What a Perceptron Reveals about Metrical Phonology

Prahlad Gupta
Department of Philosophy
Carnegie Mellon University
Pittsburgh, PA 15213
prahlad@cs.cmu.edu

David S. Touretzky*
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213
dst@cs.cmu.edu

Abstract

Metrical phonology is a relatively successful theory that attempts to explain stress systems in language. This paper discusses a perceptron model of stress, pointing out interesting parallels between certain aspects of the model and the constructs and predictions of metrical theory. The distribution of learning times obtained from perceptron experiments corresponds with theoretical predictions of "markedness." In addition, the weight patterns developed by perceptron learning bear a suggestive relationship to features of the linguistic analysis, particularly with regard to iteration and metrical feet. Our results suggest that simple statistical learning techniques have the potential to complement, and provide computational validation for, abstract theoretical investigations of linguistic domains.

Basics of metrical theory

This section outlines the structure of the syllable and the basics of metrical theory[1]. A syllable is composed of an *onset*, which contains the material before the vowel, and a *rime*. The rime is composed of a *nucleus*, which contains the vocalic material, and a *coda*, which contains any remaining non-vocalic material.

A syllable may be *open* (ends in a vowel) or *closed* (ends in a consonant). Viewing syllable structures as trees, an open syllable has a *non-branching* rime (the rime has a nucleus, but not a coda), while a closed syllable has a *branching* rime (the rime has both a nucleus and a coda). A syllable may also have a *long* vowel, in which case the nucleus is considered to branch.

In many languages, stress tends to be placed on certain kinds of syllables rather than on others; the former are termed *heavy* syllables, and the latter *light* syllables. What counts as heavy or light differs across languages but, most commonly, a heavy syllable is one with a branching rime ([Goldsmith 1990, p. 113]). However, it is possible for other properties to contribute to syllable weight. For example, in some languages, only syllables with a long vowel (i.e., branching nucleus) count as heavy. Closed syllables with short vowels do not count as heavy, as they would in the more commonly-occurring heavy/light distinction ([Goldsmith 1990, p. 179]).

Languages that distinguish between heavy and light syllables are termed *quantity-sensitive*, while languages that do not make this distinction are termed *quantity-insensitive*. There seems to be theoretical agreement that, in quantity-sensitive systems, the placement of stress is sensitive to the structure of the rime, but not the onset. We follow this assumption[2]. Thus rime structure is taken to be the basic level at which accounts of stress systems are formulated. Stress patterns are controlled by metrical structures built on top of rime structures. The version of metrical structure adopted here is *metrical feet*. We assume the *parameters* formulated by Dresher & Kaye ([Dresher & Kaye 1990, p. 142]):

(P1) The word-tree is strong on the [Left/Right]
(P2) Feet are [Binary/Unbounded]
(P3) Feet are built from the [Left/Right]
(P4) Feet are strong on the [Left/Right]
(P5) Feet are Quantity-Sensitive (QS) [Yes/No]
(P6) Feet are QS to the [Rime/Nucleus]
(P7) A strong branch of a foot must itself branch [No/Yes]
(P8) There is an extra-metrical syllable [Yes/No]
(P9) It is extra-metrical on the [Left/Right]
(P10) A weak foot is defooted in clash [No/Yes]
(P11) Feet are non-iterative [No/Yes]

As an example of the application of these parameters, consider the stress pattern of Maranungku, in which primary stress falls on the first syllable of the word and secondary stress on alternate succeeding syllables. Figure 1 shows an abstract representation of a six-syllable word, with each syllable represented as σ. The assignment of stress is characterized as follows. Binary, quantity-insensitive, left-dominant feet are constructed iteratively from the left edge of the word. Each foot has a "strong" and a "weak" branch (labeled "S" and "W," respectively, in the figure). The strong, or dominant branch assigns stress to the syllable it dominates. Since the feet are left-dominant, odd-numbered syllables are assigned stress. Over the roots of these *metrical feet*, a left-dominant *word-tree* is constructed, which assigns stress to the structure dominated by its leftmost branch. The third and fifth syllables are each dominated by the dominant branch of one metrical structure (a foot), while the first syllable is dominated by the dominant branches of two structures (a foot, and the word-tree). Even-numbered syllables are dominated only by non-dominant branches of feet. The result is that even-numbered syllables receive no stress; the third and fifth syllables receive one degree of stress (secondary stress); and the first syllable receives two degrees of stress (primary stress.) The parameter settings characterizing Maranungku are: [P1 Left], [P2 Binary], [P3 Left], [P4 Left], [P5 No], [P7 No], [P8 No], [P10 No], [P11 No]. Parameters P6 and P9 are irrelevant because of the settings of parameters P5 and P8, respectively.

Sixteen stress systems

Six quantity-insensitive (QI) languages and ten quantity-sensitive (QS) languages were examined in our experiments. The data, summarized in Table 1, were taken primarily from [Hayes 1980]. Note that the QI stress patterns of Latvian & French, Maranungku & Weri, and Lakota & Polish are mirror images of each other. The QS stress patterns of Malayalam &

*The second author was supported by a grant from Hughes Aircraft Corporation, and by the Office of Naval Research under contract number N00014-86-K-0678.

[1]For an overview of metrical phonology, see [Goldsmith 1990, chapter 4], [Kaye 1989, pp. 139-145], [van der Hulst & Smith 1982] or [Dresher & Kaye 1990, pp. 140-147].

[2]See, for example, ([Dresher & Kaye 1990, p. 141] or [Goldsmith 1990, p. 170]). However, both [Davis 1988] and [Everett & Everett 1984] present evidence that onsets may in fact be relevant to the placement of stress.

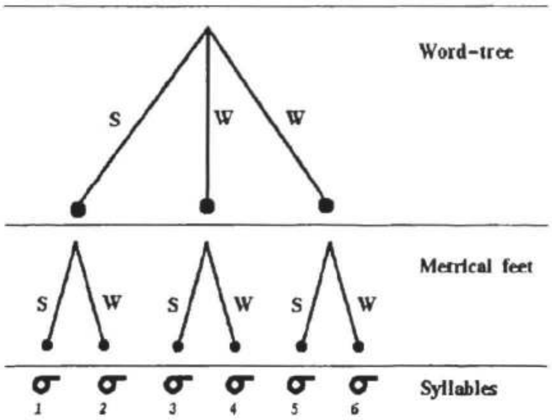

Figure 1: Metrical structures for a six-syllable word in Maranungku.

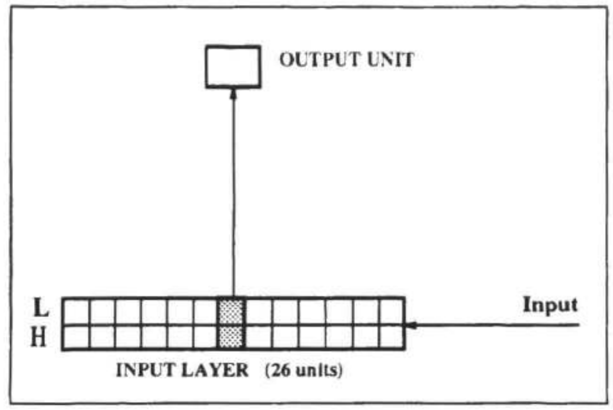

Figure 2: Perceptron model used in simulations.

Yapese, Ossetic & Rotuman, Eastern Permyak Komi & Eastern Cheremis, and Khalka Mongolian & Aguacatec Mayan are also mirror images[3].

Learning stress with a perceptron

In separate experiments, we taught a perceptron to produce the stress pattern of each of the sixteen languages. The domain was limited to single words, as in the previous learning models of metrical phonology developed by Dresher & Kaye, and Nyberg ([Dresher & Kaye 1990, Nyberg 1989]). Again as in the other models, the effects of morpho-syntactic information (such as lexical category) were ignored, and the simplifying assumption was made that the only relevant information about syllables was their weight.

Words up to 7 syllables long were slid across a 13-element input window as shown in Figure 2. At each time step, the perceptron was trained to predict the stress of the center element. In order to distinguish heavy from light syllables, each input element consisted of two units. Thus the perceptron had a total of 26 inputs, plus a bias connection. The input patterns used were [1 0] for a heavy syllable, [0 1] for a light syllable, and [0 0] for no syllable.[4] The output targets used in training were 1.0 for primary stress, 0.5 for secondary stress, and 0 for no stress.

The input data set for all stress systems consisted of all 255 word-forms of up to seven syllables. Each word was processed one syllable at a time. Connection weights were adjusted at each time step using the back-propagation learning algorithm of [Rumelhart, Hinton & Williams 1986]. One *epoch* consisted of one presentation of the entire training set. The network was trained for as many epochs as necessary to ensure that the stress value produced by the perceptron was within 0.1 of the target value, for each syllable of the word, for all words in the training set. A *learning rate* of 0.05 and *momentum* of 0.90 was used in all simulations. Initial weights were uniformly distributed random values in the range ± 0.5. Each simulation was run at least three times, and the learning times averaged. Results are shown in Table 2.

Markedness, learnability, and simulations

A universal grammar (UG) of stress should incorporate a theory of markedness, so as to predict which features of stress systems are at the core of the human language faculty and which are at the periphery. The distributional approach to markedness treats as "unmarked" those linguistic forms that occur more frequently in the world's languages. This seems to be the approach taken by, for example, Hayes[5].

Such an approach can be criticized, however, on the grounds that the frequency of occurrence of some linguistic form does not necessarily determine its status as "core" or "peripheral", and the non-occurrence of some form does not show that it is "impossible." The distribution of languages in the world is a function of all kinds of historical, non-linguistic, factors, and does not necessarily have linguistic-theoretic significance[6].

Another approach to markedness is *learnability theory*, which examines the logical process of language acquisition[7]. Thus, for example, Dresher & Kaye take iteration to be the default or unmarked setting for parameter P11, because there is evidence that can cause revision of this default if it turns out to be the incorrect setting: the absence of any secondary stresses serves as a diagnostic that feet are *not* iterative ([Dresher & Kaye 1990, p. 191]). If non-iteration were the default, their learning system might not encounter evidence that would enable it to correct this default setting, if it were in fact incorrect. It should be noted that, while this is a representative application of subset theory, the choice of default parameter values depends on the particular learning algorithm employed.

[3]We have somewhat simplified the descriptions of Polish and Malayalam compared with those in [Halle & Vergnaud 1987, pp. 57-58] and [Hayes 1980, p. 66, 109]. However, this does not detract from our discussion in any way, as stress systems corresponding to our simplifications are reported to exist: Swahili ([Halle & Clements 1983, p.17]) and Gurkhali ([Hayes 1980, p.66]), corresponding to Polish and Malayalam, respectively.

[4]Note that the distinction between heavy and light syllables is irrelevant in a quantity-insensitive language. However, this representation was used in all simulations to maintain consistency across quantity-sensitive and quantity-insensitive systems.

[5]([Hayes 1980, p. 50]): "In justifying a foot inventory as the unmarked one, a minimal requirement is to show that all the members of the inventory are attested in a fair number of languages ..."

[6]To quote Pullum ([Pullum 1982, p. 343; p. 340]): "... no one has any idea to what extent the history of the human race has skewed the distribution of [linguistic] types by skewing the distribution of people ... to postulate a default assumption that, say, *wh*-movement cannot be rightward, merely because it is commoner (in currently well-studied languages) for it to be leftward, is surely perverse as well as unnecessary. Language acquisition takes place within the infant, not within the context of a statistical survey of currently attested languages ..."

[7]As an example, the Subset Principle ([Berwick 1985], [Wexler & Manzini 1987]) has implications for markedness. Suppose that two possible settings a and b for parameter P result in the learner respectively accepting sets S_a and S_b of linguistic forms. If S_a is a subset of S_b, then, once P has been set to value b, it will never get re-set to a, even if that was the correct setting. Unmarked values for parameters should therefore be the ones yielding the most constrained system.

LANGUAGE	DESCRIPTION OF STRESS PATTERN	EXAMPLES
Latvian	Fixed word-initial stress.	$S^1S^0S^0S^0S^0S^0S^0$
French	Fixed word-final stress.	$S^0S^0S^0S^0S^0S^0S^1$
Maranungku	Primary stress on first syllable, secondary stress on alternate succeeding syllables.	$S^1S^0S^2S^0S^2S^0S^2$
Weri	Primary stress on last syllable, secondary stress on alternate preceding syllables.	$S^2S^0S^2S^0S^2S^0S^1$
Lakota	Primary stress on second syllable.	$S^0S^1S^0S^0S^0S^0S^0$
Polish	Primary stress on penultimate syllable.	$S^0S^0S^0S^0S^0S^1S^0$
Koya	Primary stress on first syllable, secondary stress on heavy syllables. (Heavy = closed syllable or syllable with long vowel.)	$L^1L^0L^0H^2L^0L^0L^0$ $L^1L^0L^0L^0L^0L^0L^0$
Eskimo	(Primary) stress on final and heavy syllables. (Heavy = closed syllable.)	$L^0L^0L^0H^1L^0L^0L^1$ $L^0L^0L^0L^0L^0L^0L^1$
Malayalam	Primary stress on first syllable except when first syllable light and second syllable heavy. (Heavy = long vowel.)	$L^1L^0L^0H^0L^0L^0L^0$ $L^0H^1L^0H^0L^0L^0L^0$
Yapese	Primary stress on last syllable except when last is light and penultimate heavy. (Heavy = long vowel.)	$L^0L^0L^0H^0L^0L^0L^1$ $L^0H^0L^0H^0L^0H^1L^0$
Ossetic	Primary stress on first syllable if heavy, else on second syllable. (Heavy = long vowel.)	$H^1L^0L^0H^0L^0L^0L^0$ $L^0L^1L^0L^0L^0L^0L^0$
Rotuman	Primary stress on last syllable if heavy, else on penultimate syllable. (Heavy = long vowel.)	$L^0L^0L^0H^0L^0L^0H^1$ $L^0L^0L^0L^0L^0L^1L^0$
Komi	Primary stress on first heavy syllable, or on last syllable if none heavy. (Heavy = long vowel.)	$L^0L^0H^1L^0L^0H^0L^0$ $L^0L^0L^0L^0L^0L^0L^1$
Cheremis	Primary stress on last heavy syllable, or on first syllable if none heavy. (Heavy = long vowel.)	$L^0L^0H^0L^0L^0H^1L^0$ $L^1L^0L^0L^0L^0L^0L^0$
Mongolian	Primary stress on first heavy syllable, or on first syllable if none heavy. (Heavy = long vowel.)	$L^0L^0H^1L^0L^0H^0L^0$ $L^1L^0L^0L^0L^0L^0L^0$
Mayan	Primary stress on last heavy syllable, or on last syllable if none heavy. (Heavy = long vowel.)	$L^0L^0H^0L^0L^0H^1L^0$ $L^0L^0L^0L^0L^0L^0L^1$

Table 1: Stress patterns: description and example stress assignment. Examples are of stress assignment in seven-syllable words. Primary stress is denoted by the superscript 1 (e.g., S^1), secondary stress by the superscript 2, and no stress by the superscript 0. "S" indicates an arbitrary syllable, and is used for the QI stress patterns. For QS stress patterns, "H" and "L" are used to denote Heavy and Light syllables, respectively.

If learnability arguments should propose the setting x for parameter P, then some explanation would be needed if 95 per cent of the world's languages could be analyzed as having the setting y for the same parameter. Although distributional observations may not be an appropriate starting point for theory construction, they do provide a set of additional data points. However, given the previously noted criticisms, it appears they can only provide a weak constraint on metrical theories. Some other source of evidence would be valuable.

It is therefore interesting to note that the simulations described in this paper do provide "learnability" results for a variety of stress patterns. By extension, they make predictions about the learnability of various linguistic forms in metrical phonology. We claim that these results provide a source of data that can complement theoretical investigations.

Table 2 shows the stress systems grouped by their theoretical analyses in terms of the parameter scheme discussed in the first section. The last column of the table shows the average learning time in epochs for each group of stress patterns. As can be seen, there appears to be a fairly systematic differentiation of learning times for groups of stress patterns with different clusters of parameter settings.

First of all, learning times appear to be significantly higher for stress systems in groups 5 through 9, which have non-iterative feet, than for those in groups 1 through 4, which either do not have metrical feet at all, or else have iterative feet. This makes the interesting prediction that non-iterative feet are more difficult to learn, and hence marked. This prediction corresponds with both Halle & Vergnaud's Exhaustivity Condition[8], and with the choice of marked and unmarked settings in Dresher & Kaye's parameter scheme (Parameter P11)[9].

Comparison of learning times for group 1 vis-a-vis groups 2, 3 and 4 suggests that a stress system with only a word-tree (i.e., with no metrical feet) is easier to learn than one with (iterative) metrical feet.

The dramatic difference in learning times between groups 8 and 9 suggests that it is marked for the dominant node to be obligatorily branching[10]. Group 8 differs from group 9 only in not having obligatory branching, and average learning times were 218 epochs vs. 2308 epochs.

This prediction agrees with the distributional view that obligatory branching is relatively marked[11]. However, it runs counter to Dresher & Kaye's choice of default values (parameter P7)[12].

However, comparison of group 6 with group 7 suggests that systems with obligatory branching may be more *easily* learned: group 6, with obligatory branching, has a learning time of 19

[8] "The rules of constituent boundary construction apply exhaustively ..." ([Halle & Vergnaud 1987, p. 15]).

[9] In Dresher & Kaye's model, iteration is the default or unmarked parameter setting, because there is evidence that can cause revision of this default. The absence of any secondary stresses serves as a diagnostic that feet are *not* iterative ([Dresher & Kaye 1990, p. 191]).

[10] This means that the strong branch of a foot must dominate a heavy syllable, and cannot dominate a light one.

[11] [Hayes 1980, p. 113]: "... the maximally unmarked labeling convention is that which makes all dominant nodes strong ... the convention that wins second place is: label dominant nodes as strong if and only if they branch ..."

[12] Obligatory branching is the default because evidence (the presence of any stressed light syllables that do not receive stress from the word-tree) can force its revision ([Dresher & Kaye 1990, p. 193]).

	LANGUAGE	CHARACTERIZATION	EPOCHS
1	Latvian, French	Word-tree, no feet	1
2	Maranungku, Weri	Word-tree, iterative binary QI feet	2
3	Koya	Word-tree, iterative unbounded QS feet	2
4	Eskimo	No word-tree, iterative unbounded QS feet	3
5	Lakota, Polish	Word-tree, non-iterative binary QI feet	9
6	Malayalam, Yapese	Word-tree, non-iterative binary QS feet, dominant node branches	19 (±1)
7	Ossetic, Rotuman	Word-tree, non-iterative binary QS feet	29 (±1)
8	Komi, Cheremis	Word-tree, non-iterative unbounded QS feet	218 (±1)
9	Mongolian, Mayan	Word-tree, non-iterative unbounded QS feet, dominant node branches	2308 (±4)

Table 2: Learning times for QI and QS stress patterns grouped by theoretical analysis. Each figure is the average learning time for languages in the group.

epochs, compared with group 7, without obligatory branching, but with a learning time of 29 epochs. This runs counter to the distributional argument, but agrees with the learnability view.

Two points are worth noting. First, it is interesting that where there is a conflict between the distributional and learnability theory predictions of markedness, there is also conflicting evidence from the perceptron simulation. Second, these conflicting perceptron results highlight the fact that it may be infeasible to analyze the effects of different settings for *individual* parameters; it may only be possible to make broader analyses of the effects of *clusters* of parameter settings. Strong interactions between parameters have also been observed in other computational learning models of metrical phonology (Eric Nyberg, personal communication).

However, in view of the greater differential in learning times between Groups 8 and 9 than between Groups 6 and 7, we conclude that the effect of obligatory branching is to *increase* learning time. That is, we view our learning results as supporting the markedness of obligatory branching. This raises the interesting possibility that learning results such as those from the present perceptron simulations can provide a new source of insight into questions of markedness. As previously noted, there is controversy over the relevance of distributional facts to theories of markedness. Moreover, the distributional view of the markedness of obligatory branching ([Hayes 1980, p. 113]) seems to conflict with the learnability view ([Dresher & Kaye 1990, p. 193]). The present simulations seem to agree with the distributional view, and, we suggest, provide a potential means of validation for theoretical analyses.

This contribution to theoretical analysis can be further illustrated for the stress systems of Lakota and Polish, which are mirror images. The analysis so far adopted for Lakota is that it has non-iterative binary right-dominant QI feet constructed from left to right, with a left-dominant word-tree[13]. Let us call this *Analysis A*. As illustrated in Figure 3, this leads to the construction of one binary right-dominant QI foot at the left edge

[13]This is based on Hayes' analysis of *penultimate* stress ([Hayes 1980, p. 55]).

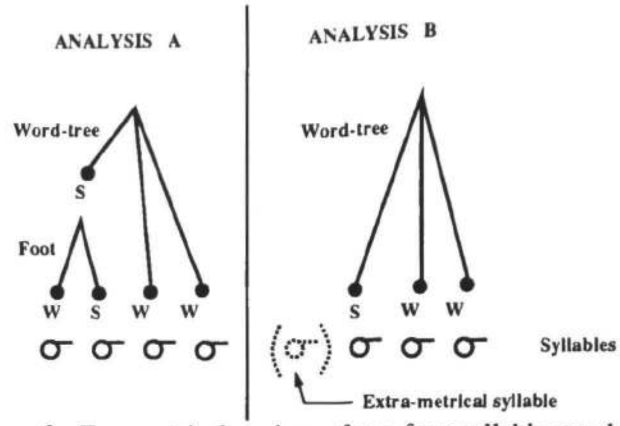

Figure 3: Two metrical analyses for a four-syllable word in Lakota. Strong branches are labeled "S", and weak branches "W".

of the word. This, together with the left-dominant word-tree, results in the assignment of primary stress to the second syllable. As has been shown, under this analysis, the perceptron learning results support the markedness of non-iteration (recall the differing learning times of Groups 1 through 4, vs. Groups 5 through 9).

However, an alternative analysis is that Lakota has a left-dominant word-tree with no metrical feet, and the first syllable is extra-metrical ([Dresher & Kaye 1990, p. 143]). Let us call this *Analysis B*. As illustrated in Figure 3, the leftmost syllable is treated as "invisible" to the stress rules, and the word-tree assigns primary stress to the leftmost of the "visible" syllables. The result is that the second syllable receives primary stress. Under this analysis, Lakota and Polish (Group 5, in Table 2) differ from Latvian and French (Group 1 in Table 2) only in having an extra-metrical syllable. The differing learning times for the two groups (1 epoch vs. 9 epochs) then suggest that extra-metricality is *marked*. However, this runs counter to both the distributional view ([Hayes 1980, p. 82]) and the learnability-theoretic view ([Dresher & Kaye 1990, p. 191]).

To summarize, *Analysis A* views Lakota and Polish as having non-iterative feet, which both the distributional/theoretical and learnability approaches treat as marked. *Analysis B* views these stress patterns as having an extra-metrical syllable, which both approaches treat as unmarked. So far, there is nothing theory-external to help choose between the analyses. We claim that the present simulation results provide such a means: since the learning results are consistent with the theoretical markedness of non-iteration, but not with the unmarkedness of extra-metricality, they provide at least weak support for preferring *Analysis A* over *Analysis B*.

In summary, the present learning results are from a model whose initial state is devoid of any information about the constructs of metrical theory that characterize different stress systems. Nevertheless, the learning results exhibit interesting correspondences with theoretical predictions. These results suggest that computational modeling may have something to contribute to the development of a markedness theory and, more generally, to aspects of linguistic analysis.

Analysis of weight patterns

In learning a stress pattern, the perceptron has acquired and encoded in its connection weights its "knowledge" of that pattern. Connection weights for all the languages studied are shown in Figure 4. Each display is a representation of the network as

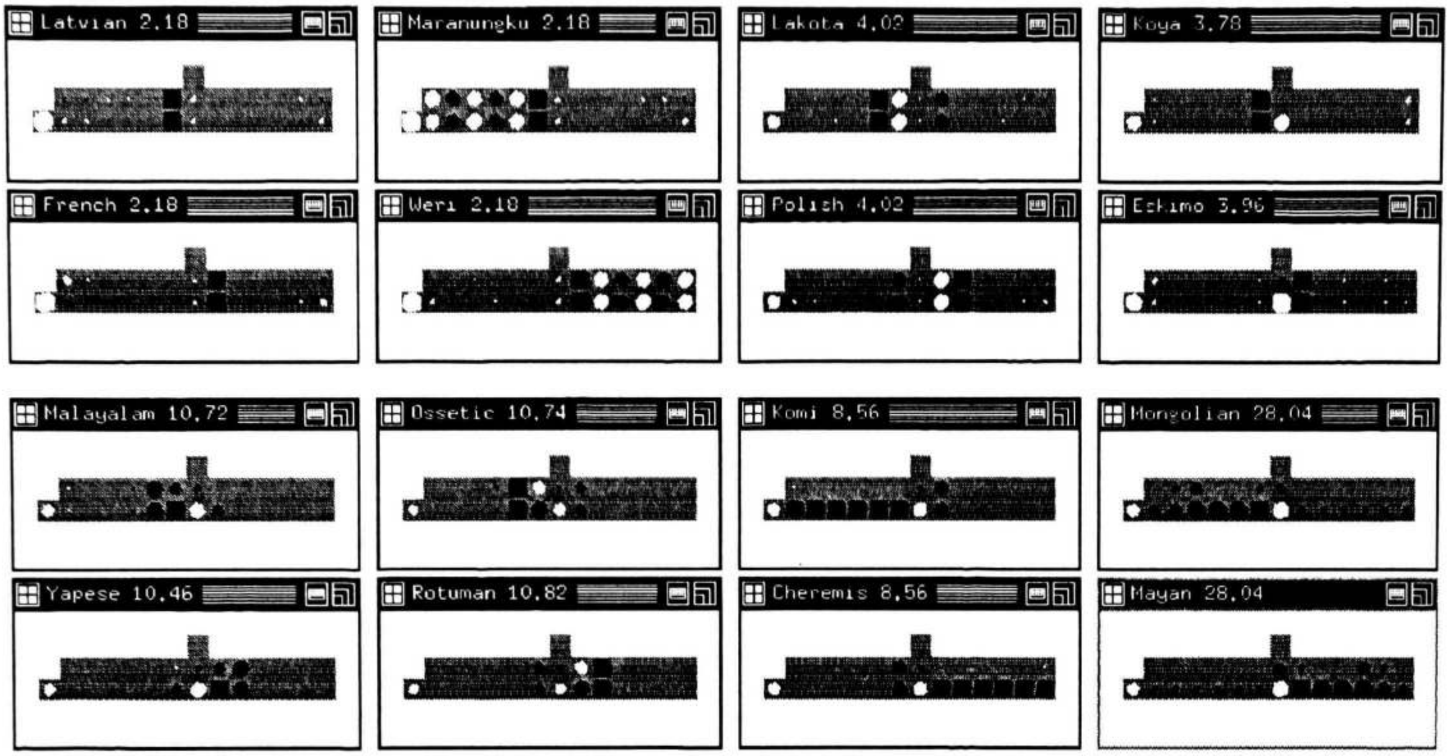

Figure 4: Learned connection weights for the sixteen stress patterns.

a whole. The large grey shaded rectangle represents the input layer of the network, with its two rows of input units. The single square protruding from the left of the input layer is the *bias* connection. The output unit is represented by the protruding square at the top of the input layer.

A blob in a particular position denotes a weight from the unit in that position to the output unit. White blobs denote positive weights, and black blobs negative weights. The size (area) of the blobs is proportional to the absolute magnitude of the weight. Weights are scaled so that the largest absolute magnitude is depicted in each display as a perfect square; other weights in that display appear as blobs of proportionate size. The scale (i.e., absolute magnitude of the largest weight) is shown in the title bar of each display. Thus, for Maranungku, the absolute magnitude of the largest weights is 2.18; these are the large (black) negative weights left of center in the input layer.

Just as with learning times, the fact that two stress patterns are mirror images of each other is reflected in the connection weights. Moreover, there seem to be correspondences between the form of the encoded knowledge and the characterization of the stress pattern in terms of parameters. Maranungku and Weri are the only stress systems with iterative binary feet (Group 2, Table 2). For these systems, but for no others, there is a very clear binary alternating pattern of positive and negative weights (see the alternation of black and white blobs in the weight displays for Maranungku and Weri). If, as is natural, we take a positive weight to correspond to the strong branch, and a negative weight to correspond to the weak branch of a foot, then for Maranungku we see left-dominant binary feet, and for Weri right-dominant binary feet – just as in the theoretical analysis[14]. It does not seem too far-fetched to say that the perceptron has discovered a version of iterative binary feet.

The single set of negative weights for Latvian and French (immediately to the left and right of center, respectively) can perhaps be interpreted as a left-dominant and right-dominant word-tree.

Recall that Lakota has non-iterative binary right dominant QI feet constructed from left to right, and that Polish has non-iterative binary left dominant QI feet constructed from right to left. That is, there will be a single binary tree, constructed at the left edge of the word for Lakota, and at the right edge, for Polish. Under this analysis, the weights to left and right of center for Lakota and Polish can be interpreted respectively as (single) right-dominant and left-dominant binary QI feet.

The weight patterns for Koya and Eskimo are close to mirror images, but not completely symmetric. Koya assigns primary stress to the first syllable and secondary stress to non-initial heavy syllables, while Eskimo assigns only one level of stress to final and heavy syllables. The chief theoretical difference between the two languages is that the former, but not the latter, has a word-tree. This difference is reflected in the fact that there are two magnitudes, or levels, of center connection weights for Koya (the large negative, and the smaller positive weights), whereas for Eskimo, there is only one level of weights (the positive and negative weights at the center are approximately equal.) This can be viewed as analogous to the two levels of metrical structure in Koya (metrical feet and word-tree) vs. the single level of structure in Eskimo (metrical feet only.)

Table 2 shows that Malayalam, Yapese, Ossetic and Rotuman (Groups 6 and 7) are the only languages with non-iterative binary QS feet. These are also the only patterns that have more than two large negative weights grouped together to the left (for Malayalam and Ossetic) or right (for Yapese and Rotuman) of center. We can take these three-or-four negative weight struc-

[14] As discussed previously, Maranungku has binary, left-dominant QI feet constructed iteratively from the left edge of the word. Weri has binary, right-dominant QI feet constructed iteratively from the right edge of the word.

tures to correspond to a non-iterative binary QS foot. There is a clear structural difference as compared with the (analogues of) non-iterative binary QI feet in the weights for Lakota and Polish.

Komi, Cheremis, Mongolian and Mayan are the only languages with non-iterative unbounded QS feet (Groups 8 and 9, Table 2). The connection weights for these systems show a pattern of nearly-identical negative weights spanning a set of several input units, and such a pattern does not occur for any other of the stress systems. Such a set of "spanning" weights seems analogous to an unbounded foot. The pattern of weights for Komi seems to correspond to an unbounded right-dominant QS foot, while weights for Cheremis seem to correspond to an unbounded left-dominant QS foot (note the single positive weight at the *right* and *left*, respectively, of the sets of weights, similar to the dominant branch of the foot). The difference in analysis between Komi & Cheremis and Mongolian & Mayan is that feet in the latter pair have obligatory branching, meaning the strong node of the foot must dominate a heavy syllable. As for Komi and Cheremis, the weights for Mongolian and Mayan show a pattern that can be interpreted as an unbounded QS foot. However, they additionally have a set of weights adjacent to the positive weight (i.e., to the "dominant branch" of the unbounded foot), which are not present for Komi and Cheremis; these additional weights can loosely be interpreted as corresponding to a branching dominant node.

It should not be too surprising that correspondences can be found between learned weight patterns and metrical constructs: both representations are derived from the same data in an attempt to explain or model the same phenomena. Nonetheless, we find it somewhat encouraging that these correspondences exist.

Conclusions

One of the supposed attractions of connectionist learning is that networks can start out with no theoretical preconceptions, developing their representations *de novo*. In truth, though, any computational model is founded on preconceptions, and ours is no exception. We have preprocessed the data to remove all detail from the words except for the heavy/light syllable distinction. In addition, we assume that all the syllables of the word are simultaneously accessible, i.e., that the word fits in a buffer, and that the stress assignment process can access the entire buffer in parallel. These are not unreasonable assumptions, but neither are they indisputable.

We have tried to show that, within the range of observed human stress systems, our model exhibits interesting behaviors, providing a fresh source of insight about linguistic stress. However, the model does not accurately reflect human capabilities for processing stress systems. For example, it is incapable of learning some existing human stress patterns (e.g., Warao, Southern Paiute), and is capable of learning stress patterns that are unlikely ever to exist amongst human languages. (However, as has been argued here, distributional evidence is in itself only weakly suggestive of what is "impossible"). Thus, our model is not suitable for defining the boundaries of human abilities, but this is scarcely surprising given that we are merely optimizing a linear threshold function.

Nevertheless, this simple model has produced interesting results. Although the simulations did not incorporate the foot and word tree constructs of metrical phonology as computational primitives, there appear to be correspondences between the performance of the model and the analyses of metrical theory. First, the learning results correspond with theoretical predictions of markedness. The theoretical predictions, as they stand, may be based either on distributional facts or on learnability theory. Our simulation results provide predictions of markedness that are based on actual "learning." Second, there seem to be parallels between theoretical characterizations of various stress systems and the knowledge of those stress systems encoded in the perceptron's weights. These parallels are at least suggestive of a mapping between the level of representation at which the perceptron performs its computations and the level of investigation at which metrical phonology is formulated.

In conclusion, these experiments suggest that simple statistical learning techniques such as the perceptron model have the potential to provide computational validation for, and to complement, theoretical investigations of domains of language such as stress systems.

Acknowledgments

We would like to acknowledge the feedback provided by Deirdre Wheeler throughout the course of this work. The first author would like to thank David Evans for access to exceptional computing facilities at Carnegie Mellon's Laboratory for Computational Linguistics, and Dan Everett, Brian MacWhinney, Jay McClelland, Eric Nyberg, Brad Pritchett and Steve Small for helpful discussion of earlier versions of this paper. Of course, none of them is responsible for any errors.

References

[Berwick 1985] Berwick, R. 1985. *The Acquisition of Syntactic Knowledge*. Cambridge, Mass.: MIT Press.

[Davis 1988] Davis, S. 1988. *Topics in Syllable Geometry*. New York: Garland.

[Dresher & Kaye 1990] Dresher, B. E. & Kaye, J. D. 1990. A Computational Learning Model for Metrical Phonology. *Cognition* 34: 137-195.

[Everett & Everett 1984] Everett, D. & Everett, K. 1984. On the Relevance of Syllable Onsets to Stress Placement. *Linguistic Inquiry* 15: 705-711.

[Goldsmith 1990] Goldsmith, J. A. 1990. *Autosegmental and Metrical Phonology*. Oxford, England: Basil Blackwell.

[Halle & Clements 1983] Halle, M. & Clements, G. 1983. *Problem Book in Phonology*. Cambridge, Mass.: MIT Press.

[Halle & Vergnaud 1987] Halle, M. & Vergnaud, J.-R. 1987. *An Essay on Stress*. Cambridge, Mass: MIT Press.

[Hayes 1980] Hayes, B. 1980. A Metrical Theory of Stress Rules. Ph.D. diss., Dept. of Linguistics, Massachusetts Institute of Technology.

[Kaye 1989] Kaye, J. D. 1989. *Phonology: A Cognitive View*. Hillsdale, NJ: Lawrence Erlbaum.

[Nyberg 1989] Nyberg, E. 1989. Weight Propagation and Parameter Setting, Ph.D. thesis proposal, Dept. of Philosophy, Carnegie Mellon University.

[Pullum 1982] Pullum, G. 1982. Letter, *Linguistics* 20: 339-344.

[Rumelhart, Hinton & Williams 1986] Rumelhart, D., Hinton, G. & Williams, R. 1986. Learning Internal Representations by Error Propagation. In D. Rumelhart, J. McClelland & the PDP Research Group. *Parallel Distributed Processing, Volume 1: Foundations*. Cambridge, Mass.: MIT Press.

[van der Hulst & Smith 1982] van der Hulst, H. & Smith, N. 1982. An Overview of Autosegmental and Metrical Phonology. In H. van der Hulst & N. Smith (eds.). *The Structure of Phonological Representations, Part I*. Dordrecht, The Netherlands: Foris.

[Wexler & Manzini 1987] Wexler, K., & Manzini, M. 1987. Parameters and Learnability in Binding Theory. In T. Roeper & E. Williams (eds.). *Parameters and Linguistic Theory*, Dordrecht, The Netherlands: Reidel.