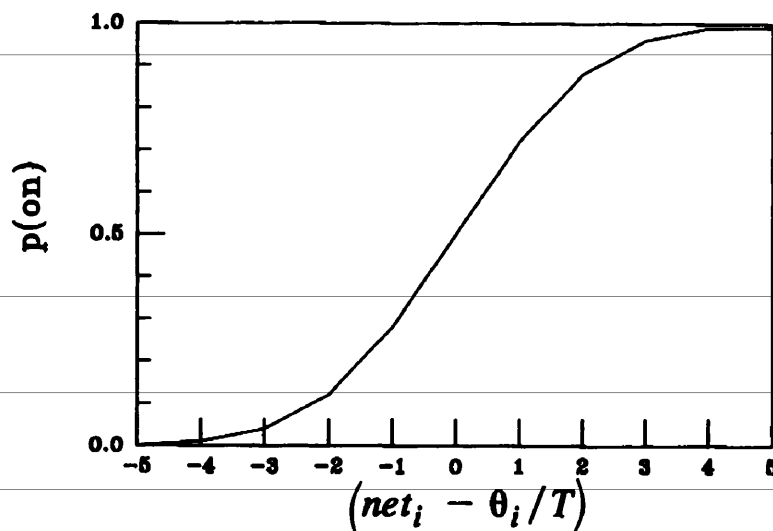


Figure 24.2 The logistic function used to calculate probability of activation. The x-axis shows values of $(net_i - \theta_i/T)$, and the y-axis indicates the corresponding probability that unit i will be activated.

target pattern of activation for the output units, based on the same encoding scheme used for the input units. We then compare the obtained pattern with the target pattern on a unit-by-unit basis. If the output perfectly reproduces the target, then there should be a 1 in the output pattern wherever there is a 1 in the target. Such cases are called *hits*, following the conventions of signal detection theory (Green and Swets 1966). There should also be a 0 in the output whenever there is a 0 in the target. Such cases are called *correct rejections*. Cases in which there are 1s in the output but not in the target are called *false alarms*, and cases in which there are 0s in the output that should be present in the input are called *misses*. A variety of measures of performance can be computed. We can measure the percentage of output units that match the correct past tense, or we can compare the output to the pattern for any other response alternative we might care to evaluate. This allows us to look at the output of the system independently of the decoding network. We can also employ the decoding network and have the system synthesize a phonological string. We can measure the performance of the system either at the featural level or at the level of strings of phonemes. We shall employ both of these mechanisms in the evaluation of different aspects of the overall model.

Learning

On a learning trial, the model is presented with both the root form of the verb and the target. As on a test trial, the pattern associator network computes the output it would generate from the input. Then, for each output unit, the model compares its answer with the target. Connection strengths are adjusted using the classic *perceptron convergence procedure* (Rosenblatt 1962). The perceptron convergence procedure is simply a discrete variant of the delta rule presented earlier in this work. The exact procedure is as follows: We can think of the target as supplying a teaching input to each output unit, telling it what value it ought to have. When the actual output matches the target output, the model is doing the right thing and so none of the weights on the lines coming into the unit are adjusted. When the computed output is 0 and the target says it should be 1, we want to increase the probability that the unit will be active the next time the same input pattern is presented. To do this, we increase the weights from all of the input units that are active by a small amount η . At the same time, the threshold is also reduced by η . When the computed output is 1 and the target says it should be 0, we want to decrease the probability that the unit will be active the next time the same input pattern is presented. To do this, the weights from all of the input units that are active are reduced by η , and the threshold is increased by η . In all of our simulations, the value of η is simply set to 1. Thus, each change in a weight is a unit change, either up or down. For nonstochastic units, it is well known that the perceptron convergence procedure will find a set of weights that will

allow the model to get each output unit correct, provided that such a set of weights exists. For the stochastic case, it is possible for the learning procedure to find a set of weights that will make the probability of error as low as desired. Such a set of weights exists if a set of weights exists that will always get the right answer for nonstochastic units. . . .

The Simulations

The simulations described in this section are concerned with demonstrating three main points:

- That the model captures the basic three-stage pattern of acquisition.
- That the model captures most aspects of differences in performance on different types of regular and irregular verbs.
- That the model is capable of responding appropriately to verbs it has never seen before, as well as to regular and irregular verbs actually experienced during training.

In the sections that follow we will consider these three aspects of the model's performance in turn.

The corpus of verbs used in the simulations consisted of a set of 506 verbs. All verbs were chosen from the Kucera and Francis (1967) word list and were ordered according to frequency of their gerund form. We divided the verbs into three classes: 10 high-frequency verbs, 410 medium-frequency verbs, and 86 low-frequency verbs. The ten highest frequency verbs were: *come* (/kʌm/), *get* (/get/), *give* (/giv/), *look* (/lʊk/), *take* (/tak/), *go* (/go/), *have* (/hav/), *live* (/liv/), and *feel* (/fi:l/). There is a total of 8 irregular and 2 regular verbs among the top 10. Of the medium-frequency verbs, 334 were regular and 76 were irregular. Of the low-frequency verbs, 72 were regular and 14 were irregular.

The Three-Stage Learning Curve

The results described in this and the following sections were obtained from a single (long) simulation run. The run was intended to capture approximately the experience with past tense of a young child picking up English from everyday conversation. Our conception of the nature of this experience is simply that the child learns first about the present and past tenses of the highest frequency verbs; later on, learning occurs for a much larger ensemble of verbs, including a much larger proportion of regular forms. Although the child would be hearing present and past tenses of all kinds of verbs throughout development, we assume that he is only able to learn past tenses for verbs that he has already mastered fairly well in the present tense.

To simulate the earliest phase of past-tense learning, the model was first trained on the 10 high-frequency verbs, receiving 10 cycles of

training presentations through the set of 10 verbs. This was enough to produce quite good performance on these verbs. We take the performance of the model at this point to correspond to the performance of a child in Phase 1 of acquisition. To simulate later phases of learning, the 410 medium-frequency verbs were added to the first 10 verbs, and the system was given 190 more learning trials, with each trial consisting of one presentation of each of the 420 verbs. The responses of the model early on in this phase of training correspond to Phase 2 of the acquisition process; its ultimate performance at the end of 190 exposures to each of the 420 verbs corresponds to Phase 3. At this point, the model exhibits almost errorless performance on the basic 420 verbs. Finally, the set of 86 lower-frequency verbs were presented to the system and the transfer responses to these were recorded. During this phase, connection strengths were not adjusted. Performance of the model on these transfer verbs is considered in a later section.

We do not claim, of course, that this training experience exactly captures the learning experience of the young child. It should be perfectly clear that this training experience exaggerates the difference between early phases of learning and later phases, as well as the abruptness of the transition to a larger corpus of verbs. However, it is generally observed that the early, rather limited vocabulary of young children undergoes an explosive growth at some point in development (Brown 1973). Thus, the actual transition in a child's vocabulary of verbs would appear quite abrupt on a time-scale of years so that our assumptions about abruptness of onset may not be too far off the mark.

Figure 24.3 shows the basic results for the high frequency verbs. What we see is that during the first 10 trials there is no difference

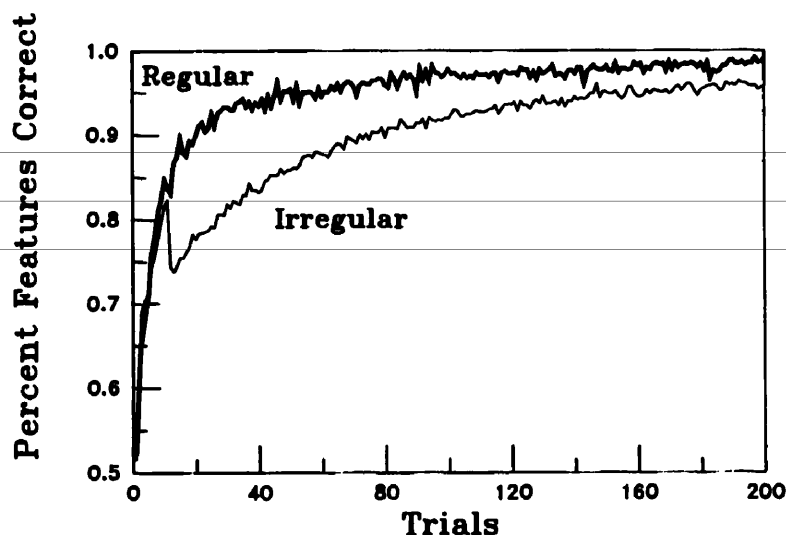


Figure 24.3 The percentage of correct features for regular and irregular high-frequency verbs as a function of trials.

between regular and irregular verbs. However, beginning on Trial 11 when the 410 midfrequency verbs were introduced, the regular verbs show better performance. It is important to notice that there is no interfering effect on the regular verbs as the midfrequency verbs are being learned. There is, however, substantial interference on the irregular verbs. This interference leads to a dip in performance on the irregular verbs. Equality of performance between regular and irregular verbs is never again attained during the training period. This is the so-called U-shaped learning curve for the learning of the irregular past tense. Performance is high when only a few high-frequency, largely irregular verbs are learned, but then drops as the bulk of lower-frequency regular verbs are being learned.

We have thus far only shown that performance on high-frequency irregular verbs drops; we have not said anything about the nature of the errors. To examine this question, the response strength of various possible response alternatives must be compared. To do this, we compared the strength of response for several different response alternatives. We compared strengths for the correct past tense, the present, the base+ed and the past+ed. Thus, for example with the verb *give* we compared the response strength of /gAV/, /giv/, /givd/, and /gAVd/. We determined the response strengths by assuming that these response alternatives were competing to account for the features that were actually turned on in the output. For present purposes, suffice it to say that each alternative gets a score that represents the percentage of the total features that it accounts for. If two alternatives both account for a given feature, they divide the score for that feature in proportion to the number of features each accounts for uniquely. We take these response strengths to correspond roughly to relative response probabilities, though we imagine that the actual generation of overt responses is accomplished by a different version of the binding network, described below. In any case, the total strength of all the alternatives cannot be greater than 1, and if a number of features are accounted for by none of the alternatives, the total will be less than 1.

Figure 24.4 compares the response strengths for the correct alternative to the combined strength of the regularized alternatives.² Note in the figure that during the first 10 trials the response strength of the correct alternative grows rapidly to over .5 while that of the regularized alternative drops from about .2 to .1. After the midfrequency verbs are introduced, the response strength for the correct alternative drops rapidly while the strengths of regularized alternatives jump up. From about Trials 11 through 30, the regularized alternatives together are stronger than the correct response. After about Trial 30, the strength of the correct response again exceeds the regularized alternatives and continues to grow throughout the 200-trial learning phase. By the end, the correct response is much the strongest with all other alternatives below .1.

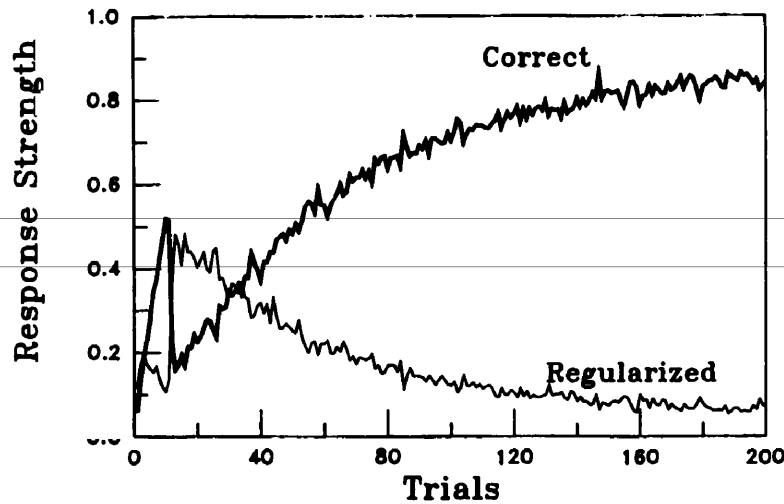


Figure 24.4 Response strengths for the high-frequency irregular verbs. The response strengths for the correct responses are compared with those for the regularized alternatives as a function of trials.

The rapidity of the growth of the regularized alternatives is due to the sudden influx of the medium-frequency verbs. In real life we would expect the medium-frequency verbs to come in somewhat more slowly so that the period of maximal regularization would have a somewhat slower onset.

Figure 24.5 shows the same data in a slightly different way. In this case, we have plotted the ratio of the correct response to the sum of the correct and regularized response strengths. Points on the curve below the .5 line are in the region where the regularized response is greater than the correct response. Here we see clearly the three stages. In the first stage, the first 10 trials of learning, performance on these high-frequency verbs is quite good. Virtually no regularization takes place. During the next 20 trials, the system regularizes and systematically makes errors on the verbs that it previously responded to correctly. Finally, during the remaining trials the model slowly eliminates the regularization responses as it approaches adult performance.

In summary, then, the model captures the three phases of learning quite well, as well as the gradual transition from Phase 2 to Phase 3. It does so without any explicit learning of rules. The regularization is the product of the gradual tuning of connection strengths in response to the predominantly regular correspondence exhibited by the medium-frequency words. It is not quite right to say that individual pairs are being stored in the network in any simple sense. The connection strengths the model builds up to handle the irregular forms do not represent these items in any separable way; they represent them in the way they must be represented to be stored along with the other verbs in the same set of connections. . . .

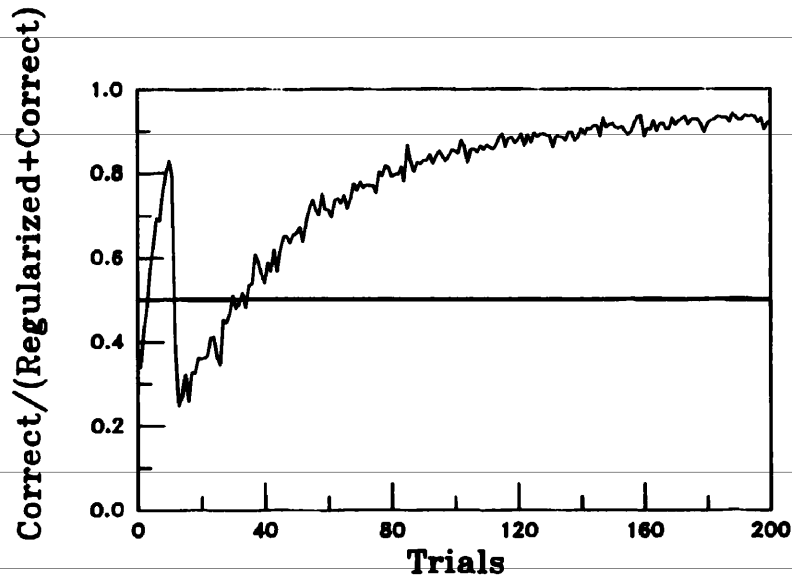


Figure 24.5 The ratio of the correct response to the sum of the correct and regularized response. Points on the curve below the .5 line are in the region where the regularized response is greater than the correct response.

Conclusions

We have shown that our simple learning model shows, to a remarkable degree, the characteristics of young children learning the morphology of the past tense in English. We have shown how our model generates the so-called U-shaped learning curve for irregular verbs and that it exhibits a tendency to overgeneralize that is quite similar to the pattern exhibited by young children. Both in children and in our model, the verb forms showing the most regularization are pairs such as *know/knew* and *see/saw*, whereas those showing the least regularization are pairs such as *feel/felt* and *catch/caught*. Early in learning, our model shows the pattern of more no-change responses to verbs ending in *t/d* whether or not they are regular verbs, just as young children do. The model, like children, can generate the appropriate regular past-tense form to unfamiliar verbs whose base form ends in various consonants or vowels. Thus, the model generates an */d/* suffix for verbs ending in *t/d*, a */t/* suffix for verbs ending in an unvoiced consonant, and a */d/* suffix for verbs ending in a voiced consonant or vowel.

In the model, as in children, different past-tense forms for the same word can coexist at the same time. On rule accounts, such *transitional* behavior is puzzling and difficult to explain. Our model, like human children, shows a relatively larger proportion of past+ed regularizations later in learning. Our model, like learners of English, will sometimes generate past-tense forms to novel verbs which show sensitivities to the subregularities of English as well as the major regularities. Thus, the past of *cring* can sometimes be rendered *crang* or *crung*. In short,

our simple learning model accounts for all of the major features of the acquisition of the morphology of the English past tense.

In addition to our ability to account for the major *known* features of the acquisition process, there are also a number of predictions that the model makes which have yet to be reported. These include:

- We expect relatively more past+ed regularizations to irregulars whose correct past form *does not* involve a modification of the final phoneme of the base form.
- We expect that early in learning, a no-change response will occur more frequently to a CVC monosyllable ending in *t/d* than to a more complex base verb form.
- We expect that the double inflection responses (/dript^d/) will occasionally be made by native speakers and that they will occur more frequently to verbs whose stem ends in /p/ or /k/.

The model is very rich and there are many other more specific predictions which can be derived from it and evaluated by a careful analysis of acquisition data.

We have, we believe, provided a distinct alternative to the view that children learn the rules of English past-tense formation in any explicit sense. We have shown that a reasonable account of the acquisition of past tense can be provided without recourse to the notion of a "rule" as anything more than a *description* of the language. We have shown that, for this case, there is no *induction problem*. The child need not figure out what the rules are, nor even that there are rules. The child need not decide whether a verb is regular or irregular. There is no question as to whether the inflected form should be stored directly in the lexicon or derived from more general principles. There isn't even a question (as far as generating the past-tense form is concerned) as to whether a verb form is one encountered many times or one that is being generated for the first time. A uniform procedure is applied for producing the past-tense form in every case. The base form is supplied as input to the past-tense network and the resulting pattern of activation is interpreted as a phonological representation of the past form of that verb. This is the procedure whether the verb is regular or irregular, familiar or novel.

In one sense, every form must be considered as being derived. In this sense, the network can be considered to be one large rule for generating past tenses from base forms. In another sense, it is possible to imagine that the system simply stores a set of rote associations between base and past-tense forms with novel responses generated by "on-line" generalizations from the stored exemplars.

Neither of these descriptions is quite right, we believe. Associations are simply stored in the network, but because we have a *superpositional* memory, similar patterns blend into one another and reinforce each

other. If there were no similar patterns (i.e., if the featural representations of the base forms of verbs were orthogonal to one another) there would be no generalization. The system would be unable to generalize and there would be no regularization. It is statistical relationships among the base forms themselves that determine the pattern of responding. The network merely reflects the statistics of the featural representations of the verb forms.

We chose the study of acquisition of past tense in part because the phenomenon of regularization is an example often cited in support of the view that children do respond according to general rules of language. Why otherwise, it is sometimes asked, should they generate forms that they have never heard? The answer we offer is that they do so because the past tenses of similar verbs they are learning show such a consistent pattern that the generalization from these similar verbs outweighs the relatively small amount of learning that has occurred on the irregular verb in question. We suspect that essentially similar ideas will prove useful in accounting for other aspects of language acquisition. We view this work on past-tense morphology as a step toward a revised understanding of language knowledge, language acquisition, and linguistic information processing in general.

Acknowledgments

This research was supported by ONR Contracts N00014-82-C-0374, NR 667-483 and N00014-79-C-0323, NR 667-437, by a grant from the System Development Foundation, and by a Research Scientist Career Development Award MH00385 to the second author from the National Institute of Mental Health.

Notes

1. The notation of phonemes used in this chapter is somewhat nonstandard. It is derived from the computer-readable dictionary containing phonetic transcriptions of the verbs used in the simulations.

2. Unless otherwise indicated, the regularized alternatives are considered the base+ed and past+ed alternative. In most cases the base+ed alternative is much stronger than the past+ed alternative.

References

Bates, E. (1979). *Emergence of symbols*. New York: Academic Press.

Berko, J. (1958). The child's learning of English morphology. *Word* 14, 150-177.

Brown, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.

Bybee, J. L., and D. I. Slobin (1982). Rules and schemas in the development and use of the English past tense. *Language* 58, 265–289.

Ervin, S. (1964). Imitation and structural change in children's language. In E. Lenneberg, ed., *New directions in the study of language*. Cambridge, MA: MIT Press.

Green, D. M., and J. A. Swets (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Kohonen, T. (1977). *Associative memory: A system theoretical approach*. New York: Springer.

Kohonen, T. (1984). *Self-organization and associative memory*. Berlin: Springer-Verlag.

Kucera, H., and W. Francis (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Kuczaj, S. A. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior* 16, 589–600.

McClelland, J. L., and D. E. Rumelhart (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review* 88, 375–407.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Rosenblatt, F. (1962). *Principles of neurodynamics*. New York: Spartan.

Rumelhart, D. E., and J. L. McClelland (1981). Interactive processing through spreading activation. In A. M. Lesgold and C. A. Perfetti, eds., *Interactive processes in reading*. Hillsdale, NJ: Erlbaum.

Rumelhart, D. E., and J. L. McClelland (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review* 89, 60–94.