Predicting Irregular Past Tenses

Comparing Symbolic and Connectionist models against Native English Speakers

Charles X. Ling
Department of Computer Science
University of Western Ontario
London, Ontario, Canada N6A 5B7
Email: ling@csd.uwo.ca

Abstract

Learning the past tense of English verbs has become a landmark task for testing the adequacy of cognitive modeling. We review a set of intriguing psychological phenomena that any modeling of past-tense acquisition has to account for. Traditional grammatical theories fail to explain phenomena of irregular verbs, while connectionist models, which require no symbols and explicit rules, fail on regular verbs. We present a general-purpose symbolic pattern associator (SPA for short) which learns a set of sufficient and necessary symbolic rules for both distinguishing and predicting regular and irregular verbs. Our all-rule theory is similar in spirit to Pinker's (1991, 1993) modular hypothesis, and is able to account for most psychological phenomena in past-tense acquisition. Even on the task of *irregular* past-tense generalization, the SPA is judged to be slightly more plausible than the connectionist model by adult native English speakers. Our results support the view that language acquisition and processing should be better modeled by symbolic, rather than connectionist, systems.

Introduction

Learning the past tense of English verbs, a minor component of language acquisition and processing, has received extensive study in the last few years. One of the reasons is that the task is quite isolated from syntax, semantics, and other morphological modules of the language learning[1], and thus can be studied in great depth. More importantly, there is a set of intriguing psychological phenomena and properties that any acquisition system must account for. Traditional grammatical theories are unable to explain many critical aspects of irregular inflections. In 1988, Rumelhart and McClelland (1986) design and implement a connectionist system for the modeling of past-tense acquisition. Claims have been made that such connectionist models, while requiring no symbol processing, grammatical rules, or explicit representation as in the traditional grammatical theories, are better models for the past-tense acquisition, and for the language acquisition in general. Over the years a number of criticisms of connectionist modeling appeared (Pinker & Prince, 1988; Lachter & Bever, 1988; Prasada & Pinker, 1993), and there has been a heated debate over the symbolic and connectionist modeling of the task. A challenge for a better symbolic model is posed by MacWhinney and Leinbach (1991), and is answered by us (Ling & Marinov, 1993). Past-tense acquisition has become a benchmark for the adequacy of cognitive modeling of language acquisition and processing.

The traditional grammatical theory proposes rule-like grammars for the regular inflection, and rote memory for the individual irregular inflection. Such theory, although explaining the phenomena of regular verbs quite well, fails for irregular verbs. Some irregular verbs (such as *go* $\mapsto$ *went*) are idiosyncratic, but other *so-called* irregular verbs form phonologically similar clusters, such as *sing* $\mapsto$ *sang*, *ring* $\mapsto$ *rang*, *spring* $\mapsto$ *sprang*, *swim* $\mapsto$ *swam*, or *grow* $\mapsto$ *grew*, *blow* $\mapsto$ *blew*, *throw* $\mapsto$ *threw*. Irregular inflection is *partially productive* depending on the similarity between the novel verb stem and the clusters of phonologically similar irregular verb stems. For example, because of the similar phonological patterns in *-ing* $\mapsto$ *-ang*, children occasionally use *brang* as the past tense of *bring*. In an experiment of predicting pseudo-verb past tenses by adults, the likelihood of irregular past-tense responses of pseudo-verbs increases with the phonological similarity from clusters of English irregular verbs. For example, subjects more willingly give irregular past-tense inflection *splung* from *spling*, less willingly, *shunk* from *shink*, and even less willingly, *sud* from *sid*. The model of past-tense acquisition should also extend irregular inflections to novel verbs, and such generalizations have to be plausible. In some theories of generative grammar, e.g., (Chomsky & Halle, 1968; Halle & Mohanan, 1985), irregular verbs are also governed by rules, such as (in the spelling form) "change *-ing* to *-ang*" as in *spring* $\mapsto$ *sprang*, which are applied to *a list* of irregular verbs. However, as correctly pointed out by Pinker (1991), Prasada and Pinker (1993), such a theory leaves the phonological similarity of verbs in the list unexplained, and thus is unable to assign a novel verb into an appropriate list for producing matching irregular past tense . If such rules are applied universally to all verbs, they fail to be sufficient conditions since they can apply to other verbs (such as *bring, fling*) incorrectly. One of the major difficulties with such theories is that, because of the complicated relations between irregular verb stems and their past tenses, the sufficient and necessary "rules" that serve for both clustering and for predicting past-tense inflections, if they exist, are hard to find.

The unsatisfactory modeling of irregular past tenses by traditional grammatical theories leads to the great

[1] Except for homophones and denominal verbs (Kim, Pinker, Prince, & Prasada, 1991).

impact of a radically different approach to the task — the connectionist modeling. In connectionist models (also often called artificial neural networks, or ANNs), generalization is produced by the similarity captured by a set of collectively contributed weights distributed widely between units. However, for all connectionist models built for the task, the unitary model is used for both regular and irregular generalization. For many theoretical reasons (Prasada & Pinker, 1993) and empirical findings (Ling, 1994), such models are not suitable for the regular past-tense inflection, which is best left to symbolic rule-learning systems. However, connectionist models do pose attractive properties as a possible model for irregular verbs. After all, connectionist models are associative in nature, and are supposed to be good at learning and generalizing mappings that require *many* complicated relations between inputs and outputs. Indeed, Pinker (1991), Prasada and Pinker (1993) propose a modified classical theory in which a rule-like module is used for the regular past-tense inflection, and an ANN-like associative memory for the irregular stem and past-tense mappings, fostering analogical generalization by similarity.

An All-rule Theory of the Past-tense Learning

We have constructed a general-purpose symbolic Pattern Associator (SPA for short) that can learn very efficiently from any set of given input/output pairs of patterns (Ling & Marinov, 1993). The SPA is an N-to-M (N input attributes to M output attribute) pattern associator which essentially applies the N-to-1 decision-tree classifier ID3 (Quinlan, 1986) M times. Given any set of patterns containing N input attributes and M output attributes, M decision trees are built by the SPA, each for one of the M output attributes. Those M trees will be used collectively to predict the output pattern of a new input pattern; each tree determines one attribute value in the output. For the verb past-tense learning, the input patterns are phoneme letters in the verb base with N as the maximum number of phoneme letters in the verb base, and the output patterns are phoneme letters in the past tense with M as the maximum number of phoneme letters in the past tense. The SPA actually utilizes an improved implementation of ID3 called c4.5 (Quinlan, 1993) with several modifications (for details of SPA, see (Ling & Marinov, 1993; Ling, 1994)). ID3 uses information gain ratio as a criterion for selecting attributes as roots of the decision subtrees. This leads to choosing the locally *most relevant* or *discriminating* attributes for the roots, and thus results in building *small* decision trees.

The decision trees built by SPA can be converted directly into a set of production rules. As an example, to determine the fourth phoneme of *sprang (spr&N)* from *spring (sprIN)*, the exception-handling rule from the actual decision tree is:

If $\mathbf{I}_2 = \mathrm{p}$ & $\mathbf{I}_4 = \mathrm{I}$ & $\mathbf{I}_5 = \mathrm{N}$ & $\mathbf{I}_6 = \$$ then $\mathbf{O}_4 = \&$

where $\mathbf{I}_k$ indicates the k^{th} input phoneme, and $\mathbf{O}_k$ the k^{th} output phoneme, The following is a typical suffix-adding rule:

If $\mathbf{I}_4 = \mathrm{n}$ and $\mathbf{I}_5 = \$$, then $\mathbf{O}_5 = \mathrm{d}$

The identity rules have only one condition, and can be expressed as:

If $\mathbf{I}_5 = \mathbf{X}$, then $\mathbf{O}_5 = \mathbf{X}$

where $\mathbf{X}$ can be any single phoneme.

However, these rules look odd since they are for position-specific phonemes in the past tense. The representation of the task is positioned phoneme strings, therefore, this limitation is rooted at the representation, and is thus intrinsic to any learning systems. However, because of the transparent representation of the SPA — one of the major advantages over connectionist models — we can represent rules in *verb-wise* format, and possibly apply further generalization on them. For example, the rules for each phoneme in the past tense *sprang* are:

```
SPA's           SPA's
Input           Output
s p r I N $ -> s p r & N $  (Unibet coding for
                             spring -> sprang)
1 2 3 4 5 6    1 2 3 4 5 6  (attribute positions)
s * * * * * -> s            (identity-mapping rule)
* p * * * * ->   p          (identity-mapping rule)
* * r * * * ->     r        (identity-mapping rule)
* p * I N $ ->       &      (exception-handling rule)
* * * * N $ ->         N    (identity-mapping rule)
* * * * N $ ->           $  (exception-handling rule)
```

Since exception-handling rules override identity-mapping rules, these rules can be represented verb-wise as an associative *template*:

$\mathbf{X}\,\mathrm{p}\,\mathbf{Y}\,\mathrm{I}\,\mathrm{N}\,\$ \mapsto \mathbf{X}\,\mathrm{p}\,\mathbf{Y}\,\&\,\mathrm{N}\,\$$.

Although this template may only be applicable to the mapping *spring* $\mapsto$ *sprang*, it is still more general than rote learning. It can be used to predict other 5-phoneme-letter verb stem with this pattern. For example, by using this template our SPA produces *splang (spl&N)* for a pseudo-verb *spling (splIN)*, which is a very plausible prediction. Therefore, the prediction of irregular inflections is not a mystery in SPA; we can account for each and every irregular past tense in this fashion.

Rules for regular past tenses with the same suffix patterns can be represented in the same, verb-wise fashion. For example,

$\mathbf{X}_1\ \mathbf{X}_2\ \mathbf{X}_3\ \mathrm{n} \mapsto \mathbf{X}_1\ \mathbf{X}_2\ \mathbf{X}_3\ \mathrm{n\ d}$

$\mathbf{X}_1\ \mathbf{X}_2\ \mathbf{X}_3\ \mathbf{X}_4\ \mathrm{n} \mapsto \mathbf{X}_1\ \mathbf{X}_2\ \mathbf{X}_3\ \mathbf{X}_4\ \mathrm{n\ d}$

Since all such rules are in a similar format except the indexes for the phoneme positions, an inductive learning step based on least general generalization (Plotkin, 1970) can be applied to produce *one* rule covering the two rules above. Such a rule has a variable that replaces individual indexes:

For all (integer) k, $\mathbf{X}_1 \ldots \mathbf{X}_k\ \mathrm{n} \mapsto \mathbf{X}_1 \ldots \mathbf{X}_k\ \mathrm{n\ d}$

This rule expresses the meaning that for any *verb stem* ending with phoneme n (i.e., $\mathbf{X}_1 \ldots \mathbf{X}_k$ n), its past tense (in phoneme string) is formed by suffixing d. This is a *first-order* rule applicable to all verb stems — something connectionist models fail to produce. Note, however, that the above inductive step cannot compress or generalize most templates for the irregular verbs. Therefore, these templates will remain in the position-specific form.

The fact that the SPA can produce symbolic rules for both *distinguishing* and *predicting* regular and irregular verbs is very interesting. After all, irregular verbs can be regarded as "regular" with rules applicable to only a few verbs. However, rules for regular and irregular verbs are highly *heterogeneous* in their format. The clusters of rules for irregular verbs form positioned templates. These templates have some flavor of the associative memory proposed by Pinker (1991), Prasada and Pinker (1993), since they represent the common pattern within clusters of phonologically similar irregular verbs. On the other hand, the rules for regular verbs produced by the post inductive step are first-order rules that apply to strings of arbitrary length. Therefore, our all-rule system is very similar in spirit to Pinker (1991), Prasada and Pinker (1993)'s modular hypothesis with rules for regulars and the associative memory for irregulars. The associative memory is implemented in symbolic rule-like templates that capture the similarity among verb patterns, and allow for analogical generalization for novel verbs.

Irregular Past-tense Generalization

As we have shown, rules for regulars are very simple since they have a small number of conditions. It is the simplicity of the regular-inflection rules that makes SPA, which actually searches for small rules, a better learning model for regular verbs than connectionist models. In particular, the generalization of the SPA's rules to first-order rules for regular verbs makes the regular past-tense inflection insensible to the verb frequency and similarity of other regular verbs in the training set. Therefore, the properties of regular-verb acquisition can be explained very well in the SPA. However, as we have seen, rules (or templates) for irregular clusters are quite complicated. Indeed, some of them are too complicated to be found by humans. These rules utilize many input phonemes in the verb stem to produce the output phonemes in the past tense. Therefore, learning irregular past-tense inflections is *the* task that connectionist models may score better.

The existence of the SPA at the present time makes it possible to answer the following interesting question: Which one, the SPA or the connectionist model, would make more reasonable *irregular* past-tense generalization?

To study irregular past-tense generalization of a model, we can train the model on irregular verbs only. Such a model will not predict past tenses in regular suffix-adding form, and irregular past-tense generalization can be studied in isolation. One might think that we can train two models on a set of irregular verbs and test them on another set of irregular verbs not shown in the training set, and see which model makes more correct predictions in the testing set. This method is not quite appropriate, however, because some irregular verbs (such as *go* ↦ *went*, *be* ↦ *was*, *come* ↦ *came*) are idiosyncratic since their past tenses are beyond any reasonable prediction by native English speakers *if* they had *never* heard of them. How do we know which irregular past tenses are (un)predictable from a *given* set of training sample? Further, we *know* that the past tense of *spring* is *sprang*, but is *sprung* a reasonable prediction as well? Which one is more reasonable when judged by native English speakers? The method of splitting irregular verbs into disjoint training and testing sets requires us to to find native English speakers, who have never heard of the past tense of those irregular verbs in the testing set, to judge the plausibility of predictions from the SPA and the connectionist model. However, such a set of jurors is hard to come by (without a lifelong public ban on using these verbs!) since most irregular verbs are the ones most frequently used in everyday life.

To solve this tricky problem, we turn ourselves to a set of *pseudo-verbs* that presumably have never been heard by native English speakers. Prasada and Pinker (1993) create 30 irregular-sounding pseudo-verbs. These irregular pseudo-verbs are derived directly from clusters of English irregular verbs involving four typical vowel transformations (in UNIBET form) — [IN ↦ 6N], [i ↦ E], [o ↦ u], and [er ↦ or] — or by making additional initial or final consonant changes. Therefore, if we train both the SPA and the connectionist model on irregular verbs, and let both models predict their past tenses, we can obtain a set of pseudo-verbs whose past-tense inflections are *different* from the rival models. We can then ask native English speakers to judge which models make more reasonable predictions.

This pseudo-verb testing paradigm allows for a direct and psychologically relevant comparison between the "pure" intuitions of native speakers about English past tenses and the predictions of computational learning models on verbs that are *truly novel* for both the humans and the models.

Simulation

Our list of English irregular verbs comes from MacWhinney and Leinbach (1991). Originally the list contains 150 irregular verb pairs, each with a Kucera-Francis frequency (Kucera & Francis, 1967). To remove noise in the training set, one of the homophones (such as *hang* ↦ *hung*, *hang* ↦ *hanged*) with lower frequency is removed from the training set. Thus, we obtain a list of 140 distinctive stem and past-tense pairs; all of them are used in the training set to maximize the overlap of irregular verbs of native English speakers. Multiple copies of verbs reflecting the Kucera-Francis frequency of the verbs are used, resulting in a total of 862 verb (tokens) in the training set. Both SPA and ANN are trained on exactly the same training set (with 862 verb tokens) in the templated format. The trained models are then used to predict the past tenses of 30 pseudo verbs.

Learning is based upon the phonological information. A UNIBET (MacWhinney, 1990) representation is used in which different phonemes are represented by different alphabetic and numerical letters. There is a total of 36 phonemes. The stem and the past tense of each verb is actually represented in a left-justified template in the format of

CCCVVCCCVVCCCVVCCC

where C stands for consonant and V for vowel space holders. This representation is used in (MacWhinney & Leinbach, 1991), and a monosyllable one (CCCVVCCC) is used in (Daugherty & Seidenberg, 1993).

The SPA takes symbolic values, and therefore, the training set is simply in the templated UNIBET phoneme strings. On the other hand, ANNs have to encode multiple values of attributes in binary bits. A specific distributed representation used in (MacWhinney & Leinbach, 1991; MacWhinney, 1993) — a set of (binary) phonetic features — is used to encode all phoneme letters for the ANN. Each vowel (V in the above templates) is encoded by 8 phonetic features (front, centre, back, high, low, middle, round, and diphthong) and each consonant (C in the above templates) by 10 phonetic features (voiced, labial, dental, palatal, velar, nasal, liquid, trill, fricative and interdental). Therefore, both SPA and ANN are trained on exactly the same training set (with 862 verb tokens) in the templated format. The *only* difference is that the SPA is trained and tested with phoneme letters directly, i.e., 18 decision trees were built for each of the 18 phoneme letters in the output templates; while ANN uses the distributed representation designed by MacWhinney and Leinbach (1991).

The artificial neural network (ANN) program we used is a package called Xerion developed at the University of Toronto. It has several sophisticated search mechanisms such as the conjugate gradient with line search. We found that training with the conjugate gradient method is much faster than with the standard backpropagation algorithm with momentum. Using the conjugate gradient method also avoids the need to search for proper settings of parameters such as the learning rate. We do need to determine the proper number of hidden units, however, because the training set contains only 140 distinctive verbs, cross-validation cannot be used. On the other hand, since the training set does not have noise (contradictory output patterns with the same input patterns), we would expect that a good ANN model should be able at least to be trained to reproduce well most of the past tenses in the training set. Therefore, the number of mistakes remained in the training set is a primary criterion for choosing the network architecture. We experiment with numbers of hidden units from 60 to 180, and also vary the random seeds. It turns out that when the number of hidden units is between 70 and 90, we obtain very small training errors (less than 2% at the verb level). After we examine the predictions of the pseudo-verbs from various networks with small errors, we find that about half of the predictions are identical. Each such run ends up with a similar number of ill-formed and reasonable generalizations (see next subsection), so we choose the results of one such network and report them here. The SPA, on the other hand, has no parameter to adjust, and its behavior is deterministic. Training with the SPA always ends up with zero error, producing the same decision trees each time, and thus, the prediction on pseudo-verbs is always the same. Training SPA takes less than 1 minute on the SunSpark Model 10, on which it takes about 15 hours to train ANN to 300 epochs.

Simulation Results

Out of 30 pseudo-irregular verbs, SPA and ANN give 4 identical past-tense inflections. See Table 1. (Note that the spelling representation in this table and in other consequent tables are suggestive and for readability only.) Since the past-tense inflections are the same from both models, they are not used in the psychological experiment.

Verb Stem		Past Tense	
spelling	phonetic	spelling	phonetic
froe	fro	frew	fru
plear	plEr	plore	plor
nist	nIst	nist	nIst
meep	mip	mept	mEpt

Table 1: Equivalent past-tense predictions from SPA and ANN

Of the remaining 26 pseudo-irregular verbs, 15 have weird past-tense inflections due to *ill-formed* generalization by SPA or ANN. These inflections are deemed to be ill-formed because, when judge by two native English speakers (one fourth-year student and one psychology professor), they seem to be implausible or dissimilar (in any obvious way) to any known English irregular verbs. The ill-formed generalization includes implausible phoneme drops (such as *frink* ↦ *frun*) and implausible phoneme change (such as *preed* ↦ *brod*, *grear* ↦ *grud*, *cloe* ↦ *hli*). Among these 15, only 1 (*frink* ↦ *frun*) is produced by SPA. The rest of the 14 ill-formed generalizations are produced by ANN, and SPA's generalizations of those 14 pseudo-verbs are quite plausible except, perhaps, *goav* ↦ *wov*, which is clearly affected by *go* ↦ *went*. Therefore, SPA makes *far less* ill-formed generalization than ANN does. The problems are likely caused by the representation of the data, as well as by the biases of the learning algorithm. Since ill-formed generalizations are deemed to be implausible, most of them are not used in the psychological experiment.

For the other 11 past-tense inflections, SPA and ANN produce *plausible* and *different* past tenses. See Table 2. These are the most interesting ones since we can now ask native English speakers to judge which inflection, from SPA or ANN, is more plausible. Upon close examination of the 11 pseudo-verbs, two have inflections that have the same pronunciation as some known English words, or as the regular past tense. These two are marked by * in the table: *few* is a common English word and *blipt* sounds like a regular past tense of blip. We suspect that the past-tense inflections of these two pseudo-verbs would bias subjects, and thus, they are removed from the psychological experiment. This ends up with only 9 pseudo-verbs, which seems a little too few. We thus correct obvious phoneme drops or changes in the ill-formed past-tense inflections , and add them (marked by $^{+}$) into Table 2. Thus, We end up with 12 (9+3) pseudo-verbs which have plausible and different past-tense inflections, and which are used in the psychological experiment described in the next section.

Experiment

The past-tense acquisition is based on phonological information. Experiments with the same set of pseudo-verbs

Verb Stem		Past Tense: SPA		Past Tense: ANN	
spelling	phonetic	spelling	phonetic	spelling	phonetic
ning	nIN	ning	nIN	neng	nEN
kwair	kwEr	kwor	kwor	kwer	kw6r
cleef	klif	clef	klEf	claft	kleft
fring	frIN	frung	fr6N	frong	frON
queef	kwif	quef	kwEf	quaft	kweft
flape	flep	flept	flEpt	fleb	flEb
blafe	blef	blofe	blof	blav	blev
preek	prik	prek	prEk	prok	prok
cleed	klid	cled	klEd	clud	kl6d
*fow	fo	few	fu	faw	fO
*blip	blIp	blipt	blIpt	blip	blIp
+frink	frINk	frunk	fr6Nk	frank	fr&Nk
+spling	splIN	splang	spl&N	splung	spl6N
+preed	prid	pred	prEd	prod	prod

Table 2: Plausible but different generalizations from SPA and ANN

* Verbs with plausible generalizations but withdrawn from the experiment.

+ Verbs with ill-formed generalizations but corrected and used in the experiment (see text).

were conducted previously (Prasada & Pinker, 1993) in the written version. We suspect that, although subjects were instructed to ignore the spelling and make their judgement solely based on the pronunciation, spelling might still have a certain effect on subjects' response. Therefore, we decide to conduct our experiment in audio format. For each pseudo-verb, two sentences (A and B) are constructed using two different past-tense inflections, one from SPA and one from ANN. All sentences have the same format:

I don't like to **X**, *but yesterday I* **Y**.

where **Y** is the past-tense inflection of **X**. A female native English speaker who has a very clear voice is invited to read these sentences and is recorded on cassette tape.

The instruction for the experiment, which is also printed on the answer sheet, was read to the subjects slowly and clearly. The instruction asks the subjects to use their "gut feeling" and "first reaction" to judge the relative naturalness of irregular past tenses in the sentence A and B. On the answer sheet, scales are given for the subjects to rank the *relative* naturalness of two past-tense inflections from the two rival models. Each verb has a scale looks like the following:

A • • • • • • • **B**

At the end of the answer sheet, subjects are asked to circle if or not English is their first language. The answer sheet returned from non-native English speakers are not counted.

A small scale experiment with 5 graduate students was conducted first, and a discussion was held afterwards. The subjects reported that they could hear the recording on the tape, and the differences in sentences A and B quite well. The actual subjects used were first-year computer science students in their first semester who were taking "Computer Science Fundamentals I". They had no knowledge of the experiment, and were happily surprised when we walked into the lecture room with audio gear. After a brief explanation, they showed great enthusiasm, and cooperated very well throughout the experiment. A total of 42 students happened to be present. In the end, 34 answer sheets were returned by native English speakers, whose results are summarized in the next subsection.

Results of Comparison between SPA and ANN again Native English Speakers

For the scale on the answer sheet (**A** • • • • • • • **B**), We assign 1 to 7 to the black dots starting from A. Thus, 4 represents the score that A and B are equally natural. Then we calculate the mean and standard error of answers for each pseudo-verb across all of the 34 answer sheets. If the mean is not equal to 4, a one-tailed, paired t-test is calculated for the probability that the mean is greater (or less) than 4. If the probability is less than 0.90, we deem the difference of the mean and 4 as unreliable, and therefore, there is no significant difference between the two choices. Table 3 lists 8 pseudo-verbs and their past-tense inflections whose naturalness rating has a reliability probability greater or equal to 0.90. For the remaining 4 pseudo-verbs, some have very large variations, and some have mean values very close to 4, indicating that the subjects do not seem to have consistent preferences of one over the other.

From Table 3, we clearly do not see any advantage of the connectionist model over the SPA; both models are judged almost equally reasonable. In fact, SPA even scores slightly better. Therefore, there seems no reason to believe that connectionist systems are better models for irregular past-tense generalization, a task that they are alleged to do well. If we include the ill-formed generalization in the score, connectionist models would be deemed to generalize significantly worse than the SPA.

Verb Stem		Past Tense: SPA		Past Tense: ANN		Which is	Score diff.	Prob. in
spelling	phonetic	spelling	phonetic	spelling	phonetic	better	|SPA−ANN|	t-test
ning	nIN	ning	nIN	neng	nEN	ANN	1.23	0.99
kwair	kwEr	kwor	kwor	kwer	kw6r	SPA	2.35	0.99
queef	kwif	quef	kwEf	quaft	kweft	ANN	0.62	0.90
flape	flep	flept	flEpt	fleb	flEb	SPA	2.12	0.99
blafe	blef	blofe	blof	blev	blev	SPA	0.59	0.95
preek	prik	prek	prEk	prok	prok	ANN	1.44	0.99
cleed	klid	cled	klEd	clud	kl6d	SPA	2.26	0.99
preed	prid	pred	prEd	prod	prod	SPA	0.53	0.90

Table 3: Subjects' Rating of the reasonable generalizations from SPA and ANN

Conclusion

We present a general-purpose symbolic pattern associator, and show that it produces symbolic rules for both regular and irregular verbs. Our all-rule system even makes more plausible irregular past-tense generalization than the connectionist simulation, which, at first sight, seems to have advantage over rule-learning systems. Our results support the view that many such high-level, rule-governed cognitive tasks should be better modeled by symbolic, rather than connectionist, systems.

Acknowledgements

I gratefully thank Ken McRae (a psychology professor at UWO) for many detailed discussions in the design of the psychological experiment. I also thank Bernadette Daly for her help in recording the audio tape used in the experiment. Many discussions with Steve Pinker, Dave Touretzky, and Brian MacWhinney have been very helpful as well.

Reference

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper a Row: New York.

Daugherty, K., & Seidenberg, M. (1993). Beyond rules and exceptions: A connectionist modeling approach to inflectional morphology. In Lima, S. (Ed.), *The Reality of Linguistic Rules*. John Benjamins.

Halle, M., & Mohanan, K. (1985). Segmental phonology of modern English. *Linguistic Inquiry, 16*, 57 – 116.

Kim, J., Pinker, S., Prince, A., & Prasada, S. (1991). Why no mere mortal has ever flown out to center field. *Cognitive Science, 15*, 173 – 218.

Kucera, H., & Francis, W. (1967). *Computational Analysis of Present-Day American English*. Provedance, RI: Brown University Press.

Lachter, J., & Bever, T. (1988). The relation between linguistic structure and associative theories of language learning – a constructive critique of some connectionist learning models. In Pinker, S., & Mehler, J. (Eds.), *Connections and Symbols*, pp. 195 – 247. Cambridge, MA: MIT Press.

Ling, X. (1994). Learning the past tense of English verbs: the Symbolic Pattern Associator vs. connectionist models. *Journal of Artificial Intelligence Research, 1*, 209–229.

Ling, X., & Marinov, M. (1993). Answering the connectionist challenge: a symbolic model of learning the past tense of English verbs. *Cognition, 49*(3), 235–290.

MacWhinney, B. (1990). *The CHILDES Project: Tools for Analyzing Talk*. Hillsdale, NJ: Erlbaum.

MacWhinney, B. (1993). Connections and symbols: closing the gap. *Cognition, 49*(3), 291–296.

MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb model. *Cognition, 40*, 121 – 157.

Pinker, S. (1991). Rules of language. *Science, 253*, 530 – 535.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. In Pinker, S., & Mehler, J. (Eds.), *Connections and Symbols*, pp. 73 – 193. Cambridge, MA: MIT Press.

Plotkin, G. (1970). A note on inductive generalization. In Meltzer, B., & Michie, D. (Eds.), *Machine Intelligence 5*, pp. 153–163. Elsevier North-Holland.

Prasada, S., & Pinker, S. (1993). Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes, 8*(1), 1 – 56.

Quinlan, J. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81 – 106.

Quinlan, J. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann: San Mateo, CA.

Rumelhart, D., & McClelland, J. (1986). On learning the past tenses of English verbs. In Rumelhart, D., McClelland, J., & the PDP Research Group (Eds.), *Parallel Distributed Processing, Vol. 2*, pp. 216 – 271. Cambridge, MA: MIT Press.