Connectionist Models and Linguistic Theory: Investigations of Stress Systems in Language

Prahlad Gupta
Department of Psychology
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

December 22, 1991

Abstract

This paper discusses a perceptron model of the learning and assignment of linguistic stress, using data from nineteen human languages. First, we point out some interesting parallels between aspects of the model and the constructs and predictions of *metrical phonology*, the linguistic theory of stress. Second, we develop a novel analysis of linguistic stress in terms of ease of perceptron-learnability. These two sets of results suggest that simple statistical learning techniques have the potential to complement, and provide computational validation for, abstract theoretical investigations of language.

We then examine why such methodologies should be of interest for linguistic theorizing. Our analysis began at a high level by observing inherent characteristics of various stress systems, much as theoretical linguistics does. However, our explanations changed substantially when we included a detailed account of the model's processing mechanisms. Our higher-level, theoretical account of stress was revealed as only an *approximation* to the lower-level computational account. Without the ability to open up the black boxes of the human processor, linguistic analyses are arguably analogous to our higher-level descriptions. This highlights the need for *computational grounding* of theory-building. In addition, we suggest that there are methodological problems underlying parameter-based approaches to learnability. These problems make it all the more important to seek sources of converging evidence such as is provided by computational models.

Contents

Connectionist Models and Linguistic Theory: Investigations of Stress Systems in Language

Prahlad Gupta
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

David S. Touretzky
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

1. Introduction

This work describes the use of connectionist techniques to model the learning and assignment of linguistic stress. Our aim was to explore the ability of a simple perceptron to model the assignment of stress in individual words, and to consider, in light of this study, the relationship between the connectionist and theoretical linguistics approaches to investigating language.

We first point out some interesting parallels between aspects of the model and the constructs and predictions of *metrical phonology*, the linguistic theory of stress: (1) the distribution of learning times obtained from perceptron experiments corresponds with theoretical predictions of "markedness," and (2) the weight patterns developed by perceptron learning bear a suggestive *structural* relationship to features of the linguistic analysis, particularly with regard to iteration and metrical feet.

We use the connectionist learning data to develop an analysis of linguistic stress based on perceptron-learnability. We develop a novel characterization of stress systems in terms of six parameters. These provide both a partial description of the stress pattern itself and a prediction of its learnability, without invoking abstract theoretical constructs such as *metrical feet*. Our parameters encode linguistically salient concepts as well as concepts that have computational significance.

These two sets of results suggest that simple connectionist learning techniques have the potential to complement, and provide computational validation for, abstract theoretical investigations of linguistic domains.

We then examine why such methodologies should be of interest for linguistic theorizing. Our analysis began at a high level by observing inherent characteristics of various stress systems, much as theoretical linguistics does. However, our explanations changed substantially when we included a detailed account of the model's processing mechanisms. Our higher-level, theoretical account of stress was revealed as only an *approximation* to the lower-level computational account. Without the ability to open up the black boxes of the human processor, linguistic analyses are arguably analogous to our higher-level descriptions. This highlights the need for *computational grounding* of theory-building. In addition, we suggest that there are methodological problems underlying parameter-based approaches to learnability. These problems make it all the more important to seek sources of converging evidence such as is provided by computational models.

2. Background: Stress systems in language

2.1. Motivation for choice of domain

Stress systems are an attractive domain for investigation because, as noted by Dresher & Kaye ([Dresher 90]): (a) the linguistic theory is well-developed, so that compared with syntax, there is a relatively complete description of the observed phenomena, and (b) stress systems can be studied relatively independently of other aspects of language ([Dresher 90, page 1]). Comparing connectionist and "classical" analyses of a well-defined linguistic domain, for which theoretical analyses already provide good coverage, should prove interesting.

2.2. Evolution of the linguistic theory

The analysis of stress has evolved through a number of phases: (I) Linear analyses presented stress as a phonemic feature of individual vowels, with different levels of stress representing different levels of absolute prominence. This approach was taken in [Trager 51], and culminated in Chomsky & Halle's seminal *The Sound Pattern of English* ([Chomsky 68]). (II) Metrical theory, as developed in [Liberman 75] and [Liberman 77], introduced both a non-linear analysis of stress patterns (in terms of *metrical trees*), and the treatment of stress as a *relative* property rather than an *absolute* one; however, the *stress* feature was retained in the analysis. (III) In subsequent developments ([Prince 76], [Selkirk 80]), reliance on this feature was eliminated by incorporation of the idea that subtrees of metrical trees had an independent status (*metrical feet*), so that stress assignment rules could make reference to them. (IV) The positing of internal structure for syllables ([Vergnaud 78], [McCarthy 79a], [McCarthy 79b]) provided a means of distinguishing light and heavy syllables, a distinction to which stress patterns are widely sensitive, but which had been problematic under previous analyses. (V) An analysis of metrical tree geometries ([Hayes 80]) provided an account of many aspects of stress systems in terms of a small number of *parameters*.

Through the development of metrical theory, there has been debate over whether the auto-segmental representations for stress are *metrical trees* only ([Hayes 80]), *metrical grids* only ([Prince 83], [Selkirk 84]), or some combination of the two ([Liberman 75], [Liberman 77], [Hayes 84a], [Hayes 84b], [Halle 87a], [Halle 87b]).

2.3. Syllable structure

A syllable is analyzed as being comprised of an *onset*, which contains the material before the vowel, and a *rime*. The rime is comprised of a *nucleus*, which contains the vocalic material, and a *coda*, which contains any remaining (non-vocalic) material. For further discussion, see [Kaye 89, pp. 54-58].

A syllable may be *open* (it ends in a vowel); or *closed* (it ends in a consonant). In terms of syllable structure, an open syllable has a *non-branching* rime (the rime has a nucleus, but not a coda), and a closed syllable has a *branching* rime (the rime has both a nucleus and a coda).

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 a heavy or a light syllable may differ across languages in which such a distinction is present, but, most commonly, a heavy syllable is one that can be characterized as having a branching rime, and a light syllable can be characterized as having a non-branching rime. ([Goldsmith 90, page 113]). Languages that involve such a distinction (between heavy and light syllables, i.e., between the *weight* of syllables) are termed *quantity-sensitive*, and languages that do not, *quantity-insensitive*. Note that, in quantity-insensitive languages, syllables can occur both with and without branching rimes; but the distinction between these kinds of syllables has no relevance for the placement of stress.

2.4. Metrical phonology

There seems to be theoretical agreement that stress patterns are sensitive to information about syllable structure, and in particular, to the structure of the syllable *rime*, and not the syllable *onset*. We follow this assumption[1]. Thus rime structure is taken to be the basic level at which accounts of stress systems are formulated. (For an overview of metrical theory, see [Goldsmith 90, chapter 4], [Kaye 89, pp. 139-145], [van der Hulst 82] or [Dresher 90, pp. 1-8]). 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 90, 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 extrametrical syllable [Yes/No]
(P9) It is extrametrical 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

[1] See, for example, ([Dresher 90, p. 141] or [Goldsmith 90, p. 170]). However, both [Davis 88] and [Everett 84] present evidence that onsets may in fact be relevant to the placement of stress.

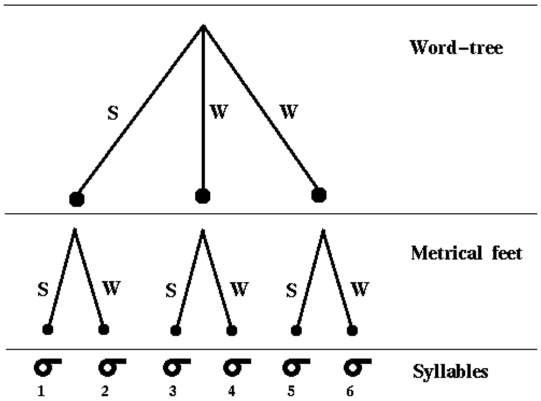

Figure 1: Metrical structures for a six-syllable word in Maranungku.

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 do not apply because of the settings of parameters P5 and P8, respectively.

2.5. Principles and parameters

Metrical theory illustrates the *principles and parameters* approach to language, one of whose central hypotheses is that language learning proceeds through the discovery of appropriate parameter settings. Every possible human language can be characterized in terms of parametter settings; once these settings are determined, the nature of structure-sensitive operations and the structures on which they operate is known, so that the details of language processing are automatically determined (at an abstract level). Subsequently, the assignment of stress in the actual *production* or *processing* of language is assumed to involve neural processes that correspond quite directly with the abstract process of application of these parameter settings as guidelines to the construction and manipulation of metrical feet[2].

2.6. Previous computational models of stress learning

Computational models of stress systems in language processing have been developed by Dresher & Kaye ([Dresher 90]) and by Nyberg ([Nyberg 90, Nyberg 92]. The focus of these models is on the learning of the *parameters* specified by metrical theory; they therefore take as a starting point the constructs of that theory, and incorporate its assumptions. What they add to the linguistic theory is what Dresher & Kaye term a *learning theory*: a specification of how data the language learner encounters in its environment are to be used to set parameters. The following features characterize these models: (1) they assume the existence of processes explicitly corresponding to the linguistic notion of parameter setting; (2) they propose a *learning* theory as an account of that parameter-setting process; (3) they assume that the process of *production* (i.e., of producing appropriate stress contours for input words after learning has occurred) involves explicit representational structures and structure-sensitive operations directly corresponding to metrical-theoretic trees and operations on those trees; (4) they assume no necessary relationship between the processing mechanisms and structures involved in *learning* vs. those involved in *production*. That is, learning is accomomplished by supplying values for the parameters defined by metrical theory; these values then form a knowledge base for stress

[2]Thus, Fodor & Pylyshyn ([Fodor 88, p. 13]): "... the symbol structures in a Classical model are assumed to correspond to real physical structures in the brain and *the combinatorial structure of a representation is supposed to have a counterpart in the structural relations among physical properties of the brain* ..." (emphasis added).

assignment, whose processing involves, for example, the construction of binary trees from right to left – an operation having no necessary correspondence with those by which the parameter values were acquired.

The work reported here differs from the Dresher & Kaye and Nyberg models with regard to these same characteristics: (1) the aim here is to explore the issue of learning of stress systems without explicit incorporation of parameters; (2) the *learning theory* employed consists of one of the general learning algorithms common in connectionist modeling, and is not an account of parameter-setting; (3) the process of production does not involve explicitly structured representations in the classical sense; (4) the processing mechanisms and structures involved in production are essentially the same as those involved in learning.

All of these differences are, of course, primarily the domain-specific manifestations of classical-connectionist contrasts.

3. Overview of model and simulations

3.1. Nineteen stress systems

Nine 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 80]. Note that the QI stress patterns of Latvian & French, Maranungku & Weri, Lakota & Polish, and Paiute and Warao are mirror images of each other. The QS stress patterns of Malayalam & Yapese, Ossetic & Rotuman, Eastern Permyak Komi & Eastern Cheremis, and Khalka Mongolian & Aguacatec Mayan are also mirror images[3].

3.2. Structure of the model

In separate experiments, we taught a perceptron to produce the stress pattern of each of the nineteen languages. The domain was limited to single words, as in the previous learning models of metrical phonology developed by Dresher & Kaye, and Nyberg ([Dresher 90, Nyberg 90, Nyberg 92]). 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.

Two input representations were used. In the *syllabic* representation, used for QI patterns only, a syllable was represented as a [1 1] vector, and [0 0] represented no syllable. In the *weight-string* representation, which was necessary for QS languages, the input patterns used were [1 0] for a light syllable, [0 1] for a heavy syllable, and [0 0] for no syllable. For stress systems with up to two levels of stress, the output targets used in training were 1.0 for primary stress, 0.5 for secondary stress, and 0 for no stress. For stress systems with three levels of stress, the output targets were 1.0 for primary stress, 0.6 for secondary, 0.35 for tertiary, and and 0 for no stress. The input data set for all stress systems consisted of all word-forms of up to seven syllables. With the syllabic input representation there are seven of these, and with the weight-string input representation, there are 255 distinct patterns[4]. The perceptron's input array was a buffer of 13 syllables; each word was processed one syllable at a time by sliding it through the buffer (see Figure 2). The desired output at each step was the stress level of the middle syllable of the buffer. Connection weights were adjusted at each step using the back-propagation learning algorithm[5] ([Rumelhart 86]). 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 were used in all simulations. Initial weights were

[3] We have somewhat simplified the descriptions of Polish and Malayalam compared with those in [Halle 87b, pp. 57-58] and [Hayes 80, 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 83, p.17]) and Gurkhali ([Hayes 80, p.66]), corresponding to Polish and Malayalam, respectively.

[4] In practice, we used weight string training sets in which there were an equal number of input patterns of each length. Thus, there was one instance of each of the 128 (= 2^7) seven-syllable patterns, and 64 instances of each of the two monosyllabic patterns. This length balancing was necessary for successful training.

[5] Note that although the architecture of the model is two-layered, with a single ouput unit, as in a simple perceptron, we used the back-propagation algorithm (BP) rather than the Widrow-Hoff algorithm (WH; [Widrow 60]). BP adds to WH the scaling of the error for each output unit by a function (the derivative) of the activation of that output unit, and thus performs a more sensitive and locally tuned weight adjustment than WH. Note that BP and WH algorithms for two layers are guaranteed by the perceptron convergence procedure to be equivalent in terms of learning capabilities, *for binary-valued outputs*. However, outputs in the present simulations are not always binary; they sometimes take on intermediate values. As a result, the different computation of the error term in BP turned out to provide better learning than that of WH.

REF	LANGUAGE	DESCRIPTION OF STRESS PATTERN	EXAMPLES
Quantity-Insensitive Languages:			
L1	Latvian	Fixed word-initial stress.	$S^1S^0S^0S^0S^0S^0S^0$
L2	French	Fixed word-final stress.	$S^0S^0S^0S^0S^0S^0S^1$
L3	Maranungku	Primary stress on first syllable, secondary stress on alternate succeeding syllables.	$S^1S^0S^2S^0S^2S^0S^2$
L4	Weri	Primary stress on last syllable, secondary stress on alternate preceding syllables.	$S^2S^0S^2S^0S^2S^0S^1$
L5	Garawa	Primary stress on first syllable, secondary stress on penultimate syllable, tertiary stress on alternate syllables preceding the penult, no stress on second syllable.	$S^1S^0S^0S^3S^0S^2S^0$
L6	Lakota	Primary stress on second syllable.	$S^0S^1S^0S^0S^0S^0S^0$
L7	Polish	Primary stress on penultimate syllable.	$S^0S^0S^0S^0S^0S^1S^0$
L8	Paiute	Primary stress on second syllable, secondary stress on alternate succeeding syllables.	$S^0S^1S^0S^2S^0S^2S^0$
L9	Warao	Primary stress on penultimate syllable, secondary stress on alternate preceding syllables.	$S^0S^2S^0S^2S^0S^1S^0$
Quantity-Sensitive Languages:			
L10	Koya	Primary stress on first syllable, secondary stress on heavy syllables. (Heavy = closed syllable or syllable with long vowel.)	$L^1L^0L^0H^2L^0L^0L^0$ $L^1L^0L^0L^0L^0L^0L^0$
L11	Eskimo	(Primary) stress on final and heavy syllables. (Heavy = closed syllable.)	$L^0L^0L^0H^1L^0L^0L^1$ $L^0L^0L^0L^0L^0L^0L^1$
L12	Malayalam	Primary stress on first syllable except when first syllable light and second syllable heavy. (Heavy = long vowel.)	$L^1L^0L^0H^0L^0L^0L^0$ $L^0H^1L^0H^0L^0L^0L^0$
L13	Yapese	Primary stress on last syllable except when last is light and penultimate heavy. (Heavy = long vowel.)	$L^0L^0L^0H^0L^0L^0L^1$ $L^0H^0L^0H^0L^0H^1L^0$
L14	Ossetic	Primary stress on first syllable if heavy, else on second syllable. (Heavy = long vowel.)	$H^1L^0L^0H^0L^0L^0L^0$ $L^0L^1L^0L^0L^0L^0L^0$
L15	Rotuman	Primary stress on last syllable if heavy, else on penultimate syllable. (Heavy = long vowel.)	$L^0L^0L^0H^0L^0L^0H^1$ $L^0L^0L^0L^0L^0L^1L^0$
L16	Komi	Primary stress on first heavy syllable, or on last syllable if none heavy. (Heavy = long vowel.)	$L^0L^0H^1L^0L^0H^0L^0$ $L^0L^0L^0L^0L^0L^0L^1$
L17	Cheremis	Primary stress on last heavy syllable, or on first syllable if none heavy. (Heavy = long vowel.)	$L^0L^0H^0L^0L^0H^1L^0$ $L^1L^0L^0L^0L^0L^0L^0$
L18	Mongolian	Primary stress on first heavy syllable, or on first syllable if none heavy. (Heavy = long vowel.)	$L^0L^0H^1L^0L^0H^0L^0$ $L^1L^0L^0L^0L^0L^0L^0$
L19	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, tertiary stress by the superscript 3, 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.

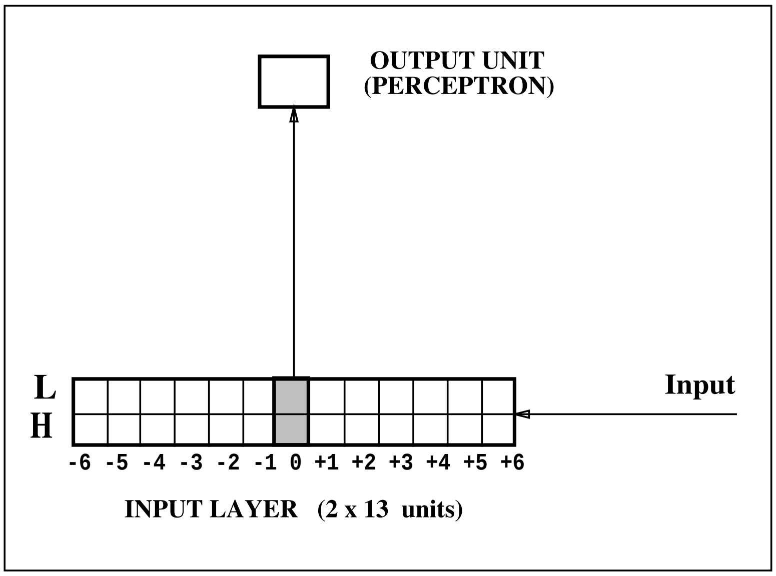

Figure 2: Perceptron model used in simulations.

uniformly distributed random values in the range ± 0.5. Each simulation was run at least three times, and the learning times averaged.

3.3. Training the perceptron

To enable precise description, let the input buffer be regarded as a 2x13 array as shown in Figure 2. Let the 7th, i.e. center, column be numbered 0, let the columns to the left of the center be numbered negatively -1 through -6 going outwards from the center, and the columns to the right of the center be numbered positively +1 through +6, going outwards from the center.

As an example of processing, suppose that the next word in the training set for some quantity-insensitive language is the four-syllable pattern *[S S S S]*, and the associated stress contour (i.e., target), is *[0 0 0 1]* (indicating that the final syllable receives stress, and all other syllables are unstressed)[6]. The first syllable enters the rightmost element of the input buffer (element +6). This is the first "time step" of processing. At the next time step, the first syllable (in the rightmost input buffer position, i.e., element +6) is shifted left to buffer element +5, and the second syllable enters the input buffer in element +6.

After two further time steps, elements +3, +4, +5, and +6 of the input buffer contain the four syllables of the current word. At the next two time steps, the leftward flow of syllables through the input buffer continues, until at time step 6, the word's four syllables are in elements +1 through +4. At time step 7, the first syllable moves into element 0 of the input buffer, and the four syllables of the word occupy elements 0 through +3 of the buffer. At this time step, training of the network occurs for the first time: the perceptron is trained to associate the pattern in its input buffer (the four syllables of the word, in buffer elements 0 through +3) with the target stress level for the syllable currently in buffer element 0. At the next time step, the syllables of the word are shifted left again, and occupy buffer elements -1 through +2. The network is trained to associate this pattern with the target stress level for the second syllable, which is now in element 0. Similarly, at the next two time steps, the perceptron is trained to produce the stress levels appropriate for the third and fourth syllables of the word, as they come to occupy element 0 of the input buffer.

At this point, the input buffer is flushed (all elements of the buffer are set to zero), and one "trial" is over. The next word in the training set can now enter, beginning the next trial.

Thus the processing of one word, syllable by syllable, constitutes one *trial*. One pass through all the words in the training set constitutes an *epoch*.

It should be noted that sequential processing of syllables is not a necessary part of this model. Exactly the same effect can be obtained by a parallel scheme of thirteen perceptron-like units whose weight vectors are tied together by

[6]To simplify discussion here, each syllable is represented as an "S" token, rather than as an "H" or "L". For quantity-insensitive syatems, this information suffices to determine placement of stress.

weight-sharing ([Hertz 91, p. 140]). All thirteen units would learn in unison, and words could be processed in one parallel step.

4. Relationship to metrical phonology

4.1. Markedness and learnability

Within the dominant linguistic tradition, a universal grammar 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 ([Hayes 80, 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 ...

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 many historical, non-linguistic, factors, and does not necessarily have linguistic-theoretic significance. To quote Pullum ([Pullum 82, 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 ...

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 90, 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.

If learnability arguments should propose the default setting x for parameter P, then some explanation would be needed if 95% 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.

4.2. Markedness and learning times

Table 2 shows the stress systems grouped by their theoretical analyses in terms of the parameter scheme discussed in Section 2.4. The last column of the table shows the average learning time in epochs for each group of stress patterns[8].

---

[7] As an example, the Subset Principle ([Berwick 85], [Wexler 87]) 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, no positive evidence can ever re-set it to a, even if that was the correct setting. Unmarked values for parameters should therefore be the ones yielding the most constrained system.

[8] As noted in Section 3.2, *weight-string* representations are necessary for QS stress patterns. For QI systems, *syllabic* representations are sufficient. Of course, weight-string representations can be used for QI systems, although the information about syllable weight will be redundant. To obtain learning times that might reflect differences between the stress patterns themselves, rather than merely reflecting differing input representations and training set sizes, we ran simulations for both QS and QI patterns using weight-string input representations. The figures in Table 2 are based on use of this weight-string representation for both QI and QS 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[9].

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[10], and with the choice of marked and unmarked settings in Dresher & Kaye's parameter scheme (Parameter P11)[11].

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[12]. Group 8 differs from group 9 only in not having obligatory branching, and average learning times were 214 epochs vs. 2302 epochs.

This prediction agrees with the distributional view that obligatory branching is relatively marked[13]. However, it runs counter to Dresher & Kaye's choice of default values (parameter P7)[14].

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 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 80, p. 113]) seems to conflict with the learnability view ([Dresher 90, p. 193]). The present simulations seem to agree with the distributional view, and, we suggest, serve as a buttress for theoretical arguments by providing an additional source of evidence[15].

This contribution to theoretical analysis can be further illustrated for the stress systems of Lakota and Polish, which are mirror images. Recall that in Lakota, primary stress falls on the second syllable of the word. 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

---

[9] The stress system of Garawa, as described in [Hayes 80, pp. 54-55], cannot be characterized in terms of the parameter scheme adopted here. The stress systems of Paiute and Warao cannot be learned by a perceptron, as will be discussed in more detail in Section 6.1. Learning results for these three stress systems have therefore been excluded from the present analysis, as either the parameterized characterization or learning times cannot be established. (They can, however, be learned using either a two-layer architecture with two output units, or a three-layer architecture with "hidden" units.)

[10] "The rules of constituent boundary construction apply exhaustively ..." ([Halle 87b, p. 15]).

[11] 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 90, p. 191]).

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

[13] [Hayes 80, 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 ..."

[14] 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 90, p. 193]).

[15] It is worth noting that learning times for these stress patterns have also been obtained with a three-layer architecture. Differences between learning times for different stress patterns were somewhat reduced. However, learning times were *ordered* exactly as in the two-layer model whose results are shown in Table 2, suggesting that the learning results just discussed are robust with regard to architectural differences. In the discussion throughout this paper, however, we focus on just results from the two-layer model. One of our aims was to examine the nature of the connection weights that develop as a result of training; these patterns are less clear when a hidden layer is involved.

	LANGUAGE	CHARACTERIZATION	EPOCHS
1	Latvian, French	Word-tree, no feet	2
2	Koya	Word-tree, iterative unbounded QS feet	2
3	Eskimo	No word-tree, iterative unbounded QS feet	3
4	Maranungku, Weri	Word-tree, iterative binary QI feet	3
5	Lakota, Polish	Word-tree, non-iterative binary QI feet	10
6	Malayalam, Yapese	Word-tree, non-iterative binary QS feet, dominant node branches	19
7	Ossetic, Rotuman	Word-tree, non-iterative binary QS feet	29 ($\pm$1)
8	Komi, Cheremis	Word-tree, non-iterative unbounded QS feet	214 ($\pm$2)
9	Mongolian, Mayan	Word-tree, non-iterative unbounded QS feet, dominant node branches	2302 ($\pm$4)

Table 2: Learning times for QI and QS stress patterns using weight-string input representations, grouped by theoretical analysis. Each figure is the average learning time for languages in the group.

left-dominant word-tree[16]. 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 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 extrametrical ([Dresher 90, 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 extrametrical syllable. The differing learning times for the two groups (1 epoch vs. 10 epochs) then suggest that extrametricality is *marked*. However, this runs counter to both the distributional view ([Hayes 80, p. 82])[17] and the learnability-theoretic view ([Dresher 90, p. 189,191])[18].

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 extrametrical 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 extrametricality, they provide at least weak support for preferring *Analysis A* over *Analysis B*.

In summary, the learning results in Table 2 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 model exhibits some

[16]This is based on Hayes' analysis of *penultimate* stress ([Hayes 80, p. 55]).

[17]Hayes argues for the importance of the device of extrametricality to the theory because of its role in accounting for a variety of stress phenomena. Thus, extrametricality is unmarked in the sense previously referred to: that of being "attested to in a fair variety of languages".

[18]In Dresher & Kaye's formulation, the presence of stress on a leftmost or rightmost syllable can rule out extrametricality. However, there is no positive cue that unambiguously determines the presence of extrametricality. Hence the default value for parameter P8 is that there is extrametricality (P8[Yes]). If the default were P8[No], but this was an incorrect setting, there would be no cue that could lead to detection that this is incorrect.

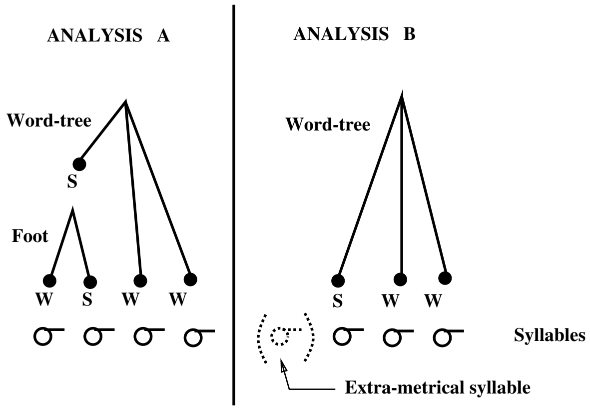

Figure 3: Two metrical analyses for a four-syllable word in Lakota. Strong branches are labeled "S", and weak branches "W". *Analysis A*: construction of one binary right-dominant QI foot at the left edge of the word, together with a left-dominant word-tree, results in the assignment of primary stress to the second syllable. *Analysis B*: the leftmost syllable is treated as "invisible" to the stress rules (*extrametrical*), and the word-tree assigns primary stress to the leftmost of the "visible" syllables. The result is that the second syllable receives primary stress.

interesting correspondences with theoretical predictions. (More detailed analyses follow.) These initial 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.

4.3. Connection weights and metrical theory

In learning a stress pattern, the perceptron has acquired and encoded in its connection weights its "knowledge" of that pattern. Connection weights for the sixteen stress patterns discussed in the previous section are shown in Figure 4. Each display is a representation of the network as a whole. The large grey shaded rectangle represents the input buffer of the network, organized as two rows of 13 values. The single square protruding from the left is the perceptron's *bias* connection. The perceptron unit itself is represented by the protruding square at the top.

A blob in a particular position denotes a weight on that input connection. 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 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[19]. 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.

[19] 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.

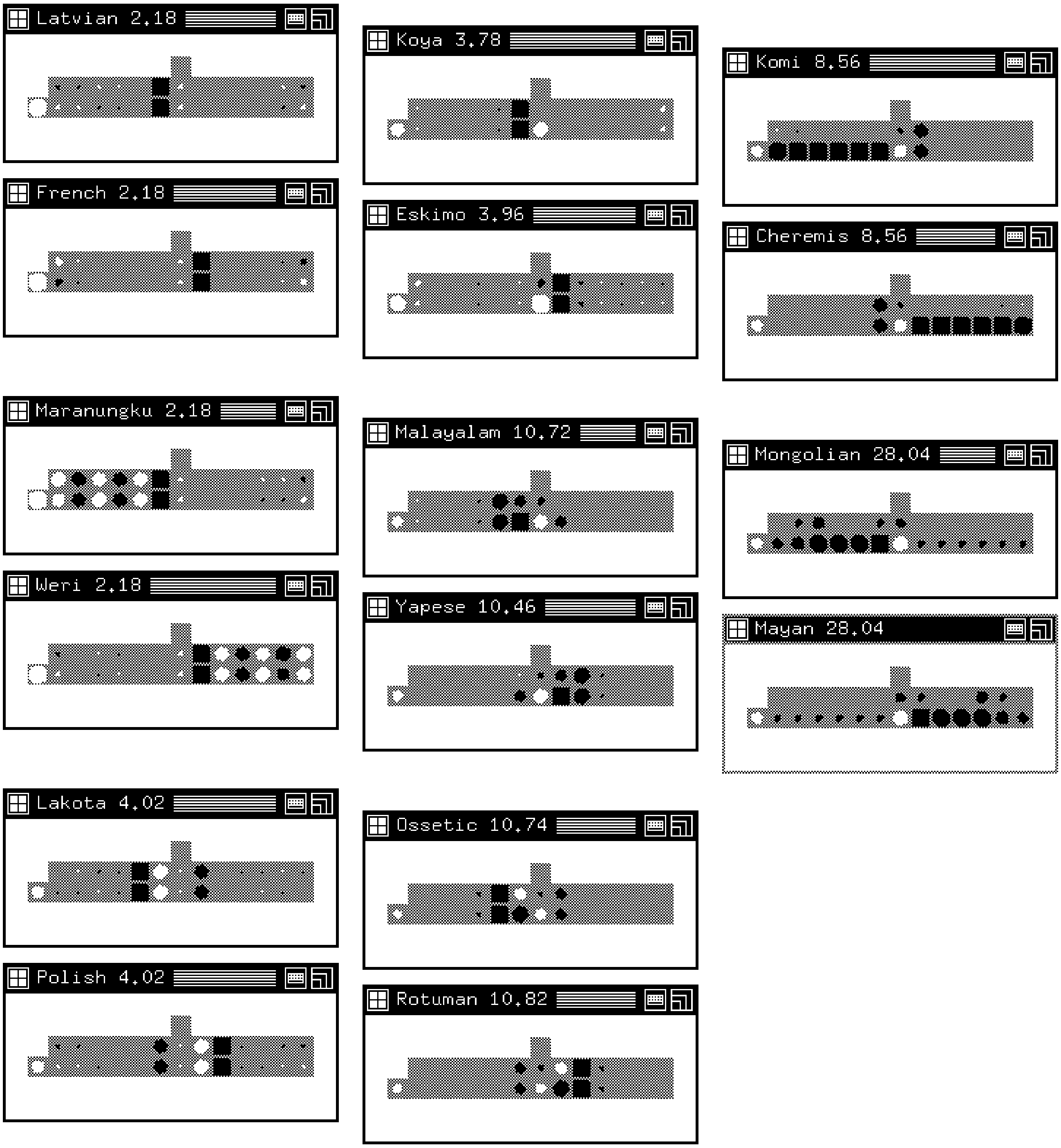

Figure 4: Learned connection weights for sixteen stress patterns. Each display is a representation of the network as a whole. The large grey shaded rectangle represents the input buffer of the network, organized as two rows of 13 values. The single square protruding from the left is the perceptron's *bias* connection. The perceptron unit itself is represented by the single protruding square at the top. 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 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 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.

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 structures 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 of the other 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.

4.4. Implication of correspondences

We have shown significant correspondences, first, between between the learning times of the network and linguistic characterizations of markedness, and second, between the structure of the network's weights when it processes various stress patterns and the linguistic characterizations of those stress patterns. Our simulations provide actual "learning" results that can be used for this kind of analysis. We find it encouraging that these correspondences exist; they suggest 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. As such, they represent a source of converging evidence and *computational validation* for investigation of the phenomena of linguistic stress. In addition, our demonstration of how these learning results can be used in consideration of theoretical issues shows how the simulations can provide *computational grounding*.

5. Computational analysis of stress learning

In this section, we focus on the nature of computations taking place in the network simulations when various stress patterns are learned. It may at this point be helpful to review the discussion of network processing in Sections 3.2 and 3.3.

Below, we first describe how the connection weights establish the trained network's ability to assign stress correctly, for a variety of the stress patterns examined. This discussion is intended to convey a flavor of how the contribution of the connection weights to network performance is interwoven with the processing dynamics of the model; this should aid in following the subsequent discussion of what factors seem to be determinants of learnability.

We next present an analysis of factors affecting learnability for the QI stress systems. Our analysis draws on inherent characteristics of the stress patterns, on learning times, and on our understanding of how the network processes the patterns. We develop an analytic scheme that enables us both to characterize the patterns, and to predict the ease or difficulty of their learning. In developing this analysis, we used results from simulations incorporating *syllabic* input representations (see Section 3.2).

We then include QS systems, and show that one analytical framework can take into account both QI and QS systems. Here, we use simulation results that employed *weight-string* representations for both QI and QS patterns (as in discussion of learning times shown in Table 2 and discussed in Section 4).

5.1. Connection weights and perceptron learning

We extend the numbering scheme for input connections given in Section 3.3. The central weights are numbered w_{0L} and w_{0H}; they are referred to collectively as w_0. The pair of weights immediately to their left is numbered w_{-1}, and so on.

For Latvian, the large negative weights w_{-1} enable detection of the left edge of a word: only the first syllable of a word passing through the input buffer from right to left will be unaffected by these weights when it is the "current input", i.e., in position 0 in the buffer (see Section 3.2 for discussion of processing in the networks). When any non-initial syllable of any word is the current input, there will be some other ("previous") syllable to its left in the buffer. Net input to the perceptron will be negative, since the magnitude of w_{-1} is greater than the magnitude of the positive bias weight. The output will therefore be low, denoting zero stress. The initial syllable of any word, however, will have no syllables to its left in the buffer, and so w_{-1} will have no effect. Net input to the perceptron will therefore be positive (from the bias connection), and so the output will be high, representing primary stress. For French, only the last syllable of a word will escape the effect of the large negative weights w_{+1}, and thus only the last syllable will receive stress. Connection weights for French are the mirror image of those for Latvian, just as the patterns themselves are mirror images of each other.

For Weri, the largest weights are w_{+1}; these are large negative weights. Consider the processing of, say, a six-syllable word. When the leftmost syllable is the "current input", and the target output is therefore that for zero stress, there will be four medium-strength positive weights (w_{+2} and w_{+4}) and four medium-strength negative weights (w_{+3} and w_{+5}), roughly canceling each other out, applying to four of the representations of syllabic elements in the buffer. There is also a pair of large negative weights (w_{+1}). The net input will therefore be negative, resulting in an output of zero stress. When the second syllabic element is the current input, the large negative weights w_{+1} still apply, as do the medium positive weights w_{+2} and w_{+4}. However, the medium negative weights applicable are now only w_{+3}; and so the net input is larger than for the previous syllable, and producing an output representing secondary stress. A similar pattern of alternation continues for all the syllables of the word: in each case, there will be either a balance of medium positive and negative weights applicable (resulting in zero stress), or one more pair of positive than negative weights, resulting in secondary stress. The exception is the last syllable: when this is the current input, there will be neither positive nor negative weights applying, but there will also not be the large negative weights w_{+1}. As a result, the net input will be higher for this syllable than for any other, resulting, as desired, in an output representing primary stress. An analogous analysis can be made for Maranungku, whose weights are the mirror image of those for Weri.

For Lakota, if the "current input" is a monosyllable, the bias activation triggers primary stress. However, when the first syllable of a non-monosyllable is the current input, the negative weights w_{+1} override the bias activation. If the current input is the second syllable of a word, the perceptron receives high positive activation from w_{-1} in addition to the bias; this is sufficient to overcome the negative weights w_{+1}. However, any syllable after the second triggers the strong inhibitory contribution of w_{-2}, and so cannot receive stress. The analysis for Polish is very similar.

The connection weights indicate systematic encoding of knowledge of the patterns by the networks. The patterns of weights for stress patterns which are mirror images of each other are themselves mirror images.

5.2. Computational analysis of learnability: QI systems

The learning times differ considerably for {Latvian, French}, {Maranungku, Weri}, {Lakota, Polish} and Garawa, as shown in the last column of Table 3[20]. Moreover, Paiute and Warao were unlearnable with this model. (They *were* learnable using either a two-layer architecture with two output units, or a three-layer architecture. We present the finding of non-learnability within the present model in order to maintain a consistency of analysis in one model. As noted previously, the two-layer model with a single output unit facilitates analysis of connection weights and their contribution to processing).

Examination of the inherent features of these stress patterns suggests various factors as being relevant to learning. The references in Table 3 are to pattern descriptions in Table 1, which it may be helpful to consult as needed.

Alternation of stresses (as opposed to a single stress) is suggested by the difference between learning times for {Latvian, French} and {Maranungku, Weri}, which also suggest that the *number of stress levels* may be relevant.

Recall from Table 1 that, in Garawa, primary stress is placed on the first syllable, secondary stress on the penultimate

[20] These are learning times with the *syllabic* input representation. We use these learning times as the basis of our discussion because they provide a more sensitive measure than learning times with *weight-string* representations. These latter representations were used in the discussion in Section 4 since QI and QS patterns were being examined together. Note that the ordering of learning times for languages is the same with both representations (Table 2, Table 3).

syllable, and tertiary stress on alternate syllables preceding the penultimate, but that no stress appears on the second syllable. The primary, secondary and tertiary stress patterns potentially lead to stress appearing on both the first and the second syllables; however, this is avoided (stress is never placed on the second syllable). This exemplifies the tendency in human languages to avoid the appearance of stress on adjacent syllables. The greater learning time for Garawa suggests that such *stress clash avoidance* is computationally expensive.

In languages such as Latvian, French, Maranungku, Weri and Garawa, primary stress is always on a syllable at the edge of the word. In Lakota and Polish, whose learning times are substantially greater than those of the other languages, primary stress is always at a non-edge syllable, except in mono- and di-syllables. (Paiute and Warao are identical, with respect to the placement of primary stress, to Lakota and Polish, respectively, but are unlearnable.) Thus, *placement of primary stress* seems computationally relevant. In particular, it appears more difficult to learn patterns in which primary stress is assigned at the edges inconsistently.

To explore these indications more fully, and to determine what features of Paiute and Warao led to their non-learnability, a number of hypothetical stress patterns were examined. These stress patterns are described in Table 4.

The following factors emerged as determinants of learnability for the range of QI patterns considered:

1. **Inconsistent Primary Stress (IPS):** it is computationally expensive to learn the pattern if neither edge receives primary stress except in mono- and di-syllables; this can be regarded as an index of computational complexity that takes the values $\{0, 1\}$: 1 if an edge receives primary stress inconsistently, and 0, otherwise.

2. **Stress clash avoidance (SCA):** if the components of a stress pattern can potentially lead to *stress clash*, then the language may either actually permit such stress clash, or it may avoid it. This index takes the values $\{0, 1\}$: 0 if stress clash is permitted, and 1 if stress clash is avoided.

3. **Alternation (Alt):** an index of learnability with value 0 if there is no alternation, and value 1 if there is. Alternation means a pattern of some kind that repeats on alternate syllables.

4. **Multiple Primary Stresses (MPS):** has value 0 if there is exactly one primary stress, and value 1 if there is more then one primary stress. It has been assumed that a repeating pattern of primary stresses will be on alternate, rather than adjacent syllables. Thus, [Alternation=0] implies [MPS=0]. Some of the hypothetical stress patterns examined here (see below) include ones with more than one primary stress; however, as far as is known, no actually occurring QI stress pattern has more than one primary stress.

5. **Multiple Stress Levels (MSL):** has value 0 if there is a single level of stress (primary stress only), and value 1 otherwise.

It is possible to order these factors with respect to each other to form a five-digit binary string characterizing the ease/difficulty of learning. That is, the computational complexity of learning a stress pattern can be characterized as a 5-bit binary number whose bits represent the five factors above, in decreasing order of significance. Table 3 shows that this characterization captures the learning times of the QI patterns quite accurately. As an example of how to read Table 3, note that Garawa takes longer to learn than Latvian (165 vs. 17 epochs). This is reflected in the parameter setting for Garawa, "01101", being lexicographically greater than that for Latvian, "00000".

The analysis of learnability is summarized in Table 5 for all the QI stress patterns, both actual and hypothetical. It can be seen that the 5-bit characterization fits the learning times of various actual and hypothetical patterns reasonably well; there are, however, exceptions, indicating that this 5-bit characterization is only a heuristic. For example, the hypothetical stress patterns with reference numbers h21 through h25 have a higher 5-bit characterization than other stress patterns, but lower learning times.

The effect of **stress clash avoidance** is seen in consistent learning time differentials between stress patterns of complexity less than or greater than binary "1000". Learning times with complexity "001" are in the range 10 to 25 epochs, while complexity "1001" patterns are of the order of 170 epochs; complexity "010" is of the order of 30 epochs, and "1010", 190 epochs, for patterns the only difference between which is the absence/presence of stress clash avoidance (Latvian2edge and Latvian2edge-SCA, references h5 and h19 respectively). A pattern with complexity "011" (Latvian2edge2stress, reference h6) has a learning time of 37 epochs, while a pattern differing only in the addition of SCA (Latvian2edge2stress-SCA, reference h20) takes 206 epochs. Complexity "101" patterns are in the range 30 to 60 epochs, while complexity "1101" patterns are in the range 70 to 170 epochs; in particular, while Garawa (reference L5) has a learning time of 165 epochs, the same pattern without SCA has a learning time of 38 epochs (Garawa-SC, refernce h10). A stress pattern of complexity "111" takes 85 epochs to learn (Latvian2edge2stress-1alt, reference

IPS	SCA	Alt	MPS	MSL	QI LANGUAGES	REF	EPOCHS (syllabic)
0	0	0	0	0	Latvian French	L1 L2	17 16
0	0	1	0	1	Maranungku Weri	L3 L4	37 34
0	1	1	0	1	Garawa	L5	165
1	0	0	0	0	Lakota Polish	L6 L7	255 254
1	0	1	0	1	Paiute Warao	L8 L9	** **

Table 3: Preliminary analysis of learning times for QI stress systems, using the *syllabic* input representation. IPS=Inconsistent Primary Stress; SCA=Stress Clash Avoidance; Alt=Alternation; MPS=Multiple Primary Stresses; MSL=Multiple Stress Levels. References L1-L9 refer to Table 1.

REF	LANGUAGE	DESCRIPTION OF STRESS PATTERN
h1	Latvian2stress	Main stress on first syllable, secondary on second
h2	Latvian3stress	Main stress on first, secondary on second, tertiary on third syllable
h3	French2stress	Main stress on final, secondary on antepenult
h4	French3stress	Main stress on final, secondary on penult, tertiary on antepenult
h5	Latvian2edge	Main stress on first and last syllables
h6	Latvian2edge2stress	Main stress on first and last, secondary on antepenult
h7	Maranungku3stress	Main stress on first, secondary on penult, alternate preceding tertiary and secondary stresses
h8	Weri3stress	Main stress on last, secondary on antepenult, alternate preceding tertiary and secondary stress
h9	Latvian2edge2stress-alt	Main stress on first, secondary on penult and alternate preceding syllables
h10	Garawa-SC	Main stress on first, secondary on penult, tertiary on alternate preceding syllables
h11	Garawa2stress-SC	Main stress on first, secondary on penult and alternate preceding syllables
h12	Maranungku1stress	Main stress on first and alternate succeeding syllables
h13	Weri1stress	Main stress on last and alternate preceding syllables
h14	Latvian2edge-alt	Main stress on first and last and alternate preceding syllables
h15	Garawa1stress-SC	Main stress on first, and penult and alternate preceding syllables
h16	Latvian2edge2stress-1alt	Main stress on first, and antepenult and alternate preceding syllables, secondary on final
h17	Garawa-non-alt	Main stress on first, secondary on penult, tertiary on ante-antepenult, no stress on second
h18	Latvian3stress2edge-SCA	Main stress on first, secondary on last, tertiary on antepenult, no stress on second
h19	Latvian2edge-SCA	Main stress on first and last but no stress on second
h20	Latvian2edge2stress-SCA	Main stress on first and last, secondary on antepenult, no stress on second
h21	Garawa2stress	Main stress on first, secondary on penult and alternate preceding syllables, no stress on second
h22	Latvian2edge2stress-alt-SCA	Main stress on first, secondary on last and alternate preceding syllables, no stress on second
h23	Garawa1stress	Main stress on first, and penult and alternate preceding syllables, but no stress on second
h24	Latvian2edge-alt-SCA	Main stress on first, and last and alternate preceding syllables, no stress on second
h25	Latvian2edge2stress-1alt-SCA	Main stress on first, and antepenult and alt preceding syllables, secondary on last, but no stress on second
h26	Lakota2stress	Main stress on second, secondary on penult
h27	Lakota2edge	Main stress on second and penult syllables
h28	Lakota2edge2stress	Main stress on second and penult, secondary on fourth syllable
h29	Lakota-alt	Main stress on second and alternate succeeding syllables, but not on last
h30	Lakota2stress-alt	Main stress on second and penult, secondary on fourth and alternate succeeding syllables

Table 4: Descriptions of hypothetical QI stress patterns.

IPS	SCA	Alt	MPS	MSL	LANGUAGE	REF	EPOCHS (syllabic)
		0	0	0	Latvian	*L1*	17
					French	*L2*	16
		0	0	1	Latvian2stress	h1	21
					Latvian3stress	h2	11
					French2stress	h3	23
					French3stress	h4	14
		0	1	0	Latvian2edge	h5	30
		0	1	1	Latvian2edge2stress	h6	37
		1	0	0	*impossible*		
		1	0	1	Maranungku	*L3*	37
					Weri	*L4*	34
					Maranungku3stress	h7	43
					Weri3stress	h8	41
					Latvian2edge2stress-alt	h9	58
					Garawa-SC	h10	38
					Garawa2stress-SC	h11	50
		1	1	0	Maranungku1stress	h12	61
					Weri1stress	h13	65
					Latvian2edge-alt	h14	78
					Garawa1stress-SC	h15	88
		1	1	1	Latvian2edge2stress-1alt	h16	85
	1	0	0	0	*impossible*		
	1	0	0	1	Garawa-non-alt	h17	164
					Latvian3stress2edge-SCA	h18	163
	1	0	1	0	Latvian2edge-SCA	h19	194
	1	0	1	1	Latvian2edge2stress-SCA	h20	206
	1	1	0	1	Garawa	*L5*	165
					Garawa2stress	h21	71
					Latvian2edge2stress-alt-SCA	h22	91
	1	1	1	0	Garawa1stress	h23	121
					Latvian2edge-alt-SCA	h24	126
	1	1	1	1	Latvian2edge2stress-1alt-SCA	h25	129
1		0	0	0	Lakota	*L6*	255
					Polish	*L7*	254
1		0	0	1	Lakota2stress	h26	**
1		0	1	0	Lakota2edge	h27	**
1		0	1	1	Lakota2edge2stress	h28	**
1		1	0	1	Paiute	*L8*	**
					Warao	*L9*	**
1		1	1	0	Lakota-alt	h29	**
1		1	1	1	Lakota2stress-alt	h30	**

Table 5: Analysis of Quantity-Insensitive learning using the *syllabic* input representation. IPS=Inconsistent Primary Stress; SCA=Stress Clash Avoidance; Alt=Alternation; MPS=Multiple Primary Stresses; MSL=Multiple Stress Levels. References L1-L9 index into Table 1, and h1-h30 into Table 4.

h16), while addition of stress clash avoidance results in a learning time of 129 epochs (Latvian2edge2stress-1alt-SCA, reference h25).

The effect of **alternation** is seen in a contrast between learning times for patterns of complexity "001" (range: 10 to 25 epochs) and "101" (range: 30 to 60 epochs); "010" and "110" (30 epochs vs. a range of 60 to 90 epochs); "011" and "111" (37 vs. 85 epochs).

The effect of **multiple primary stresses** is seen in the contrast between the stress patterns: Latvian2stress (reference h1, complexity "001", 21 epochs) and Latvian2edge2stress (reference h6, complexity "011", 37 epochs); Latvian2edge2stress-alt (reference h9, complexity "101", 58 epochs) and Latvian2edge2stress-1alt (reference h16, complexity "111", 85 epochs).

The effect of **inconsistent primary stress** is considerable: stress patterns with the most significant bit 1 are learnable in the perceptron model only if all the other bits are 0; such patterns (Lakota, Polish, references L6, L7, complexity "10000", 255 epochs) have a higher learning time than any of the patterns with most significant bit 0. All the examined stress patterns of complexity greater than "10000" were unlearnable in the perceptron model. Recall that Paiute and Warao were unlearnable; the present framework is consistent with that result, since under the present analysis, these two patterns have complexity "10101".

The impact of **multiple stress levels** is relatively smaller, and less uniform, both of which motivate this factor's being treated as least significant. Thus though there are several instances where a stress pattern with a greater number of stress levels has a higher learning time (h1 vs. L1; h3 vs. L2; h6 vs. h5; h7 vs. L3; h8 vs. L4; h21 vs. L5), there are also cases in which a stress pattern with a higher number of stress levels has a lower learning time than one with fewer stress levels (h2 vs. h1; h4 vs. h3; h11 vs. h15; h21 vs. h23).

The effects of a particular factor seem to be reduced when a higher-order bit has a non-zero value. Thus, the effects of alternation are less clear when there is stress clash avoidance: without SCA, the range of learning times for patterns without alternation is 10 to 40 epochs, and with alternation 30 to 90 epochs; but with SCA, the range without alternation is 160 to 210 epochs, and with alternation 70 to 170 epochs.

In summary, the "complexity measure" suggested here appears to identify a number of factors relevant to the learnability of QI stress patterns within a minimal two-layer connectionist architecture. It also assesses their relative impacts. The analysis is undoubtedly a simplification, but it does provide a framework within which to relate the various learning results.

5.3. Computational analysis of learnability: QS systems

For QS patterns, information about syllable weight needs to be included in the input representation – the input has to consist of (encoded) sequences of "H" and "L" tokens. A purely *syllabic* input representation is, by definition of quantity-sensitivity, inadequate. *Weight-string* representations were therefore adopted, as discussed in Section 3.2. To maintain consistency of analysis across QI and QS stress patterns, simulations for the QI languages were re-run using the weight-string representation[21]. Note that the stress patterns for all possible weight strings of length n are the same for a QI language.

Connection weights for the learned patterns were shown in Figure 4. Interpretation of the displays is as discussed in Section 5.1. Additionally, however, let the upper and lower rows of the input buffer be designated L and H (light and heavy), respectively. Thus, the large negative weights mentioned earlier in the display for Lakota (Figure 4) in position -2 are designated w_{-2L} and w_{-2H}. With the weight-string representation for inputs, a light syllable is represented by a [1, 0] vector, and a heavy syllable by a [0,1] vector. For light syllables, therefore, there will be a 1 in the top row (row L) and a 0 in the bottom row (row H); for heavy syllables, the reverse. Thus, with the weight-string representation the contents of the two rows of the input buffer are usually not identical, and this is relevant to understanding how the connection weights encode knowledge of stress patterns.

As in the discussion of QI systems in Section 5.1, the following analysis of connection weights for QS systems is meant to provide background understanding for the analysis of factors relevant to learnability.

For Koya, the bias weight supplies a fairly high positive activation; there is also high positive activation when a heavy syllable is the "current input," arising from w_{0H}. If the current input is the first syllable, then the large negative weights w_{-1} have no effect, and the bias activation results in an output denoting primary stress. If the current input is not the first syllable, then w_{-1} produce a large negative input, whether the syllable in position -1 is heavy or light, thus offsetting positive activation from the bias connection; the net input will be low, resulting in a low output denoting zero

[21] These are the learning times shown in Table 2. The correspondence of these learning times with linguistic markedness has already been discussed, in Section 4.2.

stress, *unless* the current input is a heavy syllable, in which case the large positive weight w_{0H} contributes substantially. This positive activation plus that of the bias unit together produce a greater positive net input than is offset by the negative activation from w_{-1}, and so the output is medium, representing secondary stress. In other words, the weights encode the stress pattern: stress the first syllable, and assign secondary stress to heavy syllables.

In Malayalam, the current syllable is stressed if it is the first syllable and either it is heavy (large positive activation from w_{0H}, and the large negative weight w_{-1H} has no effect), or it is light but the second syllable is also light (in which case the negative weight w_{+1H} will have no effect). If the current syllable is the first, but is light, and the second syllable is heavy, then w_{0H} will provide no stress, and additionally, w_{+1H} will damp stress provided by the bias connection. If the current input is the second syllable, it receives stress only if it is heavy (positive activation from w_{0H}) and the previous syllable was light (no negative activation from w_{-1H}.) No syllable other than the first or second will be stressed because two of the four large negative weights in w_{-1} and w_{-2} will always be triggered. The analysis for Yapese is similar.

For Ossetic, if the current input is the first syllable, it receives stress only if it is heavy (positive activation from w_{0H}.) If the current input is the second syllable, it receives stress only if the previous syllable was light (no negative activation from w_{-1H}). All syllables other than the first and second encounter negative activation from either w_{-2L} or w_{-2H}, and so never receive stress.

The connection weights for Komi and Cheremis are interesting in that they establish a means of "scanning" the buffer. Recall the stress pattern of Komi: stress the first heavy syllable, or the last syllable if there are no heavy syllables. If the current syllable is heavy, it should be assigned primary stress only if there have been no preceding heavy syllables. A heavy current syllable receives stress from w_{0H} and from the bias term, and this is sufficient to offset the effect of the negative weights w_{+1H} and w_{+1L}; but if there is a heavy syllable to its left, this stress is overridden by the weights w_{-1H} through w_{-6H}. Thus a heavy syllable will be stressed *iff* it is the first heavy syllable.

If the current input is light, it should be assigned primary stress only if it is the last syllable, *and* there have been no heavy syllables in the word. The connection weights make no provision for positive activation from any buffer position containing a light syllable. When a light syllable is the current input, therefore, positive activation comes only from the bias unit; this positive activation, however, is offset by negative activation arising from w_{+1}, and by negative activation arising from w_{-1H} through w_{-6H}. The positive bias is not outweighed by negative activation just in case there are no syllables succeeding the current input in the buffer, and also no heavy syllables preceding the current input, i.e., just in case the current input is the last syllable in a word without any heavy syllables.

The weights w_{-1H} through w_{-6H} thus produce, in parallel, the effect of "scanning" that portion of the buffer that contains syllables that "precede" the current one; this scanning is necessary to determine the appropriate assignment of stress both to heavy and light syllables. The analysis of weights for Cheremis is analogous to that for Komi.

The learnability analysis proposed in Section 5.2 on the basis of QI patterns requires some refinement. Inconsistent Primary Stress (IPS) was hypothesized as taking binary values; a 1 value for IPS was used to indicate that primary stress was assigned inconsistently at the edge of words; a 0 value indicated that this was not the case. If this measure is modified so that its value indicates *the proportion of cases in which primary stress is not assigned at the edge of a word*, the learning results for both QI and QS patterns can be integrated, to a large extent, into a unified account. Learning times in the following discussion are those shown in Table 2.

The learning times for Malayalam and Yapese are approximately 20 epochs, while those for Ossetic and Rotuman are approximately 30 epochs. The difference between these pairs of stress patterns is: for Malayalam and Yapese, primary stress is placed at the edge *except* when the edge vowel is short and the next vowel long (i.e., except 0.25 of the time); for Ossetic and Rotuman, primary stress falls at the edge *except* when the edge vowel is short, i.e., except in 0.5 of the cases.

The five factors discussed earlier were: Inconsistent Primary Stress (IPS); Stress Clash Avoidance (SCA); Alternation (Alt); Multiple Primary Stresses (MPS); and Multiple Stress Levels (MSL). The values of these indices respectively, for both Malayalam and Yapese, are [.25 0 0 1 0], and for both Ossetic and Rotuman, [.5 0 0 1 0]. The difference between learning times for these pairs of otherwise identical patterns can then be accounted for in terms of differing values of the IPS measure.

Refinement of the IPS measure thus seems warranted. Note that the earlier analysis of QI languages remains unchanged: stress patterns that had an IPS value of 0 still do, and those that had an IPS value of 1 still do as well.

The learning times of Komi and Cheremis are substantially higher than those of Koya, Eskimo, Malayalam, Yapese, Ossetic and Rotuman. As discussed above, for Komi and Cheremis, the networks in effect simulate "scanning" of the input buffer, which requires a greater number of connection weights to reach significant magnitude. It seems reasonable to hypothesize that this requirement is computationally expensive, i.e., that the learning time for Komi and Cheremis is higher because it takes longer to establish multiple weights of the proper magnitude. The connection weight displays of Figure 4 illustrate the fact that none of the other QS stress patterns require establishment of more than two or three weights of large magnitude; for Komi and Cheremis, by contrast, there is a string of large weights across the buffer.

For Komi, a particular syllable S receives primary stress under the following conditions: (1) There are no heavy syllables to the left of S, in the syllable string; and (2) S is Heavy *or* S is the last syllable. The second clause of the conditional involves *single-positional* information: information either about the syllable S itself (S is Heavy), or about the absence/presence of a syllable right-adjacent to S in the weight-string. (If there is no syllable to the right of S in the weight-string, then S is the last syllable; if there is a syllable right-adjacent to S, then S is not the last syllable). The first clause of the conditional, however, involves *aggregative* information: information about *all* the syllables to the left of S in the weight-string. The hypothetical "scanning" referred to above provides precisely this *aggregative* information; and similarly for Cheremis.

Komi and Cheremis can therefore be analyzed as stress patterns that require *aggregative* information for the determination of stress placement; none of the other stress patterns require such information. For example, for Koya, a syllable S should receive stress if it is the first syllable (which can be determined from information about the presence/absence of a syllable in the left-adjacent weight-string position), or if it is heavy, both of which are *single-positional* kinds of information. For Ossetic, a syllable S should be stressed if (a) it is the first syllable AND it is heavy (which requires *single-positional* information about the left-adjacent weight-string position, and about S itself); or if (b) it is the second syllable (*single-positional* information about the weight-string element two positions to the left of S) AND the syllable in left-adjacent position is light (also *single-positional* information).

The difference in learning times between Komi and Cheremis on the one hand, and Koya, Eskimo, Malayalam, Yapese, Ossetic and Rotuman, on the other, can now be analyzed in terms of the differing informational requirements; as has been seen, *aggregative* information requires the building of a series of weights of proper magnitude across the buffer, and this requires greater learning times. Whether or not *aggregative* information is needed therefore seems to be a further factor relevant to the learnability of stress patterns.

The patterns of Mongolian and Mayan have very much higher learning times than those of any other stress patterns, including Komi and Cheremis. For Mongolian, if the current input is heavy, then it should receive stress if it is the first heavy syllable; thus, as for Komi, each of the weights w_{-1H} through w_{-6H} must be capable of damping the positive activation from w_{0H}. If the current syllable is light, then it should receive stress only if (a) there is no syllable to its left in the buffer (if there is, then one of w_{-1H} or w_{-1L} will override the bias activation), AND (b) there is no heavy syllable to its right in the buffer. Note that this requires a set of weights w_{+1H} through w_{+6H} to the right of the current input, to determine whether there is a heavy syllable. Thus, for Mongolian, there is *aggregative* information required about heavy syllables both to the left of the current input *and* to its right. This seems to be what makes the pattern so difficult to learn. (As a matter of fact, there is a kind of compounding of aggregative requirements: the weight w_{0H} must be large enough to overcome *all* of w_{+1H} through w_{+6H}, and so must be rather large; but also, each of w_{-1H} through w_{-6H} must be able to override w_{0H}, and so each of these must be even larger. Thus, several very large weights are needed, as evidenced by the magnitude of the largest weights for Mongolian: 28.04, as against a range of approximately 9.0 to 11.0 for the other QS patterns.)

The results from Komi, Cheremis, Mongolian and Mayan thus suggest an additional factor that is relevant for determination of learnability, but that comes into play only in the case of QS patterns: whether or not *aggregative* information is required. This can be treated as a sixth index of computational complexity, that can take the values $\{0, 1, 2\}$. We therefore have the following factor, in addition to the five previously discussed:

6. Aggregative Information (Agg): has value 0 if no aggregative information is required (*single-positional* information suffices); value 1 if one kind of aggregative information is required (Komi, Cheremis); and 2 if two kinds of aggregative information are required (Mongolian, Mayan).

5.4. Unified analysis: learnability of QI & QS systems

With these modifications (viz., refinement of the IPS measure, and addition of the Aggregative measure), the same parameter scheme can be used for both the QI and QS language classes, with good learnability predictions *within* each class, as shown in Table 6. Note that in this table, learning times for all languages are reported in terms of the weight-string representation (as in Table 2) rather than the unweighted syllabic representation used for the initial QI studies (Section 5.2; Table 5). The differences in learning times across QI patterns are less marked than the differentials in Table 5, which summarized QI learning results with the syllabic representation. This is the result of the increased

Agg	IPS	SCA	Alt	MPS	MSL	QI LANGUAGES	REF	EPOCHS (wt-string)	QS LANGUAGES	REF	EPOCHS (wt-string)
0	0	0	0	0	0	Latvian French	L1 L2	2 2			
0	0	0	0	0	1				Koya	L10	2
0	0	0	0	1	0				Eskimo	L11	3
0	0	0	1	0	1	Maranungku Weri	L3 L4	3 3			
0	0	1	1	0	1	Garawa	L5	7			
0	0.25	0	0	0	0				Malayalam Yapese	L12 L13	19 19
0	0.50	0	0	0	0				Ossetic Rotuman	L14 L15	30 29
0	1	0	0	0	0	Lakota Polish	L6 L7	10 10			
0	1	0	1	0	1	Paiute Warao	L8 L9	** **			
1	0	0	0	0	0				Komi Cheremis	L16 L17	216 212
2	0	0	0	0	0				Mongolian Mayan	L18 L19	2306 2298

Table 6: Summary of results and analysis of QI and QS learning (using *weight-string* input representations). Agg=Aggregative Information; IPS=Inconsistent Primary Stress; SCA=Stress Clash Avoidance; Alt=Alternation; MPS=Multiple Primary Stresses; MSL=Multiple Stress Levels. References index into Table 1.

training set size with the weight-string representation as compared with the syllabic representation. However, learning times are consistent with the overall learnability analysis, so that the analysis of QI systems is in no way affected.

As can be seen, differences in learning times between QS stress patterns also fit in with the analysis developed earlier, and with the analysis of *single-positional* vs. *aggregative* informational requirements developed in this section.

Thus, both the QI and QS results fall into a single analysis within this generalized parameter scheme and weight-string representation, but with a less perfect fit than the within-class results. Notice also that our characterizations of stress systems is generally consistent with their characterization in terms of the constructs of metrical theory.

The QI stress patterns of Lakota and Polish have higher complexity indexes than the QS stress patterns of Malayalam, Yapese, Ossetic and Rotuman, but lower learning times. Quantity-sensitivity thus appears to affect learning times, as seems reasonable to expect, due to the distribution of the weight-string training set (see discussion in Section 3.2). However, no "measure" of its effect will be offered here. The analytical framework developed thus far appears to hold within QI languages, and within QS languages; further analysis would be needed to relate learning results across the two kinds of stress patterns.

6. Connectionist techniques and linguistic theory

6.1. Description vs. explanation: Paiute and Warao

The preceding discussion has developed a theoretical framework within which to characterize various stress systems and their learnability, in terms of both their inherent and computational characteristics. We have also seen that this scheme is consistent with the metrical-theoretic characterizations of the various systems. We thus view our analysis as similar in spirit to linguistic theory development. But in addition, our analysis has been strongly grounded in computational issues.

For example, the discussion in Section 5.2 indicated how the parameter scheme reflects the non-learnability of Paiute and Warao: their 5-bit characterizations were greater than "10000". However, this is still a relatively abstract level of theorizing/description. We now undertake a more rigorous analysis of those non-learnability results.

The two stress patterns *were* learnable for training sets containing words of up to four syllables (a "length-4" training set); the learning time was 54 epochs. With the addition of five-syllable words, however, (a "length-5" training set), no solution could be found by the perceptron model. The connection weights established for Paiute with the "length-4" training set are displayed in Figure 5a. Figure 5b is a schematic illustration of the same weights. It

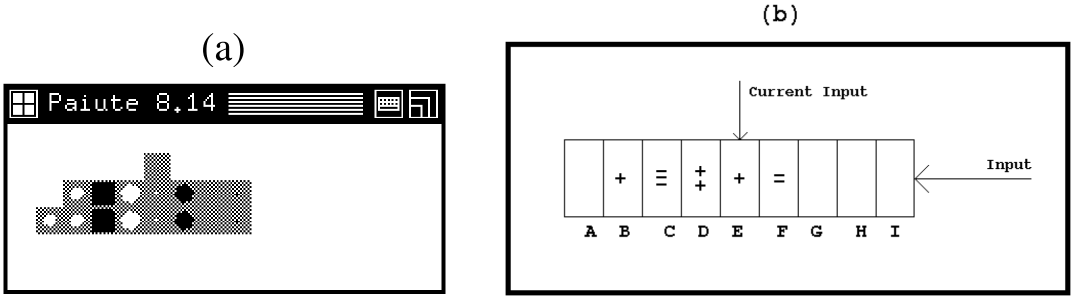

Figure 5: (a) Connection weights for the Quantity-Insensitive stress pattern of Paiute, learned for words of up to four syllables. (b) A schematic depiction of those weights in a buffer capable of processing words of up to five syllables. *E* has a modest positive weight to the output unit, and *D* a large positive weight. *F* has a large negative weight, and *C* a very large negative weight.

abstracts away from the weight-string input scheme, which is non-essential for a QI language; the figure also abstracts away the bias connection. The buffer positions *A, B, C, D, E* and *F* in Figure 5b correspond respectively to the positions w_{-4}, w_{-3}, w_{-2}, w_{-1}, w_0 and w_{+1} in Figure 5a. However, 5b depicts a larger buffer than is shown in 5a.

When the first syllable of a four-syllable word is the current input, the large negative weight *F* offsets the positive effect of *E*, resulting in a low net input (and corresponding zero stress). When the current input is the second syllable, the appropriate output is primary stress, which is ensured by the positive activation from *D* and *E* combined, which is greater than the negative activation from *F*. When the third syllable is the current input, the negative weights *C* and *F* combined offset the positive activation of *D* and *E* combined, so that the output corresponds to zero stress. For the fourth syllable, the positive weights *B*, *D*, and *E* combined are sufficiently greater than the negative weight *C* to yield an output corresponding to secondary stress, but not so much greater than *C* as to produce an output corresponding to primary stress.

The processing of the first three syllables of a *five*-syllable word is identical to that just described for a four-syllable word: the weights involved are *C, D, E* and *F*. When the fourth syllable of a *four*-syllable word is the current input, the word is spread over positions *B, C, D* and *E*; the negative weight *F* plays no role. However, when the fourth syllable of a *five*-syllable word is the current input, the five-syllable word is spread over positions *B, C, D, E* and *F*. The output should correspond to secondary stress, just as in the case of the fourth syllable of the four-syllable word, for which the weights *B, C, D* and *E* were appropriate, as described above. For the five-syllable word, however, output is affected not only by those four weights, but also by the weight *F*. For *B, C, D* and *E* to produce the appropriate output in the five-syllable case, *F* would have to be zero; however, for appropriate output on the first syllable of words of length greater than one, *F* must have the negative value shown. Thus there is a conflict in the requirements for *F* for the correct processing of “length-4” and “length-5” training sets.

This is shown more formally below[22] in an argument similar to the now famous proof of non-computability of XOR [Minsky 69]. Let θ_1 be the minimum net input required for a response corresponding to primary stress; let θ_2 be the minimum activation required for secondary stress. The following constraints must then hold:

$$
\begin{array}{lll}
(1) & & E > \theta_1 \\
(2) & & F < -E \\
(3) & & E + F < \theta_2 \\
(4) & & \theta_2 > E + F > \theta_1 + F \\
 & \Rightarrow & \theta_2 - \theta_1 > F \\
(5) & & \theta_1 > (B + C + D + E) > \theta_2 \\
 & \Rightarrow & -\theta_1 < -(B + C + D + E) < -\theta_2 \\
(6) & & \theta_1 > (B + C + D + E + F) > \theta_2 \\
 & \Rightarrow & \theta_2 < (B + C + D + E + F) < \theta_1 \\
(7) & & \theta_2 - \theta_1 < F < \theta_1 - \theta_2
\end{array}
$$

[22] We thank Geoffrey Hinton for suggesting this approach.

Inequality (1) expresses the constraint necessary for E to be able to produce primary stress for a monosyllable. (2) and (3) express the constraints necessary for E and F jointly to be able to suppress both primary and secondary stress for the first syllable of a word of length greater than one. (4) is derived from (3) as shown, and establishes an upper bound for the magnitude of F. Inequality (5) indicates the constraint that must be satisfied for B, C, D and E to produce secondary stress when the current input is the final syllable of a four-syllable word. So far, the constraints are all as required for correct assignment of stress in the "length-4" training set, and all the inequalities can be satisfied. The additional constraint needed to assign secondary stress to the fourth syllable of a *five*-syllable word is indicated in (6). This is the constraint that makes the "length-5" training set unlearnable. Summing (5) and (6) yields (7), which includes the condition $\theta_2 - \theta_1 < F$. However, we previously have (4) $\theta_2 - \theta_1 > F$. Thus, (4) and (7) impose contradictory constraints on the value of F, as was discussed above. The contradiction is responsible for the non-learnability of Paiute. An analogous demonstration can be made for Warao.

The important point here is that, while the previous analysis of Paiute and Warao in terms of Inconsistent Primary Stress is *descriptively* accurate at a gross level (i.e., languages of complexity greater than "10000" are unlearnable) based on observable properties of these and other stress patterns, the *explanation* just given of their non-learnability can be made only in terms of the interactions of properties of that stress pattern with rather specific properties of the architecture employed in the perceptron model.

6.2. Perceptron learning and human learning

There are significant similarities between the development of our six-parameter high-level analysis of perceptron stress learning and the development of the linguistic theory of human stress. Both approaches attempt to identify salient inherent characteristics of the stress systems under examination. Both approaches are concerned with the *learnability* of the stress systems by the device under study (perceptrons or humans.) Our approach, developed in Section 5, has in general been founded upon much greater knowledge of and access to the actual computational processes involved in perceptron learning than possible with investigations of human stress assignment. In the previous section, however, we showed that even a computationally informed analytical scheme may be no more than a descriptive approximation of lower-level processing.

This raises the question of how "absolute" we should regard high-level analyses as being. In particular, it raises this question about theoretical linguistic analyses. Without the ability to open up the black boxes of the human processor, linguistic analyses are, arguably, analogous to our high-level descriptions. We suggest that this highlights the need for theoretical analyses to be grounded in a consideration of computational issues.

The computational analysis developed here demonstrates that connectionist techniques enable detailed investigation of both the inherent nature and the computational characteristics of stress patterns. This, together with the correspondences between our perceptron learning results and metrical theory described in Section 4, indicate that connectionist techniques have the potential to provide the kind of computational grounding we think is required. And the difference between high-level and low-level accounts indicates that such grounding is necessary to achieve *explanatory* rather than merely *descriptive* power.

6.3. Problems with linguistic theory

We have argued the need for computational grounding and validation of linguistic theory, and suggested that connectionist techniques can provide this. We now turn to a consideration of some methodological problems with the dominant form of linguistic theorizing about stress systems. We suggest that in view of these problems, it is all the more important for investigations of language to seek new methodologies and sources of converging evidence.

6.3.1. Principles and parameters again

In the *principles and parameters* approach, an element or rule of linguistic analysis is taken to be part of an innate endowment (a "principle of Universal Grammar") if it is found to be applicable across languages, or to be so abstract that a language learner could not reasonably or logically be expected to learn it from exposure to linguistic data [Hyams 86, p. 2]. The hypothesis is that the human language faculty is so structured as to make only certain linguistic structures available to human beings. The recurrence of these limited patterns of linguistic structure in the world's languages is taken to be a reflection of the properties of the language faculty.

Language learning is taken to proceed through the discovery of appropriate parameter settings. For this approach to succeed, the relevant linguistic theory must be cast into the "parameter" mold clearly enough to specify (a) what the parameters are taken to be, and (b) what the possible settings of these parameters are, so that a given linguistic system X can be characterized in terms of parameter values P_x. The most explicitly formulated such scheme of which we are aware is that of Dresher & Kaye, discussed in Section 2.4, and adopted here as the theoretical framework.

It is clear that there are numerous stress systems that Dresher & Kaye's scheme as presently formulated cannot describe. For example, the stress patterns of Garawa and Aklan do not seem amenable to characterization in terms of these parameters. Hayes' description of Garawa ([Hayes 80, pp. 54-55]) involves three levels of stress, and his analysis involves the construction of binary feet both at the left edge of a word, and iteratively, starting at the right edge of the word. The combination of these operations has no analogue in the parameterized characterization adopted by Dresher & Kaye, whose discussion of Garawa sidesteps this difficulty by simplifying the pattern to just two levels of stress. To take another example, the stress pattern of Aklan is well-known for its complexity; the analysis given in [Hayes 80, pp. 20-33, page 59] includes conditions that cannot be expressed purely in terms of Dresher & Kaye's parameter scheme, and those authors do not discuss this pattern[23].

For an account of the learning of parameter settings P_x to provide an account of learning of the actual stress system X, a necessary condition is, obviously, that P_x should be an accurate characterization of the actual data of X. It therefore seems worth examining the original data for the variety of stress patterns examined in this paper, to determine whether these data really fit neat descriptions such as "stress the penultimate syllable and alternate preceding syllables". If they do not, then arguments about the learnability of P_x retain only a tenuous link with the learnability of the actual (messy) data of X. Moreover, these stress patterns have been central in the development of metrical theory, that is, in devising the theory's inventory of constructs. The theory therefore provides a secure descriptive formalism only in so far as the details of these stress patterns have actually been taken into account rather than glossed over.

6.3.2. Data on stress systems

We have examined three stress systems for which we could obtain source data easily: Ossetic, Koya and West Greenlandic Eskimo.

Ossetic. In Ossetic, described in [Abaev 64, pp. 10-11], (1) stress falls only on the first or second syllable of the word or word-group, (2) if there is a strong vowel[24] in the first syllable, then stress falls, with some rare exceptions, on the first syllable, (3) if there is a weak vowel in the first syllable, then stress falls on the second syllable. Morpho-syntactic factors can affect the above regularities.

This fits with the characterization of stress taken from [Hayes 80, pp. 62-63], in which primary stress is described as falling on the first syllable if it is heavy, and on the second syllable, otherwise. The description does not take into account the rare exceptions mentioned by Abaev, nor the extra-phonological determinants of stress, but this seems a reasonable simplification. Hayes cites Abaev (*op. cit.*) as the source of data.

Koya. In Koya, according to Tyler ([Tyler 69, pp. 32-33]), stress within a word occurs with long syllables, with weak stress on short syllables. Strong stress also occurs on the final syllable under certain circumstances conditioned by the intonation contour. Stress within a phrase occurs on the first syllable.

Hayes' description, adopted in this paper, and citing Tyler, is that primary stress falls on the first syllable, and secondary stress on closed syllables or syllables with a long vowel. This corresponds with Tyler's data under the assumption that phrasal stress (falling on the first syllable of the phrase, in Tyler's description) has been conflated with word-level stress. In that case, Tyler's description of stress within a word corresponds with Hayes' description of secondary stress within the word. Intonation-conditioned final-syllable stress also is ignored in Hayes' analysis.

West Greenlandic Eskimo. The data on stress in West Greenlandic Eskimo are less clear. Rischel ([Rischel 74, pp. 91-97]) states that the category of stress has no well-defined status in the language's phonology, and that it is very difficult to obtain agreement (from native speakers) on the stress patterns in a variety of word types. There is a strong tendency to hear stress on the last vowel, but Rischel suggests that this may actually be the effect of the intonational contour. Scholars have proposed various stress patterns. Thus Kleinschmidt ([Kleinschmidt 51]) suggests that the word has one main accent, and that longer words also have a subsidiary accent which tends to fall either on the first or

[23] We were able to simulate the learning of Aklan using a three-layer version of our model. For a different kind of connectionist treatment of Aklan stress assignment, see [Wheeler 91].

[24] "Strong vowel" refers to a *long* vowel, and "weak vowel" to a *short* vowel.

the last syllable. Very long words may have several subsidiary accents, which are distributed according to the principle that heavy syllables always attract the accent.

Fortescue ([Fortescue 84, p. 340]) states that there may be an auditory impression of relative stress on heavy syllables under the influence of certain intonational factors. He suggests that Kleinschmidt's account of stress can probably be reduced to the interaction of syllable weight with intonational nucleus. There may be some residual rhythmicity describable in terms of stress or pitch.

Hayes, citing [Schultz-Lorentzen 45] (which it has not been possible for us to examine), presents a much cleaner description of the stress pattern: stress syllables with branching rimes (i.e., closed syllables) and the final syllable ([Hayes 80, p. 58]). In view of the lack of agreement over Eskimo stress mentioned by both Rischel and Fortescue, this seems to be a significant simplification of the stress pattern.

Hayes ([Hayes 80, p. 58]) cites both Koya and West Greenlandic Eskimo as examples of languages that can be analyzed as having unbounded, quantity-sensitive feet, and thus as providing support for the very notion of "unbounded quantity-sensitve foot" and its inclusion in the inventory of metrical theoretic constructs. It has been seen that in both cases, the description that Hayes adopts is a simplification of the stress patterns. The validity of an analysis based on such simplified descriptions is an open question. Correspondingly, the implications for humans of a "learning" theory ultimately based on such tidied-up descriptions are unclear. While the connectionist model presented in this paper would produce different learning time results if presented with more realistic data, it is not at all clear how a learning system based on the parameters of metrical phonology would perform – for example, there may be no parameter set that can describe the actual data of Eskimo, in which case it is hard to see how a parameterized system could learn that data.

6.3.3. Difficulty of determining markedness

Within learnability theory, various proposals have been made regarding the "markedness" of particular grammars. One proposed metric is the number of intermediate grammars that have to be gone through in getting from an initial grammar G_0 to a descriptively adequate grammar G_L for language L ([Rouveret 80]). That is, the length of the sequence G_0, ..., G_L is a metric of its markedness. Similarly, Williams suggests taking the child's initial hypothesis about the language to be the "unmarked" case ([Williams 81]).

Under this view, the number of times the initial parameter settings have to be revised in arriving at the final parameter settings would indicate a language's complexity or markedness. However, as has been noted previously, the choice of initial or unmarked settings is related to the learning algorithm employed, and the nature of linguistic evidence assumed to be available in a particular model.

As an example of this, consider the notion of extrametricality. In Dresher & Kaye's formulation, the presence or absence of extrametricality is represented by parameter P8 (see Section 2.4). Dresher & Kaye implicitly take the default value for parameter P8 to be P8[Yes], meaning that there is extrametricality ([Dresher 90, p. 189,191]). This is because the presence of stress on a leftmost or rightmost syllable can rule out extrametricality; however, there is no positive cue that unambiguously determines the presence of extrametricality. If the default were P8[No], but this was an incorrect setting, there would be no cue that could lead to detection that this is incorrect. In contrast, in Nyberg's model ([Nyberg 90, Nyberg 92]), the default value of the same extrametricality parameter is taken to be P8[No], and the performance of his stress learning system indicates that the presence of extrametricality is harder to learn, or marked. Thus, we have two models, based on the same set of parameters[25], one showing that extrametricality is unmarked, and the other demonstrating that extrametricality is marked. Clearly, what is marked or unmarked is by no means an absolute, even within a parameter-based formulation.

It is therefore unclear how much this approach can contribute to determining the "markedness" of different linguistic systems. Obviously, it would be of interest to examine stages of development that children might go through in arriving at the stress pattern of their language. Work by Gerken suggests that children have "different" metrical feet ([Gerken 90]). However, as Dresher & Kaye note [Dresher 90, p. 42], there seems to be little or no data in the stress acquisition literature relevant to stages of development.

The question of how the learnability results from the perceptron simulations compare with those of metrical theory and with learnability in actual language acquisition is therefore not an easy one. As has been shown earlier, there are correspondences between the perceptron learning results and theoretical predictions, but there seems no way to "test" either the perceptron learning results or the theoretical linguistic predictions against human empirical observations. In fact, the lack of clarity in predictions of markedness from linguistic and learnability-theory arguments heightens

[25] Nyberg adopts the Dresher & Kaye scheme as well.

the potential significance of results from a learning model such as ours, which actually provides empirical learning results[26].

6.3.4. Impossible stress systems

As noted previously, the only linguistic options that can be entertained by the human mind are taken to be those consistent with the principles and parameters of Universal Grammar. Stress systems not sanctioned by the principles and parameters of metrical theory are therefore supposedly impossible. Thus Dresher & Kaye ([Dresher 90, pp. 148-151]) argue that one of the motivations for adopting parameters is that they greatly constrain the number of possible stress systems, and rule out crazy non-occurring stress systems that have never been observed. One of the criticisms frequently made of connectionist models of language processing is that they can "learn anything", and, in particular, can learn systems not sanctioned by linguistic theory.

The trouble is that, if this reasoning is carried through, and given Dresher & Kaye's parameters, the stress systems of Garawa and Aklan discussed above are impossible. Of course, it may be possible to extend the parameter scheme so as to describe these systems. But now, "possible" and "impossible" have been reduced to what has or has not been observed in the world. And, as discussed previously (Section 4.1), such distributional grounding must be viewed with caution. The point we wish to make here, then, is that the principles and parameters notion of "possibility" should be seen as having *heuristic* value rather than as providing definitive prescriptions of what is possible or impossible.

Certainly, the present perceptron model is unlikely to reflect the way that humans learn language, since it could not learn the stress systems of Paiute and Warao. Moreover, it would probably be capable of learning "outrageous" patterns quite easily. However, it is important to recognize that judging such a system to be plausible or implausible on these grounds is really an appeal to distributional evidence and intuition. A system of parameters does not provide the litmus of possibility that sometimes seems claimed for it.

A parameter-based learning model incorporates the assumptions of its underlying theory – it *implements* those notions of possibility and impossibility. But, given the theory-internal and/or distribution-based nature of parameter-based arguments for markedness and impossibility, it no longer seems clear that a parameter-based model really has much of an advantage over a model such as ours, in terms of "plausibility".

The only way to settle such questions would seem to be the ability or inability of the child to learn a particular stress system, irrespective of whether or not such a pattern is observed in the world's languages. However, as in the case of determining the *difficulty* of learning, there are no relevant data. There also seems little future possibility of obtaining evidence about the ability of children to learn "crazy non-existent" patterns. It appears that there is not much hope of sorting things out one way or another.

7. Summary

1. We have argued that the development of analytical frameworks needs to be grounded in investigation of computational processes and constraints.

    - We provided a novel characterization of 19 stress systems in terms of a set of six parameters that serve as both a partial description of the stress pattern itself and a prediction of its learnability, without invoking abstract theoretical constructs such as metrical feet. Our parameters encode linguistically salient concepts (e.g., *stress clash avoidance*) as well as concepts that have computational significance (*single-positional* vs. *aggregative* information). The same parameter scheme can be used for both the QI and QS language classes, with good learnability predictions within each class. Our ability to develop such an analytical framework indicates that connectionist learning techniques provide the tools for detailed investigation of the inherent and computational nature of the stress patterns.
    - Our parameter scheme can therefore be viewed as a formulation of the salient characteristics of these stress systems, in much the same way that linguistic analysis is such a characterization.
    - We developed our analyses by observation of environmental characteristics in the stress learning environment, much as theoretical linguistics does. We did, however, have access to the internal computational mechanisms involved in the process.

[26] Nyberg's parameter-based model also provides such learning results, in terms of *how many examples* of a stress pattern have to be presented to the stress learner [Nyberg 90].

- The form of our explanation was substantially different based on whether we took detailed account of processing mechanisms or not, as in the case of our two explanations of the non-learnability of the stress systems of Paiute and Warao (Section 6.1). Our high-level account of this non-learnability was that stress systems with a characterization lexicographically greater than "10000" are unlearnable, and this was supported by the non-learnability of the hypothetical languages h26 through h30. However, this account turned out to be an *approximation* to the low-level interactions of stress pattern with processing architecture.
- Without the ability to open up the black boxes of the human processor, linguistic analyses are, arguably, analogous to our high-level descriptions. This highlights the need for combining computational investigations with theory-building. We have demonstrated, for the domain of stress systems, that connectionist techniques provide the ability to examine computational mechanisms in considerable detail.

2. We have argued also that there are methodological problems underlying the parameter-based approach to learnability.

   - We examined some of the stress systems that have been foundational in developing the inventory of constructs in metrical phonology, and concluded that descriptions of some of these stress patterns significantly simplified the data. The implications for humans of a "learning" theory ultimately based on tidied-up descriptions are unclear.
   - The choice of initial or unmarked settings in a parameter-based scheme is related to the learning algorithm employed, and the nature of linguistic evidence assumed to be available.
   - The notion of "possibility" as defined by a system of parameters is based ultimately on appeal to distributional evidence and intuition, and should be seen as having *heuristic* value rather than as providing definitive prescriptions of what is possible or impossible.
   - Distribution-based arguments about markedness and impossibility need to be viewed with caution, as the world's distribution of linguistic forms certainly reflects extra-linguistic factors.
   - The principles and parameters approach to linguistic stress therefore does not provide as secure an analysis of learnability and markedness as it might at first glance appear to, and does not rest on a particularly solid empirical base.

3. In view of (a) the need for computational grounding for theoretical frameworks, and (b) methodological problems for parameter-based theories, it seems useful to seek techniques that can provide converging evidence to complement more conventional linguistic analysis.

   We have suggested, based on our simulations of stress learning, that connectionist techniques have the potential to provide such converging evidence and computational validation, and that they facilitate detailed investigation of computational mechanisms and processes.

   - Linguistic theory provides an abstract analysis of linguistic domains such as stress systems. This theory is cast in terms of certain kinds of constructs, which are viewed as being, at some level, processing primitives.
   - We have explored the ability of a simple perceptron model to learn a variety of stress systems that have been central in the development of the theory of linguistic stress, metrical phonology. This computational model did not incorporate the linguistic constructs.
   - Nevertheless, there turn out to be significant correspondences, first, between the structure of the network's internal states when it processes various stress patterns and the linguistic characterizations of those stress patterns, and second, between the learning times of the network and linguistic characterizations of markedness. Our simulations provide actual "learning" results that can be used for this kind of analysis. Moreover, we have shown how this source of evidence can contribute to theoretical analyses.
   - Our characterization of the stress systems in terms of a set of six parameters encodes linguistically salient concepts as well as concepts that have computational significance. The same parameter scheme can be used for both the QI and QS language classes, with good learnability predictions within each class. The predictions within each of these groups, moreover, are consistent with groupings of stress systems according to their linguistic analysis.

- All this suggests that linguistic analysis and a model such as ours are both picking out similar salient aspects of stress systems, and that theoretical linguistics and connectionist analysis might form complementary modes of enquiry for linguistic domains. That is, connectionist learning in conjunction with more traditional tools might provide the basis for a new methodology for the investigation of language.

8. A *gedanken* experiment

The Yapese Room

EXPERIMENTER: Good morning. We're ready to begin. All of you are interested in stress patterns in language, so we've arranged to have you analyze them in your own ways. Inside that room is a tape recorder, on which we will play you individual words of the language Yapese. Let's see what different analyses of the stress pattern you come up with. I must admit I'm curious to see how your analyses will reflect your various perspectives – Linguistics, Computer Science, and Computational Linguistics.

(A sequence of words is heard from inside the room)

LINGUIST: Ah, I think I see what's going on!

COMPSCI: Really? All I can tell is that stress is sometimes on the last syllable, and sometimes on the next to last.

LINGUIST: You're quite right. Stress falls on the final syllable except when the final vowel is short and the penult long, in which case stress falls on the penult.

COMPLING: Just what I was about to say myself; I agree with your analysis.

COMPSCI: Hmm, that's pretty neat. *(To the experimenter)* Do you have data from other languages?

(Many stress patterns later..)

COMPSCI: Well, now that we've all agreed on all the stress patterns, I'm off to develop a computational model of stress assignment.

COMPLING: Hey, wait! That's what I was going to do! How do you plan to go about it?

COMPSCI: It's pretty straightforward. For each pattern, I just need to put together a piece of code consisting of a couple of conditionals. I can read the syllables of the words into an array; for Yapese, for example, I need to write code to look at the last syllable in the array, and assign stress to it, unless it happens to be a short vowel, and the next-to-last syllable in the array is long, in which case I assign stress to the next-to-last element.

COMPLING: Well, that's rather naive, actually. You should be looking at the analysis that Linguist here has made. You need to incorporate those ideas, for your model to be anything other than a linguistically unsophisticated hack. Are you familiar with the *SPE* analysis?

COMPSCI: I'm not sure I know what you're talking about. Well, anyway, I'm off – I have pretty clear descriptions of the stress patterns, I think. Good luck with whatever it is you're going to put into your model.

(Some time later..)

COMPSCI: I'm back! My program's up and running. I call it GPS, for Generalized Processor for Stress.

COMPLING: My model's done, too. I call it SPE – for Stress Patterns Explained. What do you think, Linguist? Want to take a look at our models?

LINGUIST: (*Looking up from deep in thought*) What? You know, these stress patterns are really interesting. I've been thinking about them since you two went away, and I believe there's a much better linguistic analysis to be made. Better still, it provides a means of showing that these patterns differ from each other only along a small number of dimensions of variation, which can be regarded as parameters of the model. It's really quite an elegant framework. Let's call it metrical phonology. Here's how it works ...

(Some time later..)

COMPLING: That's a really neat theory. And I think I see a way to model stress assignment using it. I'll need to work out how the data interact with universal principles to establish parameter settings, but basically, your metrical stuff provides really clear representations for a computational model. I'll be back soon, with SPE-2.

COMPSCI: Er, actually, I think I'll stay with my original model. I don't see any need to change it. (*Backs out of the room.*)

(The next day)

EXPERIMENTER: And how are all of you today? I have someone here who'd like to meet you – the person who put together the speech synthesis system that produced the data you heard yesterday.

LINGUIST: Speech synthesizer? Didn't you tell us we were listening to tape recordings?

EXPERIMENTER: If you'll excuse me, I have to be going. Here's our good friend Connhacker.

CONNHACKER: I take it that our little speech system passed for human-like sound on tape? That's a great achievement for us, you know. We sincerely appreciate your co-operation yesterday.

COMPLING: Never mind all that. What about the stress patterns? Do you mean to tell me I've been developing computational models of some lousy speech synthesizer?

CONNHACKER: It's state-of-the-art, actually. For each language, we synthesize and put together syllables according to a stored list of word forms for that language; a stress contour is imposed by a connectionist network that we trained up specially.

COMPLING: I've been modeling a connectionist network! (*Turns to Linguist*) What does this do to your metrical theory? Down the tubes!

LINGUIST: No, actually the theory stands as it was – at least, if the stress contours were accurate. The theory merely provides a descriptive framework, within which to make sense of the data. Analyzing those little units in the connectionist network, or in the actual neural pathways of our brains for that matter, isn't going to give us an abstract understanding of the pattern; metrical theory can do that, and can also help uncover interesting relationships between patterns. It might even suggest constraints on the wiring of the human stress processing apparatus. But the primary benefit of the theory is in organizing the facts. If the speech synthesizer or connectionist network or whatever actually has replicated those stress facts accurately, then the metrical theory I devised is just as good as if I had devised it by listening to actual human speech.

COMPLING: But I've been using your theory to model the stress processing that I thought was being produced by humans. Obviously, it's totally inappropriate to use that theory if all that's in there is a connectionist network. There are no metrical trees or anything else going on in there that corresponds to your theory – just silly little linear threshold "units". Searle would never have stood for this!

LINGUIST: I never said there were metrical trees in the human brain; merely that there must be structures that provide a basis for the phenomena I characterized in terms of metrical trees. Your computational model is as appropriate or inappropriate as it would have been if you had modeled real speech produced by a real human with a real brain.

COMPSCI: Say, I don't know what you guys are on about, but I'd like to take a look at this speech synthesizer. (*Turns to Connhacker*) Can we?

CONNHACKER: Certainly. Since all of you seem particularly interested in the stress assignment aspects, let me show you some analyses we made of the connectionist network. While training it, we noted the learning times, and found we could pretty much predict how long it would take for the network to learn a particular kind of pattern on the basis of certain observable characteristics of the pattern. We could also tell which patterns would be learnable and which would require architectural modification of the network before they could be learned. Here are some of the factors we identified as relevant. First, whether there's alternation of stresses in the pattern. Second, whether there's just one, or more than one equal stress per word. Third, whether it's a pattern with stress clash. Fourth, ...

COMPSCI: This is really interesting. I'm pretty sure I can simulate these results: all I need is a binary string to characterize these factors you've identified, and then I'll be able to determine how long it should take to "learn" any given stress system. I can model the actual processing using Linguist's metrical tree formalism. Yep, I think I can model your connectionist network.

COMPLING: Anathema!

(Enter a Connectionist Magus, Connmag, looking enraged.)

CONNMAG: (*Boxing Connhacker's ears*) Imbecile! Is this what you've learned? I heard you spouting your spurious high-level analysis, with its "alternation" and "inconsistent primary stress" parameters. Bah! Don't you see that your "explanations" of the learning times of different stress patterns are only abstractions based on what's observable from the data? Do you see any correspondence between your analysis and the *computations* being carried out in the network? Can you prove that the Paiute stress pattern is unlearnable, based on your binary digit scheme? Of course you can't – to do that, you have to look at the computational structure of the model you've set up, and at the actual connection weights, in addition to features of the stress pattern itself. *That* analysis gives you an explanation in terms of the processing; everything else is just a correlation between a descriptive framework and observed regularities. True, your descriptive analysis helps to identify characteristics of those patterns that seem to correlate with their learnability, and thus helps to organize and make sense of the observed learning results. But don't you fall into the linguist's trap of thinking your *descriptions* are *explanations*. That way you end up inventing fallacious theories of the kind these people have been talking about. Who *are* all these people, anyway?

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.

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.

References

[Abaev 64] Abaev, V., *A grammatical sketch of Ossetic*, Indiana University Center in Anthropology, Folklore and Linguistics, Bloomington, Indiana, 1964.

[Berwick 85] Berwick, R., *The Acquisition of Syntactic Knowledge*, MIT Press, Cambridge, MA, 1985.

[Chomsky 68] Chomsky, N., & Halle, M., *The Sound Pattern of English*, Harper & Row, New York, 1968.

[Davis 88] Davis, S., *Topics in Syllable Geometry*, Garland Publishing, New York, 1988.

[Dresher 90] Dresher, B., & Kaye, J., A Computational Learning Model for Metrical Phonology, *Cognition* 34, 137-195.

[Everett 84] Everett, D., & Everett, K., On the Relevance of Syllable Onsets to Stress Placement, *Linguistic Inquiry* 15, 705-711.

[Fodor 88] Fodor, J., & Pylyshyn, Z., Connectionism and cognitive architecture: a critical analysis, in Steven Pinker & Jacques Mehler (eds.), *Connections and Symbols*, MIT Press, Cambridge, MA, 1988.

[Fortescue 84] Fortescue, M., *West Greenlandic*, Croom Helm, Beckenham, England, 1984.

[Gerken 90] Gerken, L., *Do adults and children have different feet?*, manuscript, State University of New York, Buffalo, NY, 1990.

[Goldsmith 90] Goldsmith, J., *Autosegmental and Metrical Phonology*, Basil Blackwell, Oxford, England, 1990.

[Halle 83] Halle, M., & Clements, G., *Problem Book in Phonology*, MIT Press, Cambridge, MA, 1983.

[Halle 87a] Halle, M., & Vergnaud, J.-R., Stress and the Cycle, *Linguistic Inquiry* 18, 45-84.

[Halle 87b] Halle, M., & Vergnaud, J.-R., *An Essay on Stress*, MIT Press, Cambridge, MA, 1987.

[Hayes 80] Hayes, B., *A Metrical Theory of Stress Rules*, doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1980. Circulated by the Indiana University Linguistics Club, 1981.

[Hayes 84a] Hayes, B., The Phonology of Rhythm in English, *Linguistic Inquiry* 14, 33-74.

[Hayes 84b] Hayes, B., Iambic and Trochaic Rhythm in Stress Rules, in *Proceedings of the Berkeley Linguistics Society* 11, Berkeley, CA, 1984.

[Hertz 91] Hertz, J., Krogh, A., & Palmer, R., *Introduction to the Theory of Neural Computation*, Addison-Wesley, Redwood City, CA, 1991.

[Hyams 86] Hyams, N., *Language acquisition and the theory of parameters*, Reidel, Dordrecht, Holland, 1986.

[Kaye 89] Kaye, J., *Phonology: A Cognitive View*, Lawrence Erlbaum, Hillsdale, NJ, 1989.

[Kleinschmidt 51] Kleinschmidt, S., *Grammatik der gronlandischen sprache*, 1851. (IDC Micro-Edition N-282).

[Liberman 75] Liberman, M., *The Intonational System of English*, unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1975. Circulated by the Indiana University Linguistics Club, 1978.

[Liberman 77] Liberman, M. & Prince, A., On stress and linguistic rhythm, *Linguistic Inquiry* 8, 249-336.

[McCarthy 79a] McCarthy, J., *Formal Problems in Semitic Phonology and Morphology*, unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1979.

[McCarthy 79b] McCarthy, J., On stress and syllabification, *Linguistic Inquiry* 10, 443-466.

[Minsky 69] Minsky, M. & Papert, S., *Perceptrons*, MIT Press, Cambridge, MA, 1969

[Nyberg 90] Nyberg, E., A Limited Non-Deterministic Parameter-Setting Model, *Proceedings of the North East Linguistics Society*, Université de Québec à Montréal, 1990.

[Nyberg 92] Nyberg, E., *Weight Propagation and Parameter Setting*, Ph.D. thesis, Dept. of Philosophy, Carnegie Mellon University, forthcoming.

[Prince 76] Prince, A., *Applying stress*, unpublished ms., University of Massachusetts, Amherst, 1976.

[Prince 83] Prince, A., Relating to the Grid, *Linguistic Inquiry* 14, 19-100.

[Pullum 82] Pullum, G., Letter, *Linguistics* 20, 339-344.

[Rischel 74] Rischel, J., *Topics in West Greenlandic Phonology*, Akademisk Forlag, Copenhagen, 1974.

[Rouveret 80] Rouveret, A., & Vergnaud, J-R., Specifying reference to subject, *Linguistic Inquiry* 11, 97-202.

[Rumelhart 86] Rumelhart, D., Hinton, G., & Williams, R, Learning Internal Representations by Error Propagation, in D. Rumelhart, J. McClelland & the PDP Research Group. *Parallel Distributed Processing, Volume 1: Foundations*, MIT Press, Cambridge, MA, 1986.

[Schultz-Lorentzen 45] Schultz-Lorentzen, *A grammar of the West Greenlandic language*, C. A. Reitzel, Copenhagen, 1945.

[Seidenberg 89] Seidenberg, M., Reading Complex Words, in G. N. Carlson & M. K. Tanenhaus (eds.), *Linguistic Structure in Language Processing*, Kluwer, Dordrecht, The Netherlands, 1989.

[Selkirk 80] Selkirk, E., The role of prosodic categories in English word stress, *Linguistic Inquiry* 11, 563-605.

[Selkirk 84] Selkirk, E., *Phonology and Syntax: The Relation between Sound and Structure*, MIT Press, Cambridge, MA, 1984.

[Trager 51] Trager, G. & Smith, H., *An Outline of English Structure*, American Council of Learned Societies, Washington, 1951.

[Tyler 69] Tyler, S., *Koya: An outline grammar*, University of California Publications in Linguistics, No. 54, University of California Press, Berkeley & Los Angeles, CA, 1969.

[van der Hulst 82] van der Hulst, H., & Smith, N., An overview of Autosegmental and Metrical Phonology, in H. van der Hulst & N. Smith (eds.), *The structure of phonological representations, Part I*, Foris Publications, Dordrecht, The Netherlands, 1982.

[Vergnaud 78] Vergnaud, J.-R., & Halle, M., *Metrical structures in phonology*, unpublished ms., Massachusetts Institute of Technology, Cambridge, MA, 1978.

[Wexler 87] Wexler, K. & Manzini, M., Parameters and Learnability in Binding Theory, in T. Roeper & E. Williams (eds.), *Parameters and Linguistic Theory*, Reidel, Dordrecht, The Netherlands, 1987.

[Wheeler 91] Wheeler, D. & Touretzky, D., From syllables to stress: a cognitively plausible model, in K. Deaton, M. Noske, and M. Ziolkowski (eds.), *CLS 26-II: Papers from the Parasession on The Syllable in Phonetics and Phonology, 1990.*, Chicago Linguistic Society, 1991.

[Widrow 60] Widrow, G. & Hoff, M. E., Adaptive Switching Circuits, *Institute of Radio Engineers, Western Electric Show and Convention, Convention Record, Part 4*, 96-104, 1960.

[Williams 81] Williams, E., Language acquisition, markedness and phrase structure, in S. Tavakolian (ed.), *Language acquisition and linguistic theory*, MIT Press, Cambridge, MA, 1981.