Stress Systems in Language: A Connectionist Examination

Prahlad Gupta

October 1990

Report No. CMU-LCL-90-5

Laboratory for Computational Linguistics

139 Baker Hall
Department of Philosophy
Carnegie Mellon University
Pittsburgh, PA 15213

Contents

Stress systems in language: a connectionist examination

Prahlad Gupta*

Abstract

Metrical phonology is a relatively successful linguistic theory that attempts to explain stress systems in language. This paper discusses a connectionist model that learns a variety of stress patterns without the incorporation, as processing primitives, of theoretical linguistic constructs such as *metrical foot* and *parameter*. An analysis of the learnability of various stress patterns is developed, based on learning results and connection weights developed for different stress systems. This analysis predicts that certain aspects of stress systems will be more difficult to learn, at least within the computational framework adopted. The model demonstrates an ability to generalize, and its encoding of knowledge of stress patterns indicates systematicity, with symmetries among stress patterns being reflected in the encoded knowledge.

1. Introduction

The work decribed here is an attempt to apply connectionist techniques to modeling the learning and processing of certain aspects of language. The domain of language chosen for exploration is *stress systems*; the reasons for this choice, as in the case of recent work by Dresher & Kaye ([Dresher 90]), are: (a) that the linguistic theory of stress (metrical phonology) is fairly well-developed, so that, compared with (for example) the domain of syntax, there is a relatively complete linguistic description of the observed phenomena; and (b) that stress systems can be studied relatively independently of other aspects of language ([Dresher 90, page 1]). Thus, stress systems were chosen as a relatively constrained and linguistically well-defined domain.

The focus has been on *computational* issues, such as the kinds of connectionist architecture that are minimally necessary for the task, and on analysis of the internal states of trained networks, to determine which aspects of those states contribute to the network's having learned the task. Additionally an attempt has been to examine the nature of the relationship of linguistic theory to language processing.

1.1. Linguistic structure in language processing

There is controversy over the relation of linguistic theory to the processing involved in the human use of language. For example, Treiman [Treiman 89] concludes that there is evidence for the psychological reality of syllable structure in human language processing, while Seidenberg [Seidenberg 89] draws a distinction between linguistic constructs as analytical tools, on the one hand, and the explicit representation and incorporation of these constructs into models of language processing, on the other. This suggests that the latter is not necessarily appropriate, and that results such

*I am grateful to Dave Touretzky for much helpful discussion and many valuable suggestions through the course of this work. Deirdre Wheeler provided extensive feedback throughout. I would also like to thank Brian MacWhinney, Jay McClelland, Eric Nyberg, Brad Pritchett and Steve Small for helpful comments.

as Treiman's are perhaps **better** viewed **as** epiphenomenal **effects** of **processing rather** than as evidence for any **explicit representation of syllable structure in processing. Carlson** fc Tanenhaus ([Carlson 89]) note that these **differing** perspectives reflect an underlying tension in the field of psycholinguistics. Seidenberg's views, moreover, are representative of the "connectionist" perspective: see. for example. [McClelland 86], which suggests that much of the structure perceived In-human cognitive processing may be epiphenomenal in nature, rather than explicitly present in processing representations. See also [Jackendoff 88] on differing views about the relevance of linguistic theory to models of language processing.

1.2. Modeling approaches

Those differing perspectives point to various possible approaches to language modeling.

One approach would take linguistic analyses and constructs as its starting point, and would base processing models on the manipulation, by structure-sensitive operations, of representations of such constructs. This would be. in Fodor <tc Pylyshyn's view [Fodor *HH].* the "classical" approach: it makes the assumption that explicitly structured representations are necessary in processing, and that those structures correspond to those suggested by linguistic analysis.

Another approach, exemplified by some "connectionist" models of language processing, takes as a guiding principle the requirement that the individual computing elements in a model be simple "neuron-like*' units, and seeks to simulate phenomena observed in language processing (typically a.ddrossod by the analyses of theoretical linguistics) without explicit incorporation of linguistic constructs into the models: the attempt here is to explain the observed linguistic phenomena as epiphenomena of parallel distributed processing. For example, in Elman's work [Elman Sf)]. 1 ho processing of sentences with embedded relative clauses (*The boy who chased tlu boy who kicked tlu cfit tripped)* involves no recursive syntactic structure explicitly manipulated or constructed by t ho processing at any given moment, but processing dynamics suggest recursive structure in time.

One way of viewing the relationship between these two approaches is in terms of (an interproiation of) David Marr's characterization of information-processing tasks ([Man* 82]). which proposes throe levels at which the task needs to be understood or analyzed: (1) the level of *cftmpulational theory,* at which an abstract functional description is provided of what is being computed: (2) tho lovel of *represf ntation find algorithm,* at which representations of the inputs and outputs of tho function characterized at the first level are specified, are algorithms that transform representations ol' inputs to representations of outputs: (3) the *implementation* level, at which tho instantiation of the computation in physical processing elements is considered.

In terms of this framework, the descriptions of language provided by linguistic analyses can be seen as belonging to the first **level;** they characterize **the output of the function being computed** in terms of constraints that apply to that output: this provides general constraints on the nature of strategies that may be employed in computing the function. For example, constraints on the structure of linguistic expressions are specified by syntactic analysis: and those constraints have implications for the computational mechanisms that may be involved: those mechanisms must provide a basis for the production of output conforming to the structural constraints.

Hot h classical and connectionist models (of. for example, language processing), are characterizations of the language information-processing task formulated at Man's second level, that of representation and algorithm. They differ in assumptions regarding the appropriate *nature* of representations and algorithms. Classical models assume the appropriateness of representations involving a compositional syntax, and of algorithms formulated in structure-sensitive forms to oporafo on those explicitly structured representations. Connectionist models assume representations that aro not explicitly structured, and algorithms specified in terms of primitive operations on primitive* computing elements, rather than in terms of operations on structural primitives. A further characterization that can be made of the two approaches is that classical models employ representations that explicitly incorporate the constraints suggested by level 1 analyses; representations in coniiectioiiist models are intended to incorporate to a greater extent the constraints suggested by neural science, which is concerned with issues relating to level 3. For example, classical models of language processing employ representations such as parse trees and semantic networks, which explicitly incorporate the notions of structure suggested by syntactic analyses of linguistic competence: connectionist models of language processing employ representations such as the activations of "layers of units", where the layers of units are taken to be analogues of populations of neurons. A third distinction is that, in connectionist modeling approaches, the issue of *learning is* intimately bound up with that of processing mechanisms and the construction of representations, whereas in classical approaches, there is less concern with the learning of representations, and. to the extent that learning is modelled, it is considered to involve processing mechanisms distinct from those that *apply* what has been learned. That is. in classical learning, there may be one set of computational operations involved in building up (i.e.. learning) knowledge structures: subsequently, cognitive processing is considered to involve the manipulation of those learned knowledge structure's by other operations, ones which may be distinct from those that were involved in learning.

1.3. Aims **and motivation of the present work**

Tin: rationale of the present work is that it should be more interesting and fruitful to compare "connect ionist" and "classical" approaches for a relatively well-defined domain of language (one* for which theoretical analyses provide relatively good coverage) than it would be for fragmentary aspects of a less well-defined linguistic domain. This motivated the choice of *stress systt nis* as t lie domain of exploration, rather than of more particular phenomena such as prepositional phrase attachment or particular inflectional paradigms.

Tin* aims an[1]: (a) to explore the ability of connectionist techniques to model the assignment of stress in individual words: and (b) to consider, in the light of this investigation, the assumptions and relationship of the connectionist and classical approaches—in particular, to examine the connect iouist view that the constructs of theoretical linguistics belong at the level *of description,* and that they need not form part of the inventory of computational primitives in a *prex-essing* model.

2. Background: stress systems in language

2.1. Evolution of linguistic theory

The linguistic analysis of stress systems has evolved through a number of phases: (I) Linear analyses presented stress as a phonemic feature of individual vowels, with different levels of stress represent in.u; different levels of absolute prominence: this was the approach taken in [Trager 51], and culminated in the analysis of stress presented in Chomsky *k* Halle's seminal *The Sound Pattern of English* [Chomsky f)N]. (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. **(Ill)** 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 *(nietrieal feet)-* so that stress assignment rules could make reference to them. (IV) The positing of internal structure for syllables ([Vergnaud 78]. [McCarthy 79a]. [McCarthy 70b]) 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 SO]) provided an account of many aspects of st ress systems in terms of a small number of *parameters.*

Throughout the development of metrical theory, there has been debate over whether Ihe auto-segmental representational structures 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]).

2.2. Syllable structure and stress

A syllable is analyzed as being composed of the *onset,* which contains the material before the vowel, and the *rime:* the rime is composed of the *nucleus,* which contains the vocalic material, and the *coda,* which contains any remaining (non-vocalic) material. (For further discussion, see [Kayo NO. pp. r>J-r>x]).

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 funnei- 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 00. page 113]). Languages that involve such a distinction (between heavy and light syllables, i.e.. between the *w(ighf* of syllables) are termed *quantity-sensitive,* and languages that do not. *quanfity-instnsilivt.* (Note* that, in quantity-insensitive languages, syllables can occur both with and without branching rimes: but Ihe distinction between these kinds of syllables has no relevance for the* placement of st ress).

2.3. Metrical theory

'There seems to be t heoret ical 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 *onsii.* Thus, for example, according to Dresher & Kaye. "*// is generally agreed thai ons< Is an not rthvanl to slrtss ruhs"* ([Dresher 90. page 3]: see also [Goldsmith 90. page 170]). Metrical theory thus takes rime structure as the basic level at which accounts of stress patterns can be formulated. (However, both [Davis NN] and [Everett 84] present evidence that onsets may in fact be relevant to the placement of stress).

The theory of metrical phonology further involves the notion that stress patterns are controlled by metrical structures built on top of rime structures. (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]). As discussed earlier, in one version of the theory, these metrical structures are *metrical tires,* whose construction can be characterized in terms of a number of *parameters.* One formulation of these parameters is as follows ([Dresher 90. p.4]):

(I) The word-tree is strong on the [Left/Right]
(2) I'eet are [Binary/Unbounded]
(3) Feet are built from the [Left/Right]
(4) Feet are strong on the [Left/Right]
(5) \>*W*\ are Quantity-Sensitive (QS) [Yes/No]
(0) Feel are QS to the [Rime/Nucleus]
(7) A strong branch of a foot must itself branch [No/Yes]

(8) **There is an extrametrical syllable [Yes/No]**
(0) It is **extrametrical on the [Left/Right]**
(10) A weak **foot is defooted in clash [No/Yes]**
(11) Feet are non-iterative [No/Yes]

As a result of this parametrized characterization, metrical theory is an exemplar of the *prinripU s and parameters* approach, of which a central hypothesis is that language learning proceeds through the learning of appropriate values *(settings)* for parameters. A particular set of parameter settings characterizes a particular possible human language: once the parameter settings are determined, the nature of structure-sensitive operations and the nature of the linguistic structures on which they operate is known, so that the structures and processes involved in language processing are automatically determined (at an abstract level). For example, if parameter (2) above is determined to be set to [Binary], parameter (3) to [Left] and parameter (4) to [Right], then this not only characterizes the language, but also means that the processing relevant to the imposition of stress contours on words is determined, and corresponds, in the abstract, to the assigning of stress to words by (among other things) construction of binary, left-headed, metrical trees, from left to right. *Learning* the stress pattern of the language involves determining (i.e.. learning) the appropriate parameter settings: subsequently, the assignment of stress in the actual *production* or *proassing* of language involves neural processes that correspond quite directly with the abstract application of these parameter settings as guidelines to the construction and manipulation of metrical feet.

2.4. A connectionist perspective

Thus, metrical theory: (a) constitutes a fairly well-developed system for specification of the regularities observed in the stress systems of human languages: (b) posits abstract representational structures and structure-sensitive operations on those structures as the basis of that specification system: (<•) provides an abstract account, in terms of characteristics of that specification system, of what it is that is learned in the acquisition of a stress system.

Froin the connectionist point of view, metrical theory provides one of the clearest available analyses of a relatively independent domain of language. Applied to this domain, one possible connectionist view would be (a) that the analysis provided by the theory is a Marrian level I analysis: (b) that the *parameters* of metrical phonology are not computational primitives that must necessarily be incorporated in a processing model of the learning of stress: and. (c) that the assignment of stress to words, once the stress pattern has been learned, does not necessarily involve *metrical trees/grids* **as computational/representational primitives.**

In this view, there is no necessity for processing models to directly incorporate constructs such as those of **theoretical linguistics. However, the plausibility of such a claim, and the value of such** an approach, can only be determined by investigation of whether or not connectionist models can in fact learn the correct assignment of stress, quite generally, for a variety of different stress patterns.

2.5. Other computational models of stress systems

Computational models of stress systems in language processing have been developed by Dresher <V Kaye ([Dresher 90]) and by Nyberg ([Nyberg 89]. The focus of both of those models is on the learning of the *jxirameters* 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 <t Kaye term a *learning theory,* a specification of how linguistic data the language learner encounters in its environment is to be used in order to set parameters. The following features can be said to characterize these two models: (1) they assume the existence of processes explicitly

corresponding to the linguistic notion of parameter setting; (2) the ***learning*** **theory they propose** as **an account of that parameter-setting process is couched in terms of a "classical^ model: (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 *ham ing ami production.* That is, as discussed in Section 1.2., learning involves mechanisms leading to the configuration of the various parameters of metrical theory: these then form a knowledge base for stress assignment, whose processing involves, for example, the construction of binary trees from right 1o left - an operation having no necessary correspondence with the operations by which I he parameters were learned.

The work reported here differs from the Dresher <k 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, in conjunction with the specific network architectures developed: (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-ccMined ionist emit rast s.

3. Description of the model

3.1. Scope

As in the case of work by Dresher & Kaye and Nyberg ([Dresher 90]. [Nyberg 89]). the scope of I he present work has been limited to consideration of the placement of stress in single words: moreover, again as in the other models, the effects of niorpho-syntactic information (such as lexical category of the word) on the placement of stress are ignored.

3.2. Assumptions

The following discussion applies to stress phenomena within the scope delineated above.

As mentioned earlier (Section 2.3.). it has been widely accepted that it is the material contained in **the** syllable *rime* **that is relevant to the placement of** stress, **with material in the syllable** *onset* **being irrelevant; more recently, however, there has been discussion of the role of the onset in stress** systems. **Here, the simplifying assumption is made that only information about the syllabic material** in rimes is relevant to stress systems: this follows **the** assumption **made in the computational** models developed by Dresher & Kaye ([Dresher 90. p.3]) and by Nyberg (personal communication).

As discussed in Section 2.2.. in *Quantity-Sensitive* (QS) languages, the assignment of stress is sensitive to the *weight* of syllables, while in *Quantity-Insensitive* (QI) stress systems, the weight of syllables is irrelevant. Words of a given *number* of syllables in length can differ in terms of tin* weights of the actual syllables that comprise them. In QI systems, the stress contour will be the same for all //-syllable words, since syllable weight is immaterial. In QS systems, the stress contour will *not* be the same for all //-syllable words, since the placement of stress is affected by syllable weight.

As also discussed in Section 2.2.. in QS systems, what most commonly counts as a *heavy* syllable is one with a branching rime, while a *Ught* syllable is one with a non-branching rime. However, it is possible for other properties of the syllable to be criteria! in determining weight. For example, in

Central Siberian Yupik, it is syllables with a long vowel that count as heavy; closed syllables with short vowels do not count as heavy, as they would in the more commonly-occurring heavy-light distinction ([Goldsmith 90, p. 179]).

In the parameter scheme for metrical phonology developed by Dresher fc Kaye ([Dresher 90, p.4]. summarized in Section 2.3. above), this distinction between types of quantity-sensitivity is handled by incorporation of the following parameter: (6) Feet are QS to the [Rime/Nucleus]. If parameter ((>) is set to [Rime], this amounts to specifying that the stress system has quantity-sensitivity involving a distinction between branching and non-branching rimes (or. equivalently. between closed and open syllables): thus a syllable with a branching rime is treated as *heavy*. and one without a branching rime is treated as *light*. Let this be designated "Type *A*" quantity-sensitivity. If (6) is set to [Nucleus], this means that what will count as heavy in the stress system is a long vowel, i.e.. a rime with a branching nucleus: this is exemplified by the quantity-sensitive system of Yupik, which may be designated "Type B". (A long vowel is analyzed as occupying two *(inn slots* or *motne.* which are dominated by the nucleus, while a short vowel occupies only one *mom:* thus a long vowel corresponds to a branching nucleus, and a short vowel to a non-branching nucleus).

The assumption implicit in this parameter scheme is that these are the only two kinds of quantity-sensitivity that occur: and this same assumption will be made here. In other words. the only features of a syllable that can be criterial in determining *weight* are (i) whether or not there* is a branching rime, and (ii) whether or not there is a branching nucleus.

In t he present work, the aim has been to draw upon linguistic analysis for insight regarding what information must be available in a *processing* model of stress learning/assignment. For quantity-sensitive stress systems of "Type A", information must be available about the presence/absence of syllable-final non-vocalic material: for quantity-sensitivity of "Type B". information must be available about whether vocalic material in the syllable is of one or more *worae* in duration.

(Jeneralizinq; over these informational requirements, it can be seen that the input to a stress learning/assignment system must encode the following information for each syllable: (a) whether the syllable has a coda (Yes/No): (b) whether the duration of vocalic material is one *ntom* or more (Single/Multiple). The possible values for these features set up a necessary four-way distinction between syllables, as follows: (I) ("Type 1") coda and multiple-mora vocalic: (2) ("Type 2") coda and single-mora vocalic: (3) ("Type 3") multiple-mora vocalic but no coda: (4) ("Type \") single-mora vocalic, and no coda.

Input that represents syllables categorized into one of the these four types encodes sufficient information for assignment of stress in any QI or QS language (within the scope of consideration adopted here). **In a QI language, the categorization will be immaterial - all syllables are equal; in** a "Type A'* QS language, syllables of types **(1) and** (2) **will count as heavy, and all others** as **light;** in a "Type B" QS language, syllables of types (1) and (3) will count as heavy, and all others as Itelii.

('onceptually. therefore, input to a stress learning/assigning system could be imagined as coming from ot her (unspecified) kinds of processing that syllabify the speech signal, and categorize syllables into the four above types: it seems reasonable to assume the existence of processing mechanisms that can make this four-way distinction.

In practice, in the model adopted here, a further simplification has been made: the input consists of syllables that are distinguished only in terms of whether they are *light* or *htavy.* Since what counts as *light* or *heavy* varies across stress patterns, this amounts to encoding some additional information into the input. However, encoding additional information seems reasonable in view of the fact that the kinds of representations that have been discussed here ("syllable", "heavy syllable", "light syllable") are in any case abstractions: if the input were not to be an abstraction

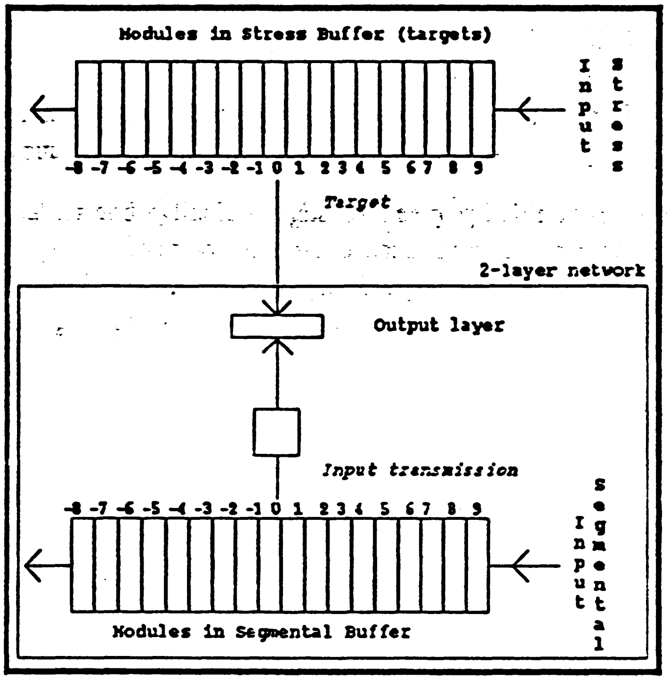

Figure 1: Architecture of the 2-layer model used in simulations

in any sense, it would have to be a speech signal.

3.3. Linguistic data

Descriptions of stress systems have been taken from [Hayes 80]. Nine Quantity-Insensitive and ten Quantity-Sensitive stress patterns were examined:

Quantity-Insensitive: Latvian, French, Maranungku, Weri, Garawa, Lakota, Polish, Southern Paiute and Warao.

Quantity-Sensitive: Koya, West Greenlandic Eskimo, Malayalam, Yapese, Ossetic, Rotuman, Eastern Permyak Komi, Eastern Cheremis, Khalkha Mongolian and Aguacatec Mayan.

3.4. Network architecture and processing

A "module" is a data structure representing a set of connectionist "units". A number of "modules" constitute the input layer of the network; each of these is connected through an array of connection weights to the output layer. The modules implement a buffer into which input is shifted from right to left; thus modules are used as elements in an input buffer which is the input layer of the network.

Figure 1 illustrates the architecture for a 2-layer network. Input representing the sequence of syllables comprising a word enters at the rightmost module (buffer element) and can be thought of as being transmitted through the buffer from each module to the module on its left. Thus if an input enters the rightmost module at time *t*, then at time *t+1*, this first module transmits to the second module, and also receives the next external input.

The center module (i.e., element of the buffer) is designated as a "monitored" element; transmission of inputs in the various modules in the buffer occurs only when the "monitored" module is "active", i.e., has some input in it. Whenever that is so, all the modules (i.e., buffer elements) that are active transmit to the output layer, to which they are connected by weighted connections. This buffering scheme amounts to a moving window.

In almost all simulations that will be described here, the output layer has consisted of a single unit. Inputs to the system are conceived of as being in two parallel streams, with the (representation of) the signal carrying information about the sound segments (syllables, really) "flowing" through the input buffer, as just described, and associated suprasegmental information (in this case, the stress associated with each syllabic unit) flowing through an identical, parallel buffer, synchronously with the syllabic-segmental information. Thus at any point, when a module in the input buffer contains a particular syllable, the corresponding module in the "suprasegmentar buffer contains the stress associated with that syllable. So. when the "monitored" element in the input buffer has input, and the contents of the buffer are therefore going to be transmitted, the module of the "suprasegmenta.r buffer corresponding to the monitored input module will contain the stress appropriate for the monitored input. The monitored module therefore contains what is treated as the "current input", and the buffer as a whole contains the whole word: at each point, the "monitored" module in the suprasegmental buffer contains the "target pattern" for the current input.

Training of the network proceeds in standard connectionist fashion: the input pattern consists of the whole word (as a stream of syllables), and the training target is the stress for that part of the input that is in the "monitored" module. Training proceeds by error correction, based on the difference between the output evoked at the output unit and what the target was: connection weights are updated appropriately.

A Her training to some criteria! point, a final pass over the input training set is made by way of "testing" of the network. Weights are not changed at this point: inputs are fed into the network, and outputs are observed and compared with targets.

Figun[1] 1 shows that the buffer is composed of modules: each module, however, represents a grouping of connectionist units, a subdivision that is not indicated in the figure. In the architectures used here, each module was taken to have two units, so that the set of modules comprising, say. the input buffer, can as well be thought of as a more typical array structure consisting of two rows of cells.

Two input representation schemes were used:

1. Syllabic information only: that is. an /?-syllable word is represented as a sequence of *n* identical tokens: the associated stress pattern is represented as a sequence of stress levels. For example, a seven-syllable word is represented as *[S S S S S S SJ*. and the associated stress pattern for a language with word-initial stress would be *[1 0 0 0 0 0 ()]*. Note that this ("syllabic") representation is insufficient for quantity-sensitive languages. Each *S* token is represented in the simulations as a *[I 1]* vector. Thus when the first syllable enters the input buffer, t he two units of the rightmost module (module 9, in Figure 1) receive inputs of "1".

2. Information about syllables and syllable weight. A heavy syllable is represented as *H*, and a light syllable as *L*. For example, a seven-syllable word consisting of alternate heavy and light syllables would be represented as *[H L H L H L H]* ("weight strings"). This information is minimally necessary for quantity-sensitive languages. An *H*token is encoded as the vector *[I 0]*, and an *L* token as the vector *[0 lj*.

In models with a single output unit, the training target for Primary Stress was a *[1]* vector, for Secondary stress a *[0.5]* vector, and for no stress a *[0]* vector. For stress systems involving three levels of stress, a slightly different scheme was used: a *[1]* vector for Primary stress, and a *[0]* vector for no stress, as before: but Secondary stress was represented by a *[0.6]* vector, and Tertiary stress by a *[0.-15]* vector.

Where two output units were used. Primary stress was represented as a *[1 1]* vector. Secondary stress as a *[1 0]* vector, and no stress as a *[0 0]* vector. Tertiary stress was represented as a *[0 I]* vector.

As a more detailed example of processing, suppose that the next word in the training set for some QI language is the four-syllable *[S S S 5/*, and the associated stress pattern (i.e., target), is *[0 0 0 1]* (indicating that the final syllable receives stress, and all other syllables are unstressed). The units in each module in both the input and target buffers ("segmentar and "stress" buffers, in Figure 1) at this point have zero activations. (A representation of) the first syllable enters the rightmost element (module 9) of the input buffer, and simultaneously, (a representation of) the target stress for that syllable (zero stress) enters module 9 of the target buffer. However, no forward propagation of activation takes place from the input buffer (layer) to the output layer, since the "monitored" module in the input layer (module 0) is inactive, i.e.. has zero activation in its units. This simulates one "time cycle"* of processing. On the next time cycle, the activations in module 9 of both the input and target buffers are transmitted, *imchanyeih* to module 8 of each buffer: the next syllable enters module 9 of the input buffer, and the associated stress enters module 0 of the target buffer. No forward propagation of activation will occur on this time cycle either, since module 0 of the input buffer is still inactive.

After two further time cycles, modules (3. 7. 8 and 9 of the input buffer contain the four syllables of the current word, and the corresponding modules of the target buffer contain the associated stress levels for those syllables. On the fifth time cycle, a vector of zeroes enters module 9 of the input buffer (indicating "no input"), and also module 9 of the target buffer. On the next four time cycles, the leftward flow of activations through the two buffers continues, still without any forward propagation from the input layer to the output layer, until, at the end of time cycle 9. the activations are in modules 1 through 4. (Modules 5 through 9 of both buffers contain zero activations, as a result of the leftward flow of the zero vectors that entered on time cycle 5). On time cycle 10. the activations move into modules 0 through 4 of the two buffers: the "monitored" module of the input buffer is for the first time "active", i.e.. its units have non-zero activation. On this cycle*, therefore, "transmission" occurs from the input buffer: all the input buffer modules that are active, propagate* their activations forward to the output layer. This is the forward propagation part, of the* connectionist processing. The target pattern for the output layer is whatever is in module* 0 of the* target buffer at this point. Computation of error, and the changing of connection weights between the input and output layers, occurs in accordance with whatever learning algorithm is being employed (in the simulations reported here, the back-propagation learning algorithm), if this is t he "training" phase of the simulation. In descriptions of processing in the rest of this paper, the eontenls e>f the "monitored" input buffer module at the time of a "transmission" from the* input bullèr will be referred to as the "current input".

On the next time cycle, the "current input" will become the second syllable of the four-syllable word, and forward propagation will occur again, as it will also on the next two time cycles (cycles 12 and 13). On time cycle 14, the leftward flow of activations through the buffers will leave module 0 in both buffers "inactive", and there will therefore be no transmission. At this point, both buffers are (IIISIHMI (activatiems of all units in all modules of the two buffers are set to zero), and one "trial" is e»ver. The next word in the training set can now enter, beginning the next trial.

Thus the processing of one word, syllable by syllable, in a number of "time cycles", constitutes e)ite» *trial.* One pass through all the words in the training set is one *epoch.*

For purposes of further discussion, the following abbreviations will be used for the various network architectures and input representations used in simulations:

- Network architectures:

  1. Two layers, one output unit, back-propagation algorithm *(N~t)*
  2. Two layers, two output units, back-propagation *(N-2)*

REF	LANGUAGE	DESCRIPTION OF STRESS PATTERN	EXAMPLE	EPOCHS
A1	Latvian	Fixed word-initial stress	51505050505050	17
A2	French	Fixed word-final stress	50505050505051	16
A3	Maraiiungku	Primary stress on initial syllable, secondary stresses on every other syllable thereafter	5I5U5L'50525O*g2*	37
A4	Weri	Main stress on final syllable, with secondary stresses on each alt. preceding syllable	***S'²5°S'²S°S'ᴶS⁽⁾S¹***	31
A5	Garawa	Main stress on initial syllable, secondary stress on penult, tertiary stress on alternating syllables preceding penult, but no stress on second syllable	<jl <^0v,'O <^'3 <^»0 <^2 <^»0	165
A6	Lakota	Fixed second stress	<^0 <^1 <^0 C,'O ^'0 ^'0 (^<)	255
AT	Polish	Fixed penultimate stress	^•0 ç«U ^'0 ^ 0 5 0 ^'1 5 0	254
A<s	Paiute	Main stress on second vowel, secondary stress on alternate succeeding vowels	***S° S¹ S° S-S° S-S°***	**
AD	Warao	Main stress normally on penult, with secondary stresses on alt. syllables before the main stress	***S⁽⁾S-S⁽⁾S-S^ij S^l S⁽⁾***	**

'Table I: Quantity-Insensitive patterns: description, example stress assignment, and learning; performance jn the *X-j/S* model.

- Input representations:

  1. Syllabic representation *(S)*
  2. Weight-string representation *(HL)*
  .'{. Kqualized weight-string representation *(HLE)*

'Thus a simulation employing a two-layer architecture with 2 output units, the back-propagation algorithm, and a weight-string representation will be abbreviated as *X-2/HL.*

Except where otherwise noted, the input set comprises words of up to seven syllables in length. for all the simulations discussed below.

4. Simulations: results and analysis

4.1. Quantity-Insensitive stress patterns

Nine Quantity-Insensitive (QI) languages were examined: Latvian. French. Maraiiungku. Weri. Lakota. Polish. Southern Paiute. Warao and Garawa.

The stress patterns of Latvian *k* French. Maraiiungku <fe Weri. Lakota <t Polish, and Sont hern Paiute <V* Warao are mirror images of each other. Table 1 summarizes the patterns for each language, and exemplifies the assignment of stress to a seven-syllable word.

Latvian <V* French: Latvian is a language with fixed word-initial stress, while French has fixed word-final stress.

Marannngku Ac Weri: In Maraiiungku. primary stress falls on the first syllable of a word, and secondary stress on alternate syllables succeeding the first: in Weri. primary stress is on 1 he last syllable* of a word, and secondary stress on alternate syllables preceding the last.

Lakota & Polish: Lakota has primary stress on the second syllable; Polish has primary stress on the penultimate syllable.

Southern Paiute & Warao: Southern Paiute assigns primary stress to the second syllable, and secondary stresses to alternate syllables succeeding the second. In Warao, primary stress falls on the penultimate syllable, and secondary stress on alternate syllables preceding the penult.

In Carawa primary stress falls on the first syllable, secondary stress on the penultimate syllable, and tertiary stress on alternate syllables preceding the penultimate. However, the second syllable never bears any stress. (The first syllable never bears any stress other than primary).

Carawa exemplifies the tendency in human languages to avoid the appearance of stress on adjacent syllables. The secondary and tertiary stress patterns, as stated above, would lead to the assignment of tertiary stress to the second syllable in all words of odd length greater than three, and to the assignment of secondary stress to the second syllable in words of length three. This would, however, lead to stress appearing on both the first and the second syllables, which is avoided.

An incidence of stress on adjacent syllables, as an anomaly among regularities, is. in linguistic analysis, termed *stress clash.*

4.1.1. Learning results

Using minimal connectionist architecture *(X-l)* and syllabic representation, seven of the nine QI stress patterns were learned with 100% accuracy. The last column of Table 1 shows the number of epochs of training necessary to achieve 100% accuracy for each stress pattern (an epoch is one presentation of the complete set of training data, in this case consisting of the seven possible words of lengths one through seven syllables). In most cases, the figures are an average over three simulation runs.

The learning time for each member of the pairs of mirror-image stress patterns is closely similar. Thus Latvian took 17 epochs and French 16: Maranungku 37 epochs and Weri 34: Lakota 255 epochs and Polish 254. Carawa took 165 epochs. Paiute and Warao were not learnable in the *X-l/S* model.

4.1.2. Connection weight displays

Connection weights for the seven learned languages are shown in Figure 2. The display for each language is a graphic depiction of the connection weights in the network when the language has been learned with 100% accuracy. These connection weights represent the "knowledge" of the stress pattern that the network has acquired: they both encode the network's "understanding" of the pattern and enable the network to perform correct processing of input representations.

Each display is a representation of the network as a whole. The **large** grey shaded rectangle at the base of each display represents the input buffer of the network, with its two layers of input units (see Section 3.4.). The single square protruding from the left of the input buffer is the *bias* unit to the single output unit, which is represented by the protruding square on top of the input buffer. A white blob in a particular position denotes a positive connection weight from the unit represented by that position to the output unit: thus a white blob in the bias unit position represents a positive connection weight from the bias unit to the output unit. A black blob denotes a negative connection (from the unit represented by the position in which the blob appears) to the output unit. The size (area) of the blobs is proportionate to the absolute magnitude of the positive/negative weight: the weight with the largest absolute magnitude is depicted in each display as a perfect square, and other weights are depicted by blobs of proportionate size. The scale (i.e.. absolute magnitude of the largest weight) is indicated by the number in the title bar of each display. Thus, for Lakota. the absolute magnitude of the largest weights is 4.02: these are the large (black) negative weights left of center in the input buffer.

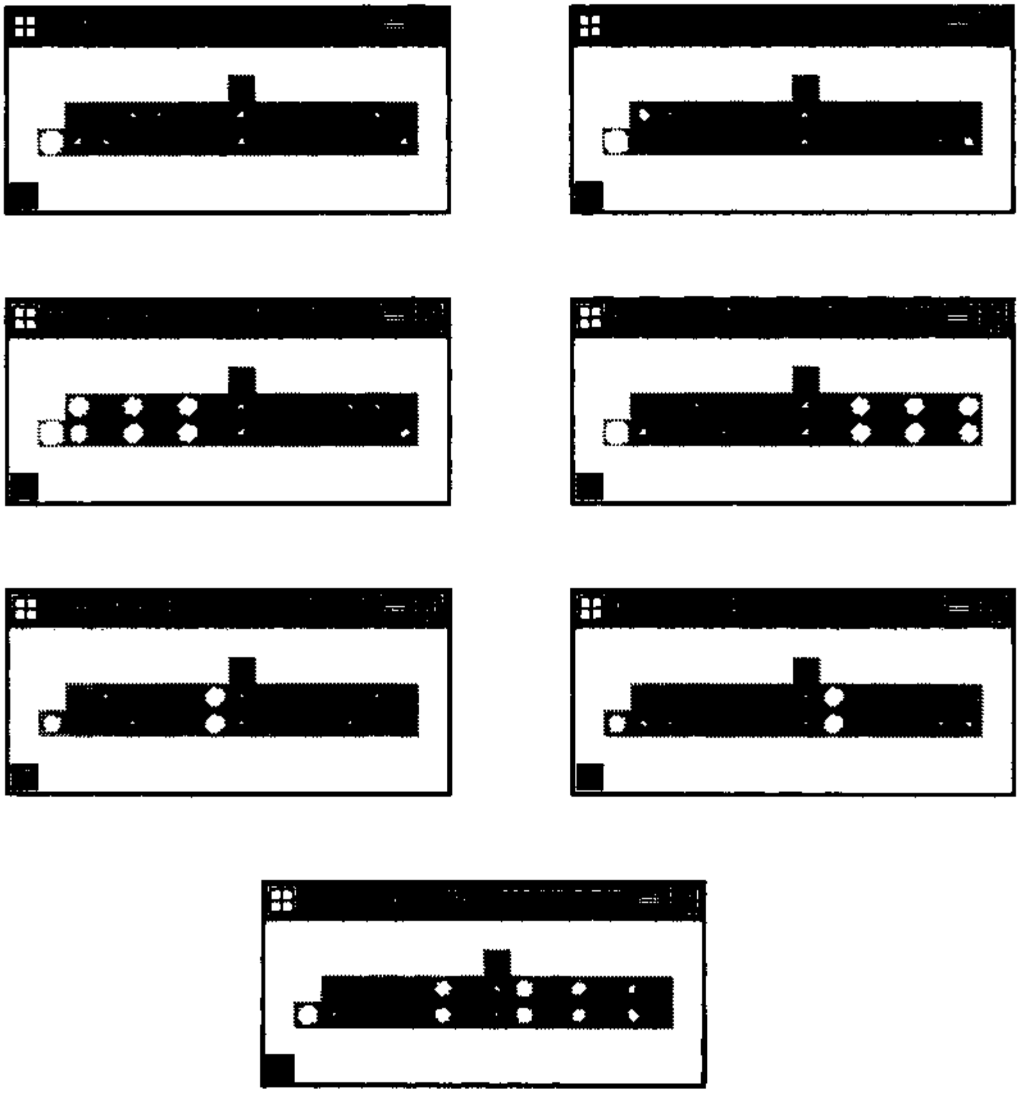

Figure 2: Connection weights for the Quantity-Insensitive stress patterns learned by the *N-I/S* model. Numbers in each title bar indicate the absolute magnitude of the largest weight: lionee *i\\r* scalr of the weights display.

Note that the weights displayed are those resulting from a single particular run, and not the average of weights over a number of runs. As can be seen from the figure, and as might be expected, the magnitude of connections weights differs across different stress patterns, but is the same for the members of each pair of mirror-image stress patterns. The magnitude seems proportional to training time, but does not in itself seem to correspond directly with any significant property of the particular pattern.

To enable precise description of individual weights, let the input buffer be regarded as a 2x13 array: 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 leftward from the center, and the columns to the right of the center be numbered positively 1 through 6. going right-ward from the center. (This numbering scheme is illustrated in Figure 1. which depicts a somewhat larger buffer size). Let the rows be /, and // (upper and lower, respectively). Thus, the large negative weights mentioned earlier in the display for Lakota are designated *w-2L* and *w-2H.*

As discussed in Section 3.4.. in the syllabic representation for inputs, a syllable is represented uniformly by a [1 1] vector. Since both elements of input vectors are 1. the two rows of the buffer have identical contents when using the syllabic representation.

4.1.3. Analysis of connection weights

l'or 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 t hose weights when it is the "'current input", i.e.. in position 0 in the buffer (see Section 3.4. 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. Activation to the output unit will be negative, since the negative magnitude of *w-1* is greater than the positive magnitude of the bias weight. The oiiipul of the network (appearing at the output unit) 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 *ir I* will have no effect. Activation to the output unit will therefore be positive (from the bias unit). and so the output of the network will be high, representing primary stress. l'or French, only the last syllable of a word will escape the effect of the large negative weights *w+L* 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 *ir+1;* these are large negative weights. Consider the processing of, say, a C-syllable word. When the leftmost syllable is the "current input" and the target output is therefore the output for zero stress, there will be two pairs of medium-strength positive weights *(ic+2* and *w+4)* and two pairs of medium-strength negative weights *(w+3* and *w+5.)* These four pairs, representing four syllabic elements in the buffer, roughly cancel each other out. There is also a pair of large negative weights (*ic-hl).* The net input to the output unit will therefore be negative, resulting in the output for 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+j.* However, the medium negative weights applicable are now only *w+3.* Therefore, the net input to the output unit is larger than it is for the previous syllable, and results in the desired 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 and positive weights applicable (resulting in zero stress) or one pair more of positive medium weights than negative medium weights, resulting in secondary stress. The exception is the last syllable: when this is the current input, there will be neither medium positive nor negative weights applied: but there will also not be the large negative weights *w+l.* Thus this is the only syllable that escapes the effect of these large negative weights. As a result, the net input to the output unit 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 unit triggers primary stress. However, when the first syllable of a polysyllable 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 output unit 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 high negative activation of *w-J*. and so cannot receive stress. The analysis for Polish is very similar.

In (iarawa. the first syllable receives stress from the bias unit. The second syllable triggers high negative activation from *w-L* and so is never stressed (the positive weights *ic-J* cannot provide positive activation for the second syllable). From the third syllable onwards, the presence or absence of secondary stress is determined by the alternating weights *w+1* through *w+6*. in conjunction with *w 2*.

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.

4.1.4. Learnability

The learning times differ considerably for {Latvian, French). {Maranungku. Weri). {Lakoia. Polish} and (iarawa. as shown in Table 1. Moreover. Paiute and Warao were unlearnable with this model.

Examination of the inherent features of these stress patterns suggests various factors as being relevant to the learning of stress patterns:

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

The greater learning time for Carawa suggests that 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 disyllables and monosyllables: Paiute and Warao are identical, with respect to the placement of primary stress, to Lakota and Polish, respectively, and 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 2. and the analysis of learnability is summarized in Table *:]* for all the QI stress patterns, both actual and hypothetical.

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

Primary Stress Placement **(PSP)** : It is computationally expensive to learn the pattern if neither edge receives primary stress except in monosyllables and disyllables. 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.

Stress Clash Avoidance (SCA) : If the components of a stress pattern can potentially lend to *stnss clash,* then the language may either actually permit such stress clash, or it may avoid

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 o
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 2: Descriptions of hypothetical stress patterns.

1 PSP	SCA	Alt	NPSJ	NSL	LANGUAGE	REF	EPOCHS
		0	0	0	Latvian	*Al*	17
					French	*A2*	16
		0	0	1	Latvian2stress	hi	21
					Latvian3stress	L2	11
					French2stress	h3	23
					French3stress	h4	14
		0	1	0	Latvian2edge	h5	30
		0	1	1	Latvian2edge2stress	U6	37
		1	0	0	*impossible*		
		1	0	1	Maranungku	*A3*	37
					Weri	•*U*	34
					Maranungku3stress	h7	43
					Weri3stress	h8	41
					Latvian2edge2stress-a.lt	h9	58
					Garawa-SC	hlO	3*
					Garawa2stress-SC	h11	50
		1	1	0	Maranungkulstress	hi 2	01
					Werilstress	hi 3	05
					Latvian2edge-alt	li 1-1	7X
					Garawa istress-SC	h15	ss
		1	1	1	Latvian2edge2stress-lalt	hlO	S5
	1	0	0	0	*impossible*		
	1	0	0	1	Garawa-non-alt	h17	MM
					Latvian3stress2edge-SC'A	his	103
	1	0	1	0	Latvian2edge-S(.'A	It 10	MM
	1	0	1	1	Latvian2edge2stress-SCA	h20	200
	1	1	0	1	Garawa	*A.j*	105
					Gara\va2stress	h21	71
					Latvian2edge2stress-al(-SCA	Ii22	91
	1	1	1	0	Garawa 1st ress	h23	121
					Latvian2edge-alt-SCA	h24	120
	1	1	1	1	Latvian2edge2stress-lalt-SCA	h25	120
1		0	0	0	Lakota	*A6*	255
					Polish	*A7*	254
1		0	0	1	Lakota2stress	h26	**
1		0	1	0	Lakota2edge	h27	**
l		0	1	1	Lakota2edge2stress	h28	**
l		1	0	1	Paiute	*AS*	**
					Warao	*A9*	**
l		1	1	0	Lakota-alt	h29	**
l		I	1	1	Lakota2stress-alt	h30	**

Tablo :$: Analysis of Quantity-Insensitive learning in the ;Y-7/S'model. PSP = Prhnary Stress Placement: S(.'A=Stress (lash Avoidance: Alt.=Alternation: NPS = No. of Primary Stresses: NSL=No. of Stress Levels. References index into Tables 1 and 2.

it. This index takes the values {0, 1}: 0 if stress clash is permitted, and 1 if stress clash is avoided.

Alternation (Alt) : An index of learnability has value 0 if there is no alternation, and value 1 if there is. Alternation denotes any pattern that repeats on alternate syllables.

Number **of primary stresses (NFS)** : The NPS value is 0 if there is exactly one primary stress. It is 1 if there are more then one primary stresses. It has been assumed that a repeating pattern of primary stresses will be on alternate rather than adjacent syllables. Thus, [Alternatioii=0] implies [NPS=0]. The hypothetical stress patterns examined here include ones with more than one primary stress: however, as far as is known, no actually occurring Ql stress pattern has more than one primary stress.

Number **of stress levels (NSL)** : The NSL value is 0 if there is a single level of stress (primary stress only): the value is 1 otherwise.

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 indicates that this complexity measure fits the learning times of various actual and hypothetical patterns rea.soua.bly well. There are. however, exceptions, indicating that this 5-bit characterization is only a heuristic. For example, the hypothetical stress patterns with reference numbers h2l through h25 have a higher 5-bit characterization than other stress patterns, but lower learning times.

SCA 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 arc* of the order of 170 epochs. Complexity "010" is of the order of 30 epochs, and "1010" is of the order of 190 epochs. (This latter contrast pair happens to represent patterns between which 1 lie presence/absence of stress clash avoidance (SCA) is the *only* difference. They are. respectively. Latvian'iedge (reference h5) and Latvian2edge-SCA (reference 19).) A pattern with complexity "ON" (Lat vian2edge2stress. reference h6) has a learning time of 37 epochs, while a pattern differing only in the addition of SCA (Latvian2edge2stress-SGA. reference h20) takes 200 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 A5) has a learning time of 1(55 epochs, the same pattern without SCA has a learning time of 38 epochs (Garawa-SC. refernce hlO). A stress pattern of complexity "111" takes 85 epochs to learn (Latvian2edge2stress-lalt, reference **h** 16), while addition of stress clash avoidance results in a learning time of 129 epochs (Latviaii2edge2stress-lalt-SCA. reference h25).

l)f Alt: The effect of **alternation** on learning times can be seen in the following three contrast pairs: Patterns of complexity "001" take 10-25 epochs, while those of "101" take 30-00 epochs: complexity of "010" takes 30 epochs while complexity of "110"takes 60-90 epochs: complexity of "Oil" takes 37 epochs, while "111" takes 80 epochs.

NPS The effect of the number **of primary stresses** is exemplified in the following two comparison pairs: Latvian2stress (reference hi. complexity "001". 21 epochs) and Latvian2edge2stress (reference h(i. complexity "011". 37 epochs): Latvian2edge2stress-alt (reference h9. complexity "101". 5S epochs) and Latvian2edge2stress-lalt (reference lil(3. complexity "111". 85 epochs).

PSP The effect of the **placement of primary stress** is considerable. Stress patterns whose most significant bit is 1 are learnable in the *X-l/S* model only if all the other bits are (). Such patterns (Lakota, Polish, references A6. A7, complexity "10000". requiring 255 epochs) have a higher learning time than any of the patterns whose most significant bit is 0. All the examined stress patterns **of** complexity greater **than "10000" were unlearnable in the *N-i/S* 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 a computational complexity** of greater than **"10000" (both have complexity of "10101").**

NSL The impact of the **number of stress levels is** relatively smaller and less uniform; both of these results make NSL the least significant factor. Thus, though there are several instances where a stress pattern with a greater number of stress levels has a higher learning time (111 vs. Al: h3 vs. A2: h6 vs. h5: h7 vs. A3; h8 vs. A4; h21 vs. A5). there also exit 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. hi: h4 vs. h3: hll vs. hl5: 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 is 10 to 40 epochs without alternation and 30 to 90 epochs with alternation. With SCA. the range is 160 to 210 epochs without alteration . and 70 to I 70 epochs with alteration.

In summary, the "complexity measure" suggested here appears to identify a number of factors relevant to the learnability of (QI) stress patterns within the minimal connectionist architecture (the *S-l/S* model): and it also assesses their relative impacts. The present analysis is undoubtedly a simplification, but it does provide some sort of framework within which to relate* the various learning results.

4.1.5. Paiute &; Warao

The Paiute and Warao stress patterns are interesting both because they serve to establish bounds on the computational capacity of the *X-l/S* model (with respect to stress patterns) and because they highlight the significance of input and output representations for computational behavior.

As already discussed, the stress patterns of Paiute and Warao proved unlearnable with **the*** A-*I/S* model (for up to 375.000 epochs, at which point training was terminated). The major problem with both of these patterns was that too low a stress was assigned to monosyllables, which was I he original pointer in the direction of the relevance of primary stress placement as a factor in learning complexity. In Paiute (Warao). primary stress falls on the first (last) syllable only in monosyllables, and in all other cases, the first (last) syllable is completely unstressed: evidently, it is difficult to adjust weights so as to deal with this exceptional stress assignment.

Modifying the stress patterns so that monosyllables are unstressed resulted in the patterns being **learned in approximately 30 epochs, thus confirming the hypothesis that exceptional monosyllabic** primary stress **was the problem. This suggested modifying the frequency distribution of training** patterns as a possible remedy, i.e., increasing the number of instances **of monosyllables in the** training sei. However, this did not enable learning. When seven additional instances of a monosyllable were included in the training pattern, stress was correctly assigned to monosyllables, but other errors occurred, reflecting the fact that the distribution of training data was skewed.

The t wo st ress patterns *were* learnable for training sets comprising words of up to only 1 syllables in length (a "length-4" training set); the learning time was 54 epochs. With the addition of 5-syliable words (a "length-5" training set), however, no solution could be found in the A-//V. *X-l/lIL* or A-*l/HLE* models. The connection weights established for Paiute with the "length—T training set are displayed in Figure 3a. Figure 3b is a schematic illustration of the same weights. It abstracts from the double buffer scheme, which is non-essential for a QI language (since t he only necessary representation is syllabic): the figure also abstracts from the bias unit, since the bias unit weights can always be re-distributed over other connection weights. The buffer positions *A. D. (\ I). E* and Fin Figure 3b correspond respectively to the positions *w-*J_h *xv-3. w-2. w~L wO* <*\\M\ wl* in

(a)

Figure 3: (a) Connection weights for the Quantity-Insensitive stress pattern of Painte. learned for words of up to 4 syllables in length, in the *N-l/S* model, (b) a schematic depiction of those weights in a buffer capable of processing words of up to 5 syllables in length. *E* has a positive connection weight to the output unit (not shown), and *D* a large positive weight: Fhas a large negative weight. and *(' 'A* very large negative weight.

Figure 3a. However. 3b depicts a larger buffer than is shown in 3a.

When the first syllable of a four-syllable word is the current input, the large negative weight *F* offsets the positive effect of if. resulting in a low activation to the output unit (corresponding to zero stress). When the current input is the second syllable, the appropriate output corresponds to primary stress, which is ensured by the positive activation from *D* and if combined, which is greater than 1**le*** negative activation from *F*. When the third syllable is the current input, the combined negative weights ('and *F* offset the positive activation of *D* and if combined, so that the output corresponds to zero stress. For the fourth syllable, the combined positive weights *B. I)* and //are sufficient |y greater than the negative weight *C to* yield an output corresponding to secondary st ress. but not so much greater than ('as to produce an output corresponding to primary stress.

The processing of the first three syllables of a/?*re-*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 /(////•-syllable word is the current input, the word is spread over positions *B. (\ I)* and /:: the negative weight *F* plays no role. However, when the fourth syllable of a /NT-syllable word is 1 he current input, the five-syllable word is spread over positions *B. C\ D.* if 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 if 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 to the first syllable of words of length greater than 1,** Fmust have the negative value shown. **Thus there is a conflict between the requirements for *F*, for the correct** processing of "length-4" and "length-5" training sets.

This is shown more formally below[1]. Let *Oi* be the threshold activation that must be delivered tot he out put unit for a response corresponding to primary stress: let *ff₂* ')e the threshold activation for secondary stress. The following constraints must then hold:

$$E > \theta_1 \quad (1)$$

$$F < -E \quad (2)$$

$$E + F < e_2 \quad (3)$$

$$\Rightarrow 0_2 > E + F$$

[1]The following analysis was formalized by Dave Touretzky. following a suggestion by (.ieoff Hinton.

$$\Rightarrow 0_2 > O_t + F$$

$$=> 0_2 - 0i > F \quad (I)$$

$$0i > (B + C + D + E) > 0_2 \quad (^r>)$$

$$=\bullet\ -0i < \{B + C + D + EX - 0_2 \quad (6)$$

$$0i > (B + C + D + E + F) > 0_2 \quad (\text{Y})$$

$$\Rightarrow \theta_2 < (B + C' + D + E + F) < \theta_1 \quad (8)$$

$$\theta_2 - \theta_1 < F < \theta_1 - \theta_1 \quad (9)$$

Inequality (i) expresses the constraint necessary for iT to be able to produce primary stress for a monosyllable. Statements (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. Statement (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. I)* and *E* to produce secondary stress when the current input is the final syllable of a four-syllable word: Statement ((3) is derived from (5). 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 ///Y-syllable word is indicated in (7). from which (8) can be derived. This is the constraint that makes the ""leiigth-.V" training set unlearnable. Adding (6) and (8) yields (9). which includes the condition ti_2 — *0* < '·• However, we previously have (4) *02* — *0* > *F*. Thus. (4) and (9) impose contradictory constraints on the value needed for *F*. as was discussed above: this leads to the non-learnability of Paiufe. An analogous demonstration can be made for Warao.

It is worth noting that the previous analysis of Paiute and Warao in terms of Primary Stress Placement (PSP) is *d(scriptin ly* accurate, at a gross level, based on observable properties of those and othor stress patterns. The *explanation* just given of the non-learnability of Painte can be made only in terms of the interactions of the properties of that stress pattern with rather specific properties of t he processing mechanisms and architecture employed in the A-/ model.

Paiute and Warao having proved unlearnable in the A'-*1* model, the next-minimal conned ionist architect ure examined was a 2-layer network with 2 output units: training targets had to be modified accordingly for this architecture. Essentially, target patterns for this architecture are |ength-2 vectors: the representations used were the fairly natural [0 0] for zero stress. [1 0] for secondary stress, and [1 1] for primary stress.

The two-output-unit architecture reduces the complexity of the task of each output unit, in **comparison with the task of a single output unit. A single output unit has to output three different** responses, **corresponding to zero, secondary and primary** stress. **With two output units (and given** the output representation adopted), the first output unit has only **to make a** two-way **distinction:** ihe response to the "current input" should be 1 if it receives (primary *or* secondary) stress, and 0. otherwise. The second output unit has to make a 2-way distinction between whether the "current input" should receive Primary stress (response 1) or not (response 0). The establishment of weights to each output unit should be simpler when only binary decisions are required, and thus the overall learning task should be simpler for Paiute and Warao in the A'-i/S'than in the *N-l/S* model.

In fact, in the *N-2/S* model. Paiute was learned in 732 epochs, and Warao in 757 epochs: connection weights are displayed in Figure 4. Each display is a representation of connection weights to each of the two output units, which are depicted separately. Thus, in the Paiute display, the weights shown in the leftmost of the two-network silhouettes are to the first output unit, and those shown in the rightmost silhouette are to the second output unit.

For Paiute. the weights enable correct stress assignment as follows: for the second output unit, only the second syllable will trigger the positive activation of *w-1* while not triggering the negative*

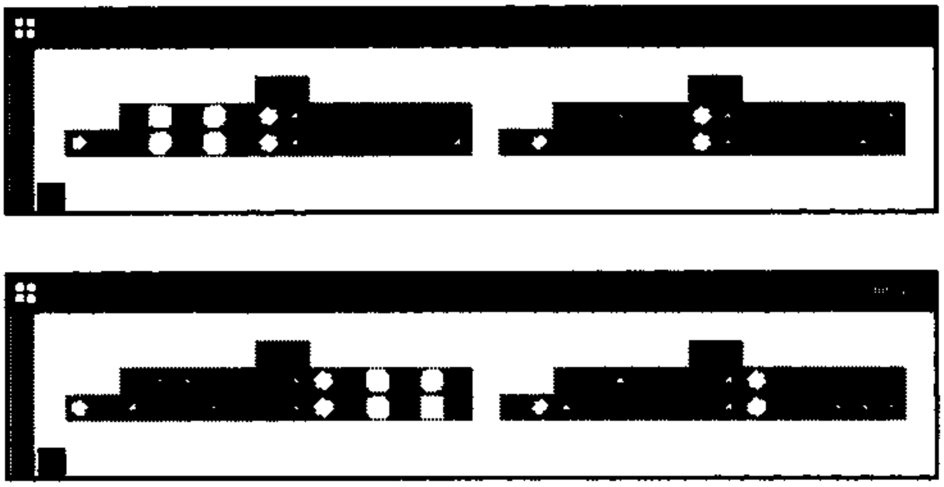

Figure 4: Connection weights for the Quantity-Insensitive stress patterns of Soutliern Paiute and Warao. learned by the *N-3/S* model. Numbers indicate the absolute magnitude of the largest weight, and hence the scale of the weights display.

activation of *w-3:* thus the second output unit 2 outputs a 1 when the second syllable is the current input, and a 0 for all other syllables. Weights to the first output unit enable it to produce a 1 in response to all even-numbered current inputs, and a 0 in response to all other syllables. Wrights for Warao can be analyzed similarly, since it is the mirror image of the weights for Paiute.

4.1.6. From **QI to QS stress patterns**

Consistent with the aim of exploring the minimal architectures necessary for learning of various kinds of stress patterns, simulations with Quantity-Sensitive (QS) patterns were also made with the 2-layer. i-output unit model. However, for QS patterns, information about syllable weight has necessarily to bo included in the input representation - the input has to consist of (encoded) sequences of "II" and *i/* tokens: a *syllabic* input representation is (by definition of quantity-sensitivity) inadequate. IIV *iyht-striny* representations were therefore adopted, as discussed in Section IJ.-I..

The use of weight strings greatly increases the size of training data that needs to be considered: for a maximum syllable length of /?. there are only /? input patterns using the syllabic representation, but there are V^_i 2" = (2"$^{+1}$ — 2) possible weight strings. For // = 7. this is 7 vs. 25 f possible patterns.

In order to obtain learnability results (regarding minimal architectures, and learning times) that might be consistent across both QI and QS stress patterns (i.e., that might reflect differences between the language patterns themselves, rather than merely **reflecting differentials in** the complexity of the learning task arising from the differing input representations **and** training set sizes), simulations for the QI languages were re-run using the weight-string representation. (The stress patterns for all possible weight strings of length *n* are the same for a QI language).

The training set initially used included exactly one instance of each possible weight string, for weight strings of up to length 7 (the *N-l/HL* model). Simulations with this input representation were unsuccessful for several of the QI patterns of complexity greater than **"1000".** Since all weight-strings of a given length have the same associated stress pattern for a QI language, the weight-st ring training set has a distribution that is heavily skewed in favor of longer words, which can distort the error gradient and trap the back-propagation learning procedure in a local minimum in weight space. To correct this, the distribution was adjusted so that there was an equal number of instances of weight-strings of each length up to 7. and the QI simulations re-run with this "equalized"* training set (the . \\-*I/HLE* model). All the QI patterns learnable in the *N-l/S* model were also Jearnable in this model, thus providing a basis for comparison of QI learning times with learning times for

RBF	LANGUAGE	DESCRIPTION OF STRESS PATTERN	EPOCHS
AH)	Koya	Main stress on first, secondary stress on heavy syllables	2
All	Eskimo	Stress the final syllable and closed (heavy) syllables	3
A12	Maiayalam	Stress first syllable except where first vowel short and second vowel long	20
A13	Yapese	Stress final syllable except where final vowel short and penult long	18
Ail	Ossetic	Stress first vowel if long, else second vowel	28
AI5	Rotuman	Stress last vowel if long, else penult	21)
A10	Koini	Stress first long vowel, else last vowel	217
A17	Cheremis	Stress last long vowel, else first vowel	211)
A18	Mongolian	Stress first long vowel, else first vowel	2312
AH)	Mayail	Stress last long vowel, else last vowel	2304

Table 4: Quantity-Sensitive patterns: description and learning performance in the *N-l/HLE* model.

QS patterns.

4.2. Quantity-Sensitive stress patterns

Ten Quantity-Sensitive (QS) languages were examined: Koya. West Greenlandic Eskimo. Malayalam. Yapese. Ossetic. Rotuman. Eastern Permyak Komi. Eastern Cheremis. Khalkha Mongolian, and Aguaeatec Mayan.

The stress patterns of Malayalam fc Yapese. Ossetic <k Rotuman. Komi V* Cheremis. and Mongolian <V Mayan are mirror images of each other.

In Koya. primary stress is on the initial syllable, and secondary stress falls on closed syllables and syllables with a long vowel. Eskimo stresses the final syllable and closed syllables.

In Malayalam. primary stress falls on the initial syllable except when the initial vowel is short and the second long, in which it falls on that second syllable. In Yapese. primary stress is on the final syllable except when the final vowel is short and the penultimate vowel long, in which case stress falls on the penult.

Ossetic assigns primary stress to the first vowel if it is long, otherwise to 1 he second vowel. Hotiimaii st resses the last vowel if it is long, otherwise the penultimate vowel.

In Komi, stress falls on the first long vowel, and on the final vowel in words without long vowels. In Cheremis. stress falls on the last long vowel, and on the initial vowel in words without a long vowel.

Mongolian stresses the first syllable with **a** long vowel, and the first syllable if there are no long vowels. Mayan stresses **the last syllable with a long vowel, and the last syllable if there are no long** vowels.

Table I summarizes the stress patterns and learning times.

4.2.1. Learning results

Using the minimal possible connectionist architecture *(N-l)* and the equalized weight-string representation, all ten QS stress patterns were learned with 100% accuracy. The last column of Table 4 shows the number of epochs of training necessary to achieve 100% accuracy for each stress pattern.

Learning times for each of the pairs of mirror image patterns are very close to each other: 20 epochs for Malayalam and 18 epochs for Yapese: 28 epochs for Ossetic and 29 epochs for Rot iimaii; 217 epochs for Komi and 219 epochs for Cheremis: 2312 epochs for Mongolian, and 2304 epochs for Mayan. The learning time for Koya was 2 epochs, and 3 epochs for Eskimo.

4.2.2. Analysis of connection weights

Connection weights for the learned patterns are displayed in Figure 5. Interpretation of the displays is as discussed in Section 4.1.2.. However, with the HLE representation for inputs, a heavy syllable is represented by a [1 0] vector, and a light syllable by a [0 1] vector. For heavy syllables, therefore, there will be a 1 in the bottom row (row *H*) and a 0 in the top row (row *L*): for light syllables, a 0 in the bottom row (*H*) and a 1 in the top row (*L*). Thus 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 the stress patterns.

For Koya, the bias weight enables a fairly high positive activation to the output unit; there is also high positive activation when a heavy syllable is the "current input", arising from *w0b*. 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 the pair *w–1* produces a large negative activation to the output unit, whether the syllable in *–1* is heavy or light, thus offsetting positive activation from the bias unit. Net activation to the output unit will be low, resulting in a low output denoting zero stress, *unless* the current input is a heavy syllable, in which case the large positive weight *w0H* enables a large activation to the output unit. This positive activation, plus that of the bias unit, together produce greater positive activation to the output unit than is offset by the negative activation from *w–1*; therefore 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 one of two conditions exist: Either it is heavy (large positive activation from *w0H*, 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 *w0H* will provide no stress, and additionally, *w+1H* will damp stress provided by the bias unit. If the current input is the second syllable, it receives stress only if it is heavy (positive activation from *w0H*) 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 *w0H*. 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 input" is heavy, it should be assigned primary stress only if there have been no preceding heavy syllables. A heavy current syllable receives stress from *w0H* 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 to the output unit from any buffer position containing a light syllable. When a light syllable is the current input, therefore, positive activation to the output unit 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 no heavy syllables preceding the current input—i.e., just in case the current input is the last

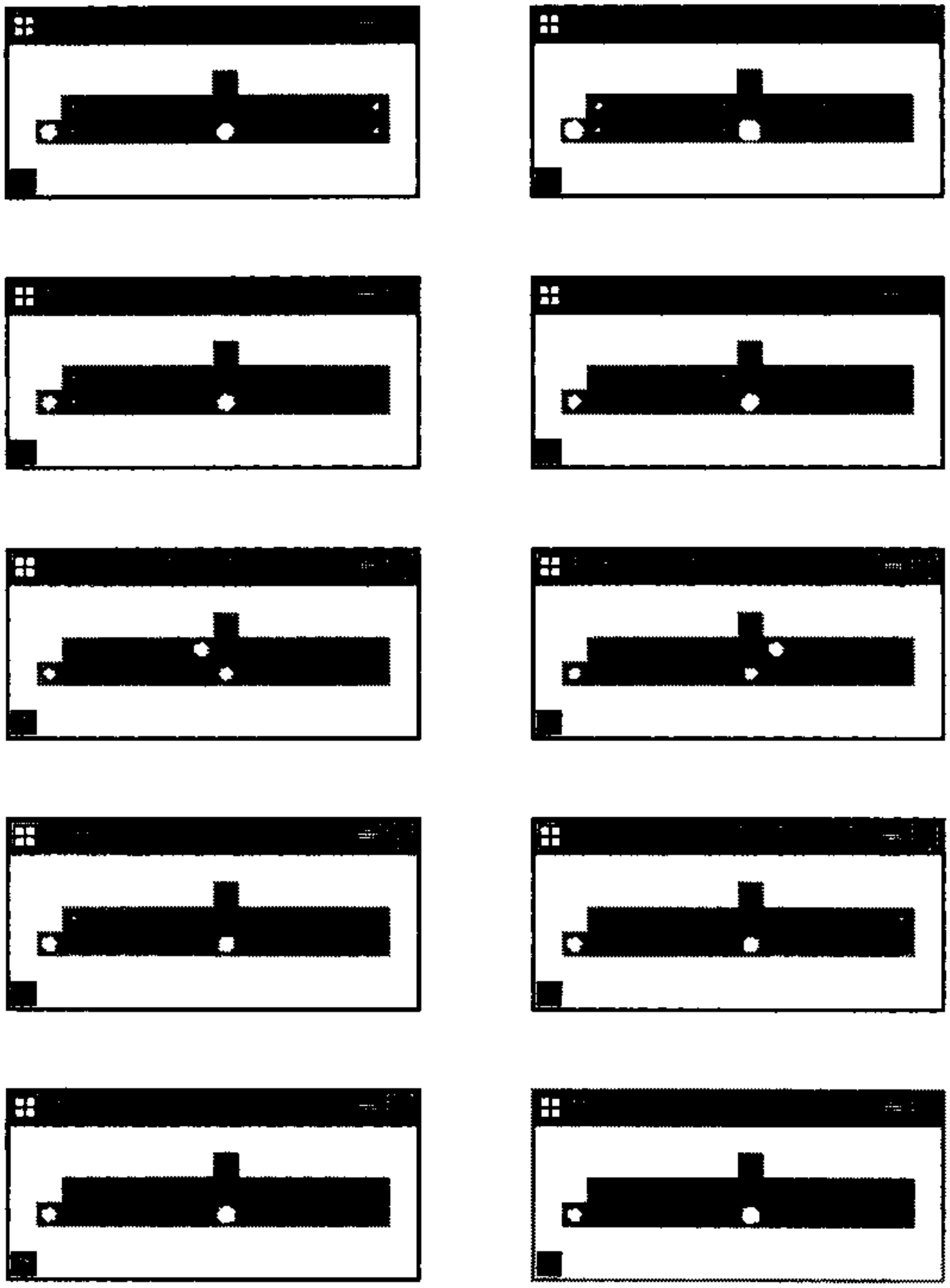

Figure 5: ('onnection weights for Quantity-Sensitive stress patterns learned by the *N-l/IILEmoilvl.* Numbers in each title bar indicate the absolute magnitude of the largest weight, and hence the scale of the weights display.

syllable in a word without any heavy syllables.

The weights *w-lH* through *iv-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.

4.2.3. Learnability

Tlie learnability analysis proposed in Section 4.1.4. on the basis of QI patterns requires some refinement. Placement of primary stress (PSP) was hypothesized as taking binary values; a 1 value for PSP 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 projHrrtion of cases in which primary stress is not assigned at the edge of a wont*, the learning results for both QI and QS patterns can be integrated, to a large extent, into a unified account.

The learning times for Malayalam and Yapese are approximately 20 epochs, while those for Ossotic and Rotuinaii are approximately 30 epochs. The difference between these pairs of stross patterns is as follows: For Malayalam and Yapese. primary stress is placed at the edge *execpt* 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 0.50 of I**e*** cases.

The five factors discussed earlier were: Primary Stress Placement (PSP): Stress Clash Avoidance (S('A): Alternation (Alt): Number of Primary Stresses (NPS): and Number of Stress Levels (NSL). TIK[1] values of these indices, respectively, for both Malayalam and Yapese. are [0.25 0 0 1 ()]: and for both Ossetic and Rotuman. [0.50 0 0 10]. The difference between learning times for those pairs of otherwise identical patterns can then be accounted for in terms of differing values of the PSP measure.

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

Tho learning times of Komi and Cheremis are substantially higher than those of Koya. Kskimo. Malayalam. Yapese. Ossetic and Rotuman. As discussed above, for Komi and Cheremis. tho networks in effort 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 large magnitude. The connection weight displays of Figure 5 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[2]: (I) Thoro aro 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-positiexnal* 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 *aggngative* information: information about *all* the syllables to the left of S in the weight-string. The simulated "scanning" referred to above provides precisely this *aggregative* information: and similarly for Cheremis.

[2]Thf following analysis owes much to discussion with Dave Touretzky.

Komi and C'heremis 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 *singlc-ltositional* 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 *singlf-positional* in format ion).

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 large magnitude across the buffer, and this requires greater learning times. Whether or not *aggreyative* information is needed, therefore, seems to be an additional 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 III* through *w 6H* must be capable of damping the positive activation from *icOH.* 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-lH* or *w-lL* 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 to the right of the current input, to determine whether there is a heavy syllable. Thus, for Mongolian, there is *aggtrgativt* information required about heavy syllables both to the left of the current input, *and* to its right. 'This seems |(> |>p what makes the pattern so difficult to learn. (As a matter of fact. 1 here is a. kind of compounding of aggregative requirements: the weight *wOH* must be large enough to overcome *all* of *ir+lH* through *w+6H.* and so must be rather large: but also, each of *w-lH* through //• *(ill* must he able to override *ivOIL* 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: 2N.(M. 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—whether or not *aggirgative* information is required but that comes into play only in the case of QS patterns. This can be treated as a sixth index of computational complexity that can take the values {0, 1,2}: 0 if no *aggregative* information is required *(single-positional* information suffices); 1 if one kind of aggregative information is required (Komi. Cheremis): and 2 if two kinds of aggregative information are required (Mongolian, Mayan).

The learning results and learnability analysis for both QI and QS stress patterns are summarized in Table I>: learning times for both QI and QS patterns are with the *X-l/HLE* model, which should make them comparable.

The differences in learning times across QI patterns are less marked in comparison with the differentials in Table 3. which summarized QI learning results with the *N-l/S* model. This is the result of the increased training set size with the equalized weight-string representation as compared with the syllabic representation. Learning times are. nevertheless, consistent with the overall learnability analysis.

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.

Agg	PSP	SCA	Alt	NPS	NSL	QI LANGUAGES	REF	EPOCHS	QS LANGUAGES	REF	EPOCHS
0	0	0	0	0	0	**Latvian** French	Al A2	1 1			
0	0	0	0	0	1				Koya	A10	2
0	0	0	0	1	0				Eskimo	All	3
0	0	0	1	0	1	Maranungku Weri	A3 A4	2 2			
0	0	1	1	0	1	Garawa	A5	7			
0	0.25	0	0	0	0				Malayalain Yapese	A12 A13	20 18
0	0.50	0	0	0	0				Ossetic Rotuman	AJ4 A 15	28 29
0	1	0	0	0	0	Lakota Polish	A6 AT	9 9			
0	1	0	1	0	1	Paiute Warao	A8 A9	** **			
1	0	0	0	0	0				Komi Chereinis	AJO A17	217 219
2	0	0	0	0	0				Mongolian Mayan	A18 A19	2312 2301

Table* r>: Summary of results and analysis of QI and QS learning in the *N-l/IILE* model. Agg=Aggregative Information: PSP=Primary Stress Placement: S(.'A=Stress Clash Avoidance: AM=Altenmtion: NPS=No. of Primary Stresses: NSL=No. of Stress Levels. References index into Tables I and 1.

The QI stress patterns Lakota and Polish have higher complexity indexes than the QS stress paitertis Malayvalam. Yapese. Ossetic and Rotuman. but lower learning times. Quantity-sensitivity thus ap|>ears to affect learning times, as seems reasonable to expect, due to the distribution of 1 he //A/:'I raining set. 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.

4.3. Generalization

The ability of trained networks to generalize has been examined for QI stress patterns. **All** the **simulations described so far involved training sets representing words of up to seven syllables in length (a "length-7" training set). To test generalization, networks that had been trained to 100%** accuracy on those ("length-7**) **training** sets **were re-trained on training** sets **of all words of up to** length X syllables (a "length-8" training set). The criterion for generalization was that the time taken (in epochs) to learn the "length-8" training set after prior training on the "length-7" training sot should be less than the time taken to learn the "length-8" training set *ab initio.* i.e.. without any such prior training.

Results are summarized in Table 6. which shows learning times for a "length-S" training set. for all the QI stress patterns learnable in the *N-l/Smodel.* The first column shows learning time from scratch, i.e. without prior training on the "length-7" set. and the second column shows learning time for the "length-*" set when the network has already been trained on the "length-7" set.

For Latvian. French. Lakota and Polish, no extra training is required to learn stress assignment for the "length-8" set. once the "length-7" set has been learned. Turning back to the weight displays in Figure 2. this can be seen as a natural outcome of the fact that these stress patterns have a single, fixed stress: the weights that are established with the "length-7" training set will, clearly.

		EPOCHS TO LEARN UP-TO-LENGTH-8 WORDS	
REF	LANGUAGE	From scratch	Pre-training
A1	Latvian	22	0
A2	French	19	0
A3	Maranungku	46	35
A4	Weri	48	35
A5	Garawa	161	19
A6	Lakota	303	0
A7	Polish	299	0

Table 6: Generalization for QI stress patterns. Column 3 shows the number of epochs required to learn a training set of words of up to 8 syllables in length. Column 4 shows the number of epochs required to learn the same training set after the network has already been trained on words of up to 7 syllables in length. References are made to Table 1. which describes these stress patterns.

enable the networks to process words of arbitrary length. For languages with a single fixed stress, therefore, there is perfect generalization.

For Maranungku and Weri. the weight displays of Figure 2 (which show weights established with a "length-7" training set) indicate that additional weights would need to be established to handle N-syllable words, since stress is neither fixed nor single, but alternating, and with two levels. The same observation seems true of Garawa. For Maranungku and Weri. there is in fact a reduction in learning time, from about 45 epochs to learn the "length-8" training set from scratch to about 35 epochs to learn it after training on the "length-7" training set. While this is itself a non-trivial reduction, the relatively low number of epochs for *ab initio* training on the "length-N" set may in fact be masking the generalization effect: in the case of Garawa. for instance, there is a very substantial reduction in learning time, from 161 epochs to 19 epochs.

An alternative, and more stringent, criterion of generalization would define "generalization" as the ability to assign stress to a word of greater length than previously encountered, *on first presentation*. As will be discussed in the next section, it is by no means clear whether such a criterion is truly appropriate in the present context. In any case, it seems reasonable to conclude that the present results indicate some ability to generalize previous learning.

5. Discussion

5.1. Performance of the model

5.1.1. Generality

The simulation results reported in the previous section indicate that the connectionist models considered can learn stress-assignment phenomena quite generally. A total of nineteen actual stress patterns was examined: seventeen of them were learnable by the minimal connectionist two-layer model, and the remaining two (Paiute and Warao) were learnable by a two-layer architecture with one additional output unit. Thus all the stress patterns examined were learnable by relatively simple connectionist models.

The linguistic analysis of stress patterns in terms of parameters is. in fact, incapable of providing an account of all the stress patterns considered here. For example, the stress pattern of Garawa does not seem amenable to characterization in terms of parameter settings alone. Thus the description of Garawa given by Hayes ([Hayes 80. pp. 54-55]) 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 purely parametrized characterization adopted by Dresher *k* Kaye (see Section 2.3.)? and, in fact, Dresher *k* Kaye do not report any examination of the stress pattern of Garawa. 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 a parameter setting scheme, and Dresher *k* Kaye do not discuss this pattern either.

One of the achievements of metrical theory is considered to be its relatively good coverage of the phenomena, in its domain; this in turn has yielded parameterized characterizations, and these have formed the basis of computational models such as those of Dresher fc Kaye. and Nyberg. Where the parametrized characterization breaks down, therefore, a model based on a direct mapping from these abstract parameters to computational operations has no basis for modeling stress assignment. *In order to model the learning of non-parameterized stress systems, arbitrary computational operations would have to be introduced (as. in fact, they are in the linguistic analysis of these stress patterns). This would, however, defeat the purpose of the parameterized approach, one of whose motivations is to identify a constrained set of universal operations.

Thus, current parametrized treatments of metrical phonology do not have fully general coverage: neither, correspondingly, do computational models employing those parameters (and related constructs) as processing primitives. It is therefore interesting that the stress patterns of both Carawa and Aklan were learned by connectionist networks in the present simulations. (iarawa has already been discussed (Section 4.1.1.. Table 1): the stress pattern of Aklan was learned with difficulty (50.4X1 epochs of training), in a three-layer architecture.

In a sense, therefore, it appears that the processing primitives of the connectionist models employed here provide a broader and more general basis for stress assignment than do direct mappings from the constructs of (parametrized) metrical phonology.

5.1.2. Generalization

As discussed in Section 4.3.. the ability of the trained networks to generalize was examined for Ql stress patterns. In all cases, the time taken to learn a set of words of up to S syllables in length (a "lengt h-X" training set) when the network had previously been trained on words of up to 7 syllables in length (a "length-7" training set) was significantly less than the time taken to Irani a "lengl h-X" training set without prior training on a "length-7" training set. This indicates some ability to generalize from previous training.

It might be argued that, for true generalization, the networks should have been able to assign stress correctly to 8-syllable words on first encounter, once they had learned a "length-7" training set: certainly, this is the prediction of the linguistic model, in which parameter settings, once determined, provide the basis for processing (i.e.. assigning of stress to) words of arbitrary length. However, as will be argued below, the fact that this is what linguistic theory predicts does not in itself render the present generalization results inadequate.

An obvious question is that of how generalization occurs in human processing. To conform to the predictions of linguistic analysis, a human language learner should be able to assign stress correctly to words of greater length than it has previously encountered, even on first exposure. It has not-been possible to determine whether there are research results that bear directly on this question: Dresher <? Kaye state that the stress acquisition literature does not appear to provide evidence that can be directly related to parameter setting ([Dresher 90. page 42]. In general, developmental evidence relating to stress acquisition seems to be limited (Peter Gordon, personal communication).

A further relevant question here is: What kinds of information does a human language learner bring to bear on the acquisition of stress? Linguistic theory presents syllable structure as the only necessary **informational** basis **for** *descriptions* **of** stress **patterns. This does not mean,** however, that this is **the only information utilized in actual human processing: potentially, all the information** available **in the** speech **signal (which encodes more than just syllable structure) is available. It** seems quite likely that human processing utilizes **information about the** *segmental content* of words. For example, in generalizing stress from words of length three syllables to words of length four syllables, it is at least plausible that the human language learner is guided by analogies with previously encountered words: thus, familiarity with the three-syllable words *charity* and *clarity,* and the four-syllable word *capacity* could provide the knowledge that stress appears two syllables before /// in words suffixed with *ity.* This could aid in generalizing to the five-syllable word *genemlity.*

The proposition that the *human* processing underlying stress learning/assignment makes reference only to syllabic information may be true or false, even though it is assumed to be true in linguistic theory.

Now. suppose that evidence were to reveal that children are. in fact, able to assign stress correctly to words of greater syllable-length than they have previously encountered, even the very first time they are exposed to them (i.e. they are able to generalize "instantaneously").

If it is *false* that human stress processing makes reference only to syllabic information, then "Instantaneous" generalization does not reveal inadequacies in the connectionist architecture of the present model as much as it reveals inadequacies in the input representation. Input representation, according to linguistic analysis, has been taken to encode syllabic information only.

If it is *trm* that human stress processing makes reference only to syllabic information, then "instantaneous" generalization would reveal inadequacies in the present models, since they incorporate the necessary information, but do not produce instantaneous generalization.

In sum. the generalization results from the present models are not necessarily in accord with the predictions of linguistic theory: however, as has just been argued, any difference does not in hself invalidate those results. The results would be inadequate only if (a) human processing does exhibit "instantaneous" generalization in the domain of stress, and (b) human processing in stress phenomena makes reference only to syllabic information. Both of these are empirical questions, and answering them does not fall within the purview of formal linguistics.

5.1.3. Systematicity

Fodor <V Pvlvshvn have argued ([Fodor 88]) that *systematicity is* an intrinsic aspect of the way humans use language. Thus, it is not possible for a human language user to be able to understand the meaning of *John loves Mary* without being able to understand the meaning of *Mary loves John:* **this understanding implies the ability to grasp the predicate relations inhering between the verb, subject and object—to grasp structural relations in general. Since connectionist models do not** incorporate such structural relations, they **can** provide **no** basis **for** such **systematicity.**

The metrical theory analysis of stress systems sets up. for each system, an inventory of represenintional structures, and operations on those structures, that are employed in characterizing the assignment of stress to words in that stress system. In the example of systematicity given above, it is awareness of the structures of the sentences that underlies the human inability to understand one wit hunt being able to understand the other, or to know that one is grammatical without also knowing that the other is grammatical. In analogous fashion, it could be argued that the descriptions of stress patterns involve reference to structures such as *binary quantity-insensitive feet, unbounded quantity-sensitive feet,* and so on. The ability of a child to learn a stress system described by a set of structures deployed in one particular configuration (e.g.. *binary, quantity-insensitive feet, constructed from right to left)* implies the ability to learn a stress system described by those same structures, configured in some other fashion (eg., *binary, e/uantity-inse nsitive feet, constructed from left to right).* More generally, it implies the ability to learn some other stress system which draws on the same inventory of representational and operational primitives. Viewed this way, systematicity is a by-product of the notion of Universal Grammar.

In this view, there should be no reason to expect a connectionist model to be able to learn groups of systematically related stress patterns except in an arbitrary way: none of the structure that establishes those relationships has been incorporated in the model. Thus, if a connectionist model can learn stress system A, it will be entirely coincidental if it can also learn stress pattern B, which is the mirror image of A.

In fact, as has been seen, the results of the simulations are remarkably systematic in this regard. Several pairs of mirror-image patterns were examined, and learning times for the members of the pairs were almost identical (see Tables 1 and 4). Moreover, the "knowledge" of the stress pattern that is encoded in the connection weights reflects very systematically the fact that the pairs of patterns are mirror-images of each other, as can be seen from Figures 2 and 5.

The performance of the models in the simulations examined here therefore seems to exhibit some kind of systematicity, despite the fact that none of the structured representations of metrical theory have been directly incorporated.

5.2. Levels of analysis

5.2.1. An existence proof

As discussed earlier (Sections 1.3., 2.4.), there is a clearly defined set of theoretical constructs in terms of which the linguistic analysis of the domain of *stress systems* in language is stated. That analysis, moreover, provides fairly good coverage of the observed phenomena of the domain. For the "classical" approach, therefore, there is a well-defined set of putatively computational primitives in terms of which to formulate processing models.

The connectionist models described here have in fact proved capable of learning a wide variety of stress patterns, *without* the incorporation of theoretical constructs such as *metrical tree* and *parameter* as processing primitives. In other words, the models have been able to learn correctly the stress-assignment phenomena described by the linguistic analysis without recourse to the structured representations employed by that analysis. The connectionist models do employ insights from linguistic theory, and therefore do, to an extent, incorporate linguistic constructs; nevertheless, this differs substantially from a model such as that of Dresher & Kaye, which is essentially a computer implementation of linguistic theory and learnability theory.

To the extent that current simulations are considered successful, their success can be viewed as constituting something like an existence proof (in the domain of stress systems) for the connectionist view that the notions of linguistic theory need not map directly onto the computational primitives of a processing account. Conversely, these results weaken the classical argument that connectionist techniques are *in principle* inappropriate simply because their representation schemes do not incorporate the structure characteristic of linguistic analysis. The results also support the argument that linguistic theory properly belongs to a different level of analysis from that of computational modeling, and that it is a crucial *assumption* of classical modeling (rather than a matter of necessity) that the mapping between these two levels should be rather direct. In other words, having Marrian-Level-Two representations and algorithms directly reflect the structures posited by Marrian-Level-One linguistic theory is not the only way in which a processing model can incorporate, and be guided by, the insights of that theory.

5.2.2. *Gedanken* experiment

The Yapese Room

EXPERIMENTER: Well, 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 individual words of the language Yapese. Let's see what different analyses of the stress pattern you come up with.

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

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

COMSCI: Really? All I can tell is that stress is sometimes on the last syllable, and sometimes on the second from 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.

COMSCI: Hmm. that's pretty neat. *(To the experimenter)* Do you hav<j (lain froin other languages?

(Many stress patterns later....)

COMSCI: 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?

COMSCI: 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 second-last syllable in the array is long, in which case I assign stress to the second-last element.

COMPL1NG: Well, that's rather naive, actually. In reality, you should be looking at the linguistic 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?

COMSCI: I'm not sure I know what you're talking about. Well, anyway, Tin 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....)

COMSCI: I'm back! My program's up and running. I call it Artificially Intelligent Stress Processor—AISP. for short.

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 thought)* What? You know, these stress patterns are really interesting. I've been thinking about them since you two went away, and I think there's a much better linguistic analysis to be made. Bettor still, it is one that provides a means of showing that these patterns differ from each other along only a small number of dimensions of variation, which can be regarded as parameters of the model. It's really quite an Hega.nt 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 the way in which the data interact with universal principles to establish parameter settings. But basically, your metrical stuff provides really clear representations for a computational modH. I'll be back soon, with SPE-2.

COMSCI: 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 neural pathways in the brain, isn't going to give you a particularly good understanding of the pattern. Metrical theory can do that, and can also help uncover interesting relationships among patterns. It might even suggest constraints on the wiring of human stress processing apparatus. But the primary benefit of the theory is in organizing the fact s. 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 1 hat I < hough!, 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 "units".

LINGUIST: I never said there were metrical trees in the human brain: merely that then[1] 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 modelled real speech produced by a real human with a real brain.

C'OMSCI: Say. I don't know what you guys are going on about, but I'd like to take a look at this speech synthesizer. (*Turns to Connlwekfr)* 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 stresses per word. Third, whether it's a pattern with stress clash. Fourth.....

COMSCI: This is really interesting. Tin pretty sure I can simulate these results. All I need is a binary counter 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. Connniag. looking enraged.)

CONNMAG: *(Boxing Connhacker's ears)* Imbecile! Is this what you've learned? I heard you spouting your spurious analysis of "scanning" and "primary stress placement". Bah! Don't you see that your "explanations" of the different learning times of different stress patterns are only abstractions based on what's observable about the stress patterns? 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 *\hr* architecture of the model you've set up. 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. Your descriptive analysis, stated in terms of observable properties of the stress patterns, such as "alternation" and "primary stress placement" helps to identify characteristics of those patterns that seem to correlate with their learnability. Thus, it helps organize and make sense of the observed learning results. Hut don't you fall into the trap of thinking your *descriptions* are *explanations.* If you do that, you'll end up inventing more theories of the kind these people have been talking about. Who *arc* all these people, anyway?

[Carlson 89] Carlson, G. N. *k* Tanenhaus, M. K., Introduction, in G. N. Carlson *k* M. K. Tanenhaus (eds.), *Linguistic Structure in Language Processing,* Kluwer, Dordrecht, The Netherlands, 1989.

[Chomsky (38] Chomsky, N. *k* Halle, M., *The Sound Pattern of English,* Harper &: Row, New York, 1968.

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

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

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

[Elmaii 89] Eliiian. J. L., *Representation and Structure in Conneetiemist Models,* CRL Technical Report 8903. Center for Research in Language. University of California. San Diego. August 1989.

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

[C«olclsniiUi 90] Goldsmith. J. A.. *Autosegmental and Metrical Phonology,* Basil BlackwelL Oxford. England. 1990.

[Halle N7a] Halle. M. k Yergnaud. J.-R.. Stress and the Cycle. *Linguistic Inquiry* IS. 15 SI.

[Halle S7b] Halle. M. &' Yergnaud. J.-R.. *.4/? Essay on Stress.* MIT Press. Cambridge. MA. 1987.

[Hayes SO] Hayes. B.. .4 *Metrical Theejry of Stress Rules,* doctoral dissertation. Massachusetts Institute of Technology. Cambridge. MA. 1980. Circulated by the Indiana University Linguistics Club. 19X1.

[Hayes SI a] Hayes. B.. The Phonology of Rhythm in English. *Linguistic Inquiry* II. 33 71.

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

[JarkendoffSS] Jackendoff, R., Why are they saying these things about us? *Topic...Comment,* in *Natural Language and Linguistic Theory* 6 (1988), 435-442.

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

[Libennan 75] Liberman. M. Y.. *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. Y. tt Prince. A. S.. On stress and linguistic rhythm. *Linguistic Inquiry* 8. 249-336.

[Liberman 86] Liberman, M. Y., Prospects for Connectionist Models of Natural Language: The Connectionist Boom, in *Proceedings of the 24th Meeting of the ACL,* 181—183, 1986.

[Marr 82] Marr, D., *Vision.* Freeman. San Francisco, CA. 1982.

[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.

[McClelland 86] McClelland. J. L.. Rumelhart, D. E., *k* Hinton. G. E.. The Appeal of PDP. in D. E. Rumelhart. J. L. McClelland *k* the PDP Research Group, *ParalUI Distributed Processing. Volume 1: Foundations.* MIT Press. Cambridge. MA. 1986.

[McClelland 88] McClelland. J. L. fc Rumelhart. D. E.. *Explorations in Parallel Distributed Processing: A Handbook of Models. Progretms. and Exercises** MIT Press. Cambridge. MA. 1988.

[Nyberg 89] Nyberg. E. H.. *Weight Propagation and Parameter Setting.* Ph.D. thesis proposal. Dept. of Philosophy. Carnegie Mellon University. 1989.

[Prince 70] Prince. A. S.. *Applying stress,* unpublished ms.. University of Massachusetts. Amherst. 1976.

[Prinro s:{] Prince. A. S.. Relating to the Grid. *Linguistic Inquiry* 14. 19—-100.

[Rumelhart 80] Rumelhart. D. E.. McClelland. J. L.. fc the PDP Research Group. *ParalUI Distributed Processing: Explorations in the Micrejstructure of Cognition. \ólum< I: Foundations.* MIT Press. Cambridge. MA. 1986.

[Sejdenberg S9] Seidenberg. M. S.. Reading Complex Words, in G. N. Carlson k M. K. Taiienhaus (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.

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

[Treiman N9] Treiman. R.. The Internal Structure of the Syllable, in G. N. Carlson <V, M. K. Tanenhaus (eds.). *Linguistic Structure in Language Processing.* Kluwer. Dordrecht. The Netherlands. 1989.

[van der Hulst S2] van der Hulst. H. fc Smith. N.. An overview of Autosegniental and Metrical Phonology, in H. van der Hulst *k* N. Smith (eds.). *The structure of phonological representations. Part I.* Foris Publications, Dordrecht. The Netherlands, 1982.

[Vergnaud 78] Vergnaud. >*h%R.* *k* Halle. M.. *Metrical structures in phonology,* unpublished ms.. Massachusetts Institute of Technology. Cambridge. MA. 197K.