Occasional Papers in Linguistics

Dynamic Computational Networks and the Representation of Phonological Information
Gary N. Larson
November 1992

The University of Chicago

The Department of Linguistics
1010 East 59th Street
Chicago IL 60637
ACKNOWLEDGMENTS

The writing of a dissertation is always a collaborative process, involving a committee, faculty mentors, fellow students and a large number of colleagues in the broader linguistics community. As a result, I would be remiss not to thank them for their contribution to this project and to my intellectual development. But in many respects, such a brief thank you would prove grossly inadequate. During the last several years I have had the privilege of participating in a unique intellectual community at the University of Chicago, where the only party line has been the fact that there is no party line. In John Goldsmith's ongoing seminars with an aspiring group of phonologists (Diane Brentari, Anna Bosch, Caroline Wiltshire, Michael Kim, Elisa Steinberg, Manuela Noske, Karen Peterson and others), we have had the freedom to take good ideas wherever they may be found while critically evaluating both the strengths and weaknesses of competing approaches to phonology. The eclecticism evident in this dissertation is the result of such an open-minded inquiry.

John Goldsmith has not only been my teacher and mentor, but is, in this project, a collaborator in the fullest sense of the term. The use of the personal pronoun "we" throughout the paper is not simply a polite form of address but an indication that I speak for the two of us. After three and a half years of collaboration, it would be very difficult to parcel out credit as to which ideas are mine, his, or more broadly, ours. So in writing these words, I acknowledge a debt to all those who have contributed, with the proviso that all errors, whether in thought or exposition are fully my own responsibility.

The completion of a doctoral program involves much more than intellectual effort, however. I owe an even greater debt to my wife, Beth, and my children, Jeffrey and Jonathan for their patience in the face of long hours of work, resulting far too often in household chores not being shared, games not being played, and generally being taken for granted. Without their consistent support, this project would have ended before it began.

I am also thankful to Wheaton College which has supported the completion of my degree with reduced teaching loads and constant encouragement. Special conferences that discussed the application of connectionist architectures to linguistic problems at the University of New Hampshire (1990) and the Abbaye de Royaumont (1991, 1992) also proved formative to our thinking. Finally, the completion of the dissertation has been partially funded through National Science Foundation grant #BNS 9009678 and a fellowship funded by the University of Chicago Division of Humanities.
TABLE OF CONTENTS

ACKNOWLEDGMENTS ................................................................. ii

CHAPTER ONE: Introduction / Historical Review 1

1.1 Justification for a dynamic computational model — Wave Phonology 3
1.2 Evaluative criteria 7
1.3 Outline of the dissertation 8

CHAPTER TWO: Architecture of the Model 9

2.1 Network Architectures — A brief survey 10
2.2 Dynamic Computational Network 14
2.3 Autosegmental/Featural Network 16
2.4 Syllable Network
   2.4.1 Input — Inherent sonority 20
   2.4.2 Processing
      2.4.2.1 Lateral inhibition/excitiation 21
      2.4.2.2 Feedforward inhibition/excitiation 26
      2.4.2.3 Feedback inhibition/excitiation 27
   2.4.3 Recognition 28
2.5 Metrical Network
   2.5.1 Input — Positional, lexical, moraic activation 31
   2.5.2 Processing 31
   2.5.3 Recognition 32
2.6 Modular interfaces 33
2.7 Representation of temporal structure — Reprise 34

CHAPTER THREE: Syllabification 36

3.1 The syllable as a phonological constituent 36
   3.1.1 Constituency models of syllable structure 37
   3.1.2 Sonority as a phonological variable 39
3.2 Spanish
   3.2.1 Spanish onsets — Harris' account 41
   3.2.2 Spanish onsets — Network account 46
   3.2.3 Spanish rimes — Harris' account 48
   3.2.4 Spanish rimes — Network account 50
3.3 Icelandic
   3.3.1 The problem: Predicting vowel length in Icelandic
   3.3.2 Itō's solution
   3.3.3 Critique — Descriptively adequate but "implausible"
   3.3.4 Computational alternative
3.4 Imdlnawn Tashlihyt Berber
   3.4.1 Dell & Elmedlaoui's analysis
   3.4.2 Network alternative
3.5 Problem cases for sonority hierarchy
   3.5.1 Sonority plateaus
   3.5.2 Sonority reversals
   3.5.3 Syllable peaks that are not sonority peaks
3.6 Conclusion

CHAPTER FOUR: Stress Assignment

4.1 Architecture of Metrical Network
   4.1.1 Input: Positional activation (initial/final)
   4.1.2 Input: Syllable weight (heavy/light/bias)
   4.1.3 Input: Lexical stress
   4.1.4 Processing: Lateral inhibition (α/β)
   4.1.5 Identification: Local maxima, global maxima, thresholds
4.2 Alternative accounts
   4.2.1 Halle & Vergnaud — An Essay on Stress (1987)
   4.2.2 Dresher & Kaye (1990) — Principles and Parameters
4.3 Quantity-insensitive stress systems
   4.3.1 Maranungku
   4.3.2 Weri
   4.3.3 Paiute / Warao
   4.3.4 Latvian / French / Lakota / Polish
   4.3.5 Garawa
   4.3.6 Lenakel
   4.3.7 Antepenultimate stress
4.4 Quantity-sensitive stress systems
   4.4.1 Eskimo / Koya
   4.4.2 Komi / Cheremis / Mongolian / Mayan
   4.4.3 Winnebago
   4.4.4 Creek "tonal accent"
4.5 Conclusion
CHAPTER FIVE: Phonological Processes / Phonetic Realization Rules

5.1 Sonority-sensitive phonological processes
   5.1.1 Fortition / Lenition
   5.1.2 Phonetic realization of English voiceless stops
   5.1.3 Icelandic vowel length revisited

5.2 Additional phonological processes
   5.2.1 Assimilation / Dissimilation
   5.2.2 Harmony
   5.2.3 Insertion / Deletion / Reduplication

CHAPTER SIX: Learning Procedures in Dynamic Computational Networks

6.1 Methodological preliminaries
   6.1.1 Data bases
   6.1.2 Coding — Input
   6.1.3 Coding — Output

6.2 Learning algorithm — Early efforts

6.3 Proposed learning algorithm — Modified simulating annealing

6.4 Results

6.5 Potential objections

Appendix I: Computer Source Code
   Available on request

Appendix II: Language Data
   Available on request

BIBLIOGRAPHY

167
Chapter 1

Introduction / Historical Review

Forty years ago, major developments in linguistics, information theory, formal mathematics and data processing led researchers to attempt to model linguistic phenomena on digital computers. Ten years later, in 1960, the organizers of the First National Symposium on Machine Translation reported that sufficient progress had been made to offer the hope that the marriage of computer science and linguistics would soon offer working models of machine translation and automatic speech recognition (Edmundson 1960). While the next decade proved this goal to be elusive, the prospects of using computers to model linguistic theories remained promising. In 1970, Morris Halle and M. P. Schützenberger wrote in the preface to The Formal Analysis of Natural Languages, a conference at the Institut de Recherches en Informatique et Automatique:

"(the conference is organized) ... in order to bring computer scientists into contact with linguists concerned with the formalization of natural languages. In bringing these groups together it was the hope that each would benefit from the contact. Although computers have as yet been used only sparingly in linguistic research they clearly are promising tools for research ... As might be expected not all areas are equally advanced. Thus phonology appears to have reached the point where it is possible to focus fruitfully on the abstract character of the formalism and to use it as a tool for further discoveries. (emphasis mine)"

It is in this spirit that the current project is undertaken. Unfortunately, in the twenty years that have elapsed since Halle and Schützenberger offered their rosy assessment we have not witnessed the expected explosion in the computational modeling of linguistic theories. In fact, by 1980, in the judgment of much of mainstream linguistics, the entire enterprise of computational linguistics was not only dead, but thoroughly discredited (Slocum 1985).

At the same time that post mortems for computational linguistics were being written, important developments were making possible a new rapprochement between computer science and linguistics. Both autosegmental and metrical phonology have significantly expanded the representational repertoire of generative phonology as well as reducing the formal distance between representations and the rules that modify them. Natural phonology and related theoretical approaches have reintroduced the importance of sonority as a phonological variable (Vennemann 1974; Kahn 1980; Hooper 1972; Hankamer & Aissen 1974). More recently, harmonic phonology has developed a conception of phonology where rules consist of the attraction toward goals and the

---

1While Chomsky would later disavow any direct connection between his research program and that of the machine translation/computational linguistics community (LSLT 1975, pp. 6-7, 39-40), early generative theory developed in parallel with formal mathematics, computational theory and computer science (e.g., Chomsky 1956, 1958, 1963)

At the same time that the formalist approach to phonological theory represented by early generative phonology has undergone significant modification, computational theory has developed in a somewhat parallel direction. Dynamic and non-linear programming, declarative languages such as LISP and PROLOG, and parallel distributed processing all represent important developments that significantly increase the computational power of computer models while decreasing the formal distinction between program and data. More recently, a variety of connectionist and computational networks have been applied to linguistic analysis with promising results (McClelland & Rumelhart 1981, 1986; Feldman & Ballard 1982; Pinker & Prince 1988; Elman 1989). Within the domain of phonology, there has been a recent explosion of interest, with a wide variety of computational approaches currently competing to offer integrated accounts of phonological phenomena (Halle & Vergnaud 1987; Dresher & Kaye 1990; Dresher 1991; Lakoff 1989; Touretzky 1989 Touretzky & Wheeler 1989; Touretzky & Gupta 1992; Smolensky & Prince 1992). The current study flows out of the confluence of these converging streams in phonological and computational theory.

While the theoretical landscape is quite different from that of the fifties and sixties, and while the resulting changes have been generally beneficial, residual difficulties remain that make a synthesis difficult. Although recent phonological analyses frequently refer to quasi-scalar variables like sonority and degrees of stress, the formalism of generative phonology still makes the representation of such variables quite difficult. While a great deal of attention is now being paid to the issue of phonological representations, it is often accompanied by a commensurate lack of formality in the description of phonological rules. Indeed, phonologists have frequently treated rules and representations as competing ideas rather than as complementary tools within the theoretical arsenal (cf. particularly Stephen Anderson's 1985 history of 20th century phonology that is subtitled, Theories of Rules and Theories of Representations).

Computational theory in general and connectionism in particular faces a complementary problem. Whereas current phonology makes reference to numerical values or activation vectors (e.g. sonority) that can't easily be incorporated into its formalisms, connectionism has difficulty finding numerical activation vectors that can adequately represent linguistic constituents.³ Geoffrey Hinton observes in the introduction

²Ironically, these developments have relaxed precisely those features of the generative formalism that prompted Halle to be optimistic about the modeling of generative phonology on a computer in the first place: the formal distinction between syntax and semantics, between program and data.

to the 1991 MIT volume, *Connectionist Symbol Processing*, that its contributors intended to address —

"...the current tension within the artificial intelligence community between advocates of powerful symbolic representations that lack efficient learning procedures and advocates of relatively simple learning procedures that lack the ability to represent complex structures effectively. The contributors aim to extend the representational power of connectionist networks without abandoning the automatic learning that makes these networks interesting."

Implicit in the discussion is the admission that compositional symbol systems are not congenial (or even possible) domains for connectionist modeling (cf. Fodor & Pylyshyn's 1988 critique). Natural language serves as a paradigm case for the difficulty facing connectionist models, possessing a gross temporal structure that is traditionally represented in terms of a hierarchical constituent structure. Moreover, within generative frameworks, linguistic representations are generated by a combinatorial syntax and modified by derivationally ordered rules. Given the complexity of the domain and putative success of the generative formalism, both advocates and critics of connectionism generally argue that more powerful representational models will be needed in order to adequately account for linguistic phenomena.

We are left with a situation where linguists propose a quasi-numerical representation of their data (e.g. sonority) but lack the computational machinery necessary to process its phonological implications, while connectionists possess the necessary computational machinery but lack the representational theory necessary to quantize linguistic data. In other words, each discipline's problem may be the search for the other's solution. As a result, the representation of phonological phenomena such as syllabification and stress assignment using connectionist-style networks offers an excellent opportunity to test both a phonological theory that incorporates energy measures such as sonority and a computational theory that implements connectionist architectures.

For the last three and a half years, John Goldsmith and I have been examining an approach to linguistic theory construction that uses "dynamic computational networks" to develop and systematically evaluate alternative phonological representations. As a descriptive term, "dynamic computational networks" will be used to refer to simple spreading activation networks that permit the internal state of a unit to be directly influenced by neighboring units. Using this tool, syllabification, phonotactic constraints, stress assignment and a variety of phonetic realization rules have all proven to be amenable to network explanations.

### 1.1 Justification for a dynamic computational model — *Wave Phonology*

One of the tensions that often characterizes contemporary phonological discussion is the purported choice between rules and representation. As was noted above, Stephen Anderson (1985) organizes the recent history of phonology in terms of this opposition.
Contemporary approaches often attempt to distinguish themselves by reducing the role of explicitly stated rules, particularly to the extent that they need to be serially ordered within a derivational framework. For many, the ultimate goal would be to design a phonological representation that can eliminate rules altogether (e.g. Scobbie 1991). Connectionist architectures have historically aligned themselves with this approach, claiming that apparent rules are epiphenomenal to the complex patterns of connections within a network (cf. the title of the 1991 New Hampshire conference — _Language With or Without Rules?_; Rumelhart & McClelland's past-tense network, 1986).

The proposed model represents a moderating position on the rules/representations dichotomy (Goldsmith & Larson 1990). Goldsmith's conclusion in _Autosegmental & Metrical Phonology_ (1990a), for example, expresses the desire to retain but redefine the role of phonological rules:

With respect to the notion of rules, throughout most of this book we have retained the traditional generative conception, according to which rules come with a structural description and apply if that description is met. ... I believe that this notion stands in need of serious revision, although, as we have seen, ongoing research in phonological theory has been able to enunciate a powerful conception of phonological representations, independent of any changes in the theory of rules. Now, however, ... we may proceed to a novel and even more compelling picture of phonology, in which rules interact with phonotactic conditions on a small number of levels to develop representations at each level satisfying the conditions stated there.

In phonology, the model we arrive at is one that looks more like a model of chemistry than the models of classical generative phonology, in which the phonological grammar resembled nothing more than a computer program. In the model that is emerging currently, representations have a complex geometric structure, but relatively few degrees of freedom in the changes they may undergo. Rules define possible changes in the structure of the phonological material, and in each and every case, the changes are motivated by an attempt to achieve a greater satisfaction of well-formedness conditions. This bears a striking similarity to the notion that chemical systems tend toward a lower energy level, consistent with the properties that they have. The application of this kind of model has been urged elsewhere in cognitive studies by Smolensky (1986), for example, and the convergence of work in phonology with that in other areas of cognitive science offers great hope for continued advances of the sort that we have seen in phonology in the last fifteen years (pp. 331-332).

As the project was originally conceived, we imagined building connectionist models similar to Smolensky's harmony machine (Smolensky 1988; McMillan & Smolensky 1988; Legendre, Miyata & Smolensky 1990; etc.) or Touretzky's BoltzCONS network (Touretzky 1989; Touretzky & Wheeler 1989; Touretzky, Wheeler & Elvgren 1990; cf.
Lakoff 1988). Unlike the associative network models that replace rules with a simple weight matrix à la Rumelhart & McClelland's past-tense model, harmony machines incorporate "rules" as a set of soft constraints in a multiple constraint satisfaction or energy minimization problem.

While such networks hold great promise for processing systems of variable constraints (or rules), they do not, in fact, address the kinds of concerns that are introduced when we attempt to manipulate the phonological variables that are appropriate to the description of syllabification or stress assignment. Ultimately the issue seems not to be the role of rules vis-à-vis representations but rather the type of rules or representations that need to be manipulated within a phonological grammar. Connectionist models are not without their "rules," encoded in restricted patterns of connections, activation formulas and learning algorithms. A computational device that processes numerical activation vectors is different from one that manipulates formal symbols with a combinatorial syntax, but both utilize rules that act on a phonological representation.

As a result, we must look in a different direction to find the contrast between the proposed model and other competing approaches. Goldsmith (1990) contrasts two approaches to the representation of syllables, a syntactic phrase-structure model that is characterized as the external representation of the syllable, and a sonority approach that characterizes the internal representation of the syllable. In the next chapter we will similarly describe two alternative conceptions of syllable structure: a constituency model in which syllable representations and the rules that modify them are described in terms of formal symbolic atoms governed by a combinatorial syntax, and a sonority model in which the energy relationships present within and between syllables are computed.

When one considers the contrast in terms of underlying metaphors, we find ourselves moving from the domain of classical particle physics to quantum mechanics. One of the most provocative metaphors for describing linguistic phenomena is to be found in a seldom-cited essay written over 30 years ago — Kenneth Pike's 1959 essay entitled, "Language as Particle, Wave and Field." Rather than drawing a contrast between rules and representations, the essay contrasts different conceptions of the phonological "stuff." Drawing inspiration from the particle-wave duality that informs quantum mechanics, Pike identifies linguistic correlates to each of the three basic conceptions of physical reality. He writes:

These three views of language can be summarized in different terms. Language, seen as made up of particles, may be viewed as if it were STATIC — permanent bricks juxtaposed in a permanent structure, or as separate 'frames' in a moving-picture film. The view of language made up of waves sees language as DYNAMIC — waves of behavioral movement merging one into another in intricate, overlapping, complex systems. The view of language as made up of field sees language as FUNCTIONAL, as a system with parts and classes of parts so interrelated that no parts occur apart from their function in the total whole, which in turn occurs only as the product of these parts in functional relation to a meaningful social environment (p. 129).
Pike goes on in the essay to lament the fact that contemporary approaches (à la 1959) to linguistics in general and phonology in particular rely almost totally on the static *particle* view of language. While the movement from *item and arrangement* (IA) to *item and process* (IP) and ultimately to generative-transformational descriptions of phonology move from the purely static to the processual, even generative phonology and its descendants tend to view the terminals that appear in the structural descriptions of phonological rules as discreet *particle-like* entities or constituents that possess enumerable properties and which can enter into a variety of relationships with each other.\(^4\)

Pike goes on to observe that several phenomena in language can only be accounted for when one observes *wave-like* energy relationships within the linguistic utterance. Both syllables and metrical feet exemplify a periodic rise and fall in energy. The sonority hierarchy that has been observed for nearly a century describes a bell-shaped distribution where the nucleus comprises a high-energy peak with consonants tailing off to the margins (Sievers 1893; Jespersen 1904). The preference within the world's languages for CV syllables as well as the preference for binary metrical feet represent phenomena that are best accounted for using computational devices appropriate to the discussion of energy relationships. While the computational machinery necessary for the description of particles has been well established for decades, no comparable models have been offered for the description of waves.\(^5\) The dynamic computational networks proposed in the following chapters are specifically designed to fill this vacuum, offering a computational device that can process the energy relationships within the syllable and/or metrical waves.

As a result, we can perhaps label two (or perhaps three) approaches to phonology: *particle phonology* (with apologies to Sandy Schane) and *wave phonology*. While there may be a strong temptation to embrace one and criticize the other, both may, in fact, serve as complementary metaphors that are each capable of explaining certain phonological phenomena. A theory that attempts to account for wave-like phenomena in language is not necessarily inconsistent with theories that attempt to account for particle-like phenomena. Just as quantum mechanics accounts for the behavior of electromagnetic radiation by reference to a particle-wave duality, the choice between particle analyses and wave analyses in phonology may be a both-and rather than an either-or proposition. Pike does not propose that waves supplant particles as linguistic primitives but that they rather enhance our descriptions. He concludes,

The solution to our problem does not lie in this direction (i.e., eliminating particles). Rather, a view of the multiple structuring of data, in three particle hierarchies of phonology, lexicon, and grammar, must be retained.

---

\(^4\) Sanford Schane's "particle phonology" (1984) is an explicit attempt to cash out the implications of the particle metaphor with respect to phonological description.

\(^5\) With the recent development of cognitive linguistics (cf. Brugman & Lakoff 1988; Langacker 1986) with an attendant application of associative connectionist models to the problem of describing semantic fields (Cottrell 1987, 1988; Harris 1989), the *field* approach to language has also developed an appropriate computational architecture (cf. "Field Computation in the Brain," MacLennan 1992). The description of a distributed memory where information is encoded in "the pattern of connections" rather than any particular memory address also proves amenable to the field metaphor.
But this hierarchical-particle approach must be supplemented by wave and field outlooks for providing dynamic and functional components within the analysis. . . . Now following this experience, I am convinced that all forms of human behavior can be studied in terms of a hierarchy of particles, a sequence of waves of events, and a background field within which there are manifested concentrations of energy that we call events (p. 143).

In light of the possibility that particle phenomena and wave phenomena are simply alternative conceptions of the same reality, the current project will defend the need for incorporating tools that describe and manipulate energy waves. While the terms are not strictly equivalent in either extensional or intensional terms, during the course of our discussion we will alternately refer to the sonority model as a computational model or a wave model of the syllable. In the same way, we will somewhat interchangeably refer to the constituency model as a syntactic or particle model.

1.2 Evaluative criteria

The question that faces us is what would count as a demonstration of the value of the sonority/computational/wave model. We propose that meeting any of the following four conditions would provide a warrant for the approach, with each successive condition offering a weaker justification.

1. Computational (sonority-sensitive) constraints can exhaustively account for the phonological phenomena under analysis. Alleged constituency-sensitive constraints prove to be either incorrect, epiphenomenal or derivative.

2. One or more significant phenomena relating to the syllable can only be accounted for by reference to sonority-sensitive constraints or processes, even though other phenomena may well depend on constituency relationships. Such a theory should be able to translate between or at least integrate the two modes of representation.

3. While syntactic processes and constraints can exhaustively account for all of the phonological phenomena under analysis, one or more of these processes can be translated into computational terms and thereby be processed more efficiently and/or be made amenable to an automatic learning process.

4. One or more phonological processes can be accounted for using the computational model.

While it arguably would be sufficient to demonstrate the empirical adequacy of the network approach over an appropriately broad range of linguistic data, we will attempt throughout to compare network analyses to competing analyses both within the symbolic and connectionist traditions. One of the advantages of the network architecture will be
the availability of explicit measures of its power and its descriptive adequacy (cf. Chomsky & Halle 1968; Halle 1970).

1.3 Outline of the dissertation

In the following chapters, we will describe the architecture of the dynamic computational network, test it over a variety of phonological domains, and then describe a learning procedure that permits automatic training of the network given a limited corpus of data. The second chapter will describe the architecture of the various components of the model. Chapters three through five will apply the model to a variety of linguistic phenomena—syllabification and phonotactic constraints (chapter three), stress assignment (chapter four), and phonetic realization rules (chapter five). Chapter six will examine connectionist learning procedures that allow the network to discover inductively the coefficients that characterize the grammar or knowledge of the system. Readers who are more concerned with the linguistic analyses than the mathematical description of the model should scan the second chapter and then focus on chapters three through five.
Chapter 2

Architecture of the Model

Under the rubric dynamic computational networks we intend to encompass a rather broad range of models that permit the representation of phonological information through the use of a small number of numerical activation units (Rumelhart & McClelland 1986; Smolensky 1988; Goldsmith 1990). Before describing the architecture of the proposed model, several observations concerning the choice of titles are appropriate. While we have avoided the label "connectionist", the networks do, in fact, belong to this general class. Several considerations motivate the avoidance of the term. While connectionist networks can and do possess a wide variety of architectural designs (cf. contributions in Hinton 1991), in common parlance the term is typically applied to associationist learning networks that presuppose significantly less temporal and symbolic structure than the proposed networks (Rumelhart & McClelland 1986). The networks which will be proposed are particularly designed to account for the wave-like phenomena described in chapter one (syllables, metrical feet, etc.) while associative nets more naturally model the field-like properties of linguistic phenomena (semantic networks, multiple constraint satisfaction, etc.). The term "connectionism" also tends to evoke in the reader's imagination not just a computational architecture, but a host of philosophical commitments, witnessed by the Cognition debate and particularly the critique by Fodor & Pylyshyn (1988). While we are sympathetic with Smolensky's response in "On the Proper Treatment of Connectionism" (1988), we would rather sidestep the issue altogether, at least provisionally. We view computational networks first and foremost as a tool of linguistic research—a tool which may well influence how and why linguistic theories are developed, but one which can potentially model a wide variety of theoretical positions. As was noted in chapter one, we also owe a large debt to both the connectionist and symbolic paradigms, receiving from the former a powerful repertoire of computational devices and learning algorithms and from the latter a wealth of insight concerning the complexity of phonological representations and the processes by which they can be modified.

The title, dynamic computational networks, does identify several positive commitments as well. First, phonological phenomena are viewed as potentially representable by a network of relatively simple units. As noted in chapter one, current phonological theories should be amenable to the network metaphor, both in the fact that they often utilize quasi-scalar constructs such as sonority and variable degrees and stress and in the fact that segments are subject to contextual influences. Autosegmental and metrical representations are already network-ready in that they encode contextual relationships as a complex of associated units (autosegments, grid marks, etc.). Even standard derivational models typically propose a mapping between underlying and surface representations that is mediated by one or more intermediate levels. The major distinction between the current model and these theoretical precursors is the fact that the current network is a computational network. Rather than manipulating formal objects syntactically, the current approach attempts to quantize the relationships between units. Sonority is the prototypical quasi-scalar phonological variable and, as such will play an
important role in the network. Stress has also been represented as a scalar variable and is therefore also amenable to network analysis (Chomsky & Halle 1968). Even binary features can be thought of as quantized property of a segment. Finally, the network is dynamic rather than static. Rather than representing phonological generalizations as static properties of individual phonological atoms, important properties emerge from the dynamic interactions of several units.

It is also our goal to provide a naturally constrained theory that uses as little external stipulation as possible. One of the advantages of computational model is that we will be able to explicitly describe its architecture, and in so doing, provide appropriate measures of evaluative adequacy. The power of a connectionist network can vary in several dimensions: 1) the gross number of units in the network; 2) the patterns of the connectivity between units (i.e. number of connections per unit and segregation of units into layers or subpopulations); 3) the functions which determine the internal activation of a unit and the activation passed between units; 4) the number and type of exogenous inputs to the network; and 5) the learning algorithms which modify the weights in the network.\(^1\) In simulations using a traditional serial computer architecture, we can also measure variables such as CPU processing time, memory requirements, etc. By offering an explicit architectural design we will be able to propose an appropriate evaluation metric, since we can identify moves that increase power, processing time, computational complexity, etc.

2.1 Network Architectures - A Brief Survey

In order to develop a rationale and description for the proposed network architecture, we will begin with a unstructured network and then describe the structural modifications required to deal with phonological phenomena. In the least structured case, a computational network can contain an arbitrary (but fixed?) number of units exist in an \(n\)-dimensional space.

(1) \(N\)-dimensional auto-associative network

\(^1\)Strictly speaking, the adequacy of the network description is independent of the learning algorithm that is used to set weights. One could imagine a theory which stipulates certain of the weights in the network as part of a universal grammar (UG).
Each connection encodes a simple algebraic function that serves either to activate or to inhibit the affected unit's internal state. In the simplest case, the activation value of a unit represents a weighted sum across all of its connections.

\[ a_i^{t+1} = \sum_j w_{ij} o_j \]

For any unit \( i \) in the network, its activation \( a_i \) at time \( t+1 \) is equal to the weighted sum \( w_{ij} o_j \) of all units \( j \) connected to \( i \). The computational cost of such a auto-associative network would be directly proportional to the number of required units, with a superior system being determined by its ability to capture a large number of patterns relative to the number of units in the network.

Connectionist networks that have been applied to experimental tasks typically impose a variety of limitations on this general model. For example, rather than have each unit connected to potentially any other unit in the \( n \)-dimensional space, the units are often segregated into layers of units, where an arbitrary number of connections can be made between two adjacent layers (and potentially within a single layer) but limitations are placed on which layers are connected to which. Typically a network will have a layer of input units, a layer of output units, and possibly, one or more intervening hidden layers.

(3) **Feedforward Network (with hidden units)**

A recurrent network adds to the complexity of the feedforward network by allowing bi-directional connections between units on the same level as well as between units on different levels. Typically these intralevel connections are inhibitory (competitive) while interlevel connections are excitatory. In some systems, the inter-level connections are bi-directional as well, permitting feedback as well as feedforward connections (bottom-up vs. top-down processing; cf. PDP I). The segregation into layers permits specialized input and output units as well as allowing the network to develop its own representation for hidden units. Even with bidirectional recurrent connections, they also encode fewer connections than the \( n \)-dimensional network described above, and as a result, impose less computational cost. Of course, the additional structure must be
accounted for, either as a innate endowment of the system or through some self-organizational principle.

In addition to a pattern of connectivity (i.e. which units and layers are connected to which), a network is characterized by activation functions which determine how a unit’s internal activation is computed from the inputs which it receives from neighboring units. In the simple activation function in (2), a unit simply sums all of the activations that it receives from its neighbors, each weighted by an excitatory or inhibitory coefficient \( w_{ij} \) associated with the connection from \( j \) to \( i \). A host of more complex activation functions are available, however. For instance, a unit could have a \textit{threshold} associated with it so that it would be on if its input exceeded a certain amount and off otherwise. Alternatively, it could impose a \textit{logistic} function that increases the probability that a unit is on when the net input is highly positive and decreases the probability when the net input is highly negative. For example, Rumelhart, Hinton and Williams (1986) describe a thermodynamic network with the following activation function:

\[
(4) \quad p(a_i) = \frac{1}{1 + e^{-(\sum \sum w_{ij} \eta_j - \theta_j) / T}}
\]

In addition to basing the probability of a unit being on on the input from its neighbor, the formula in (4) includes a \( \eta_j \) term that permits exogenous input to the system to directly affect a unit (cf. 2.4.2), a \( \theta_j \) term that introduces a threshold (see 2.4.3) and a parameter \( T \) that indicates the \textit{temperature} of the network (see 6.4). The effect of temperature on the logistic function can be observed in (5), where we plot probability in terms of net input for four different temperatures. While high temperatures flatten the curve, making the network’s behavior more random, low temperatures cause the network to approximate a step function where positive input creates nearly 100% probability that the a unit will be on while negative input insures that it will be off. The horizontal line in the graph indicates the position where behavior is purely random (prob. = .5).
Figure 1.5 Effect of temperature on probability

A simpler logistic function can determine the probability of a unit being on solely from its input, as in (6):

\[
p(q_i) = \frac{1}{1 + e^{-\sum_{j} w_{ij} q_j}}
\]

Other alternative activation functions that can introduce non-linearity into a network could include linear threshold functions, step-functions and maxima/minima detectors (see 2.4.3), squashing functions (PDP I, pp. 485-489; Goldsmith 1990), and Boolean functors such as AND, OR, XOR, and NOT, though in the latter case, the necessary computations are likely performed by subnetworks consisting of several units (Rumelhart, Hinton & Williams 1986).

In addition to varying in complexity (and potentially, neural plausibility), the activation functions dictate the type of values that a unit might take. The simple activation function in (3) allows units to take continuous scalar (i.e. real) values. A squashing function retains the scalar or analog character of the units but places them in a subscribed range. A step-function places an additional limitation by insuring that units take integer or even whole-number values. The strongest limitation is enforced by those functions that digitize the activation value of a unit. The important observation is that computational networks can be designed to process either digital data, analog data or both. In addition to evaluating the complexity of a network based on the number of units and their pattern of connectivity, we can evaluate the processing complexity of the activation functions that the network encodes.
In addition to varying the architecture of connections within the network, connectionist frameworks propose a variety of procedures by which the connection weights can be set, modified and/or learned. In fact, as was noted in chapter one, the principal change that re-energized the connectionist enterprise was the development of powerful learning algorithms such as back-propagation and simulated annealing (see chapter six). Unfortunately, even with such powerful computational and learning processes, the standard three-layer network still lacks the representational power to deal complex symbolic domains such as those entailed by linguistic phenomena (cf. Hinton 1991).

In order to deal with the phonological representations that are of concern to linguists, several additional modifications are proposed for the general network architecture. Phonological strings are complex symbolic objects, possessing multidimensional temporal structure (cf. autosegmental phonology) and arguably, a hierarchical constituent structure and internal syntax. And while constructs like sonority and relative degrees of stress suggest scalar variables, representing a phonological string with a numerical activation vector is anything but straightforward. As a result, we propose the following three integrated modules to account for featural, syllabic and metrical phenomena:

2.2 Dynamic Computational Network

As currently devised, the dynamic computational network comprises three modules which interact with each other in both feedforward and feedback relationships, thereby permitting both bottom-up and top-down processing.
At the base of the network is an *autosegmental* (or featural) network which takes as its input a distributed representation of the segment and outputs an inherent sonority coefficient which serves as the primary input for syllabification. The core of the model is a *segmental* (or syllable) network which takes the inherent sonority communicated from the autosegmental network as its principal input and provides a unique syllabification as its primary output. The model also contains a *metrical* network that generates a surface stress pattern given lexical, positional and quantity-sensitive input activation.
2.3 Autosegmental/Featural Network

The autosegmental network has three principal functions in the current model. First, and most critically, it computes an inherent sonority value associated with each segment in a language's phonological inventory. Each column of units contributes to a sonority vector based on the featural representation of the segment. Second, it permits a computational representation for a number of phonological processes (assimilation, dissimulation, spreading, cf. chapter five). Finally, the interface between the segmental and autosegmental networks provides a potential explanation for autosegmental licensing (Goldsmith 1990, Bosch 1991).

One of the most important architectural features of the network is the fact that its units are segregated not only into horizontal layers but also into vertical columns that correspond to the string's skeletal or timing slots. Rather than adopting a fully distributed representation where both features and position are distributed among a number of units (e.g. Wickeiphones, Rumelhart & McClelland, PDP II), the proposed network adopts what Smolensky has called a semi-local representation (Smolensky 1990). In a semi-local representation, one of the dimensions (i.e. time) is represented by atomic units (i.e. segments), while the other dimension has a distributed representation (i.e. features). In the autosegmental network we have a localist representation of time (one inherent sonority vector per segment in the string) but we potentially have a distributed featural representation for each segment in the phonological inventory. McClelland & Rumelhart (1981) propose a similar architecture for interactive letter recognition. While each letter in the English alphabet is decomposed into a set of letter features, the entire set of letter features is reproduced for each position in the word string.

The autosegmental network can utilize a variety of featural representations for the segment. The most agnostic would be to provide a purely atomic or local representation for each unique segment/phoneme in a language's inventory. Assuming an inventory of 44 phonemes in a language such as English, 44 separate but comparable units would have to be replicated for each position in the string (cf. vertical columns of units in (7) above). Assuming a permanent hard-wired network that uses a fixed mapping to process word-length strings (cf. Hinton 1991), the localist proposal would result a very large number of basically redundant units since only one of every 44 would be active. Whether the sonority associated with each of units is stipulated or learned, one would also have to

---

2The use of the terms autosegment or feature to refer to the units associated with each segment reflects a somewhat idiosyncratic approach to constituency and association. Several phonologists have distinguished between lines in phonological graphs that represent constituency and those that represent association. Itô (1989) notes that melodic (or autosegmental) structure "involves relations of predication," while metrical structure illustrates constituency relations. Brentari & Bosch (1990) argue that it is a mistake to confuse the two types of relationship. Whether or not traditional symbolic analyses require a strict differentiation, the network description will not distinguish between the segment/syllable relationship and the feature/segment relationship.

3The formulas represented in generative phonological adopt a similar format, with each symbol in the string representing a feature matrix (cf. Chomsky & Halle 1968).
account for the fact that a segment has the same inherent sonority regardless of its serial position in the string. The number of units in such a network can be reduced in a variety of ways. First, it is possible to use a featural representation that would use a relatively small number of units to represent each segment. For instance, one could search for the minimum number of binary features that could account for the phonemic inventory of a language (e.g. \( \log_2 44 = 5.46 \); 6 binary features required). In the featural analyses of Jakobsen, Fant, & Halle (1952) and later, Chomsky & Halle's *Sound Pattern of English* (1968), mathematical efficiency was balanced with criteria such as articulatory and acoustic plausibility, identification of natural classes and simplicity in rule formulations to result in the postulation of roughly 18-20 distinctive features for English.

Given a rather standard set of SPE-style binary features we could then propose a sonority vector that is determined by dot product of the featural specification of a segment (+Feature represented by 1, -Feature represented by 0; cf. Goldsmith & Larson 1990) and a weight vector that determines how each feature contributes to sonority. For example, the inherent sonority for a /d/ could be represented as follows:

(8) **Sonority Vector (featural representation)**

\[
(\pm\text{syl}, \pm\text{con}, \pm\text{son}, \pm\text{cont}, \pm\text{voiced}, \pm\text{labial}, \pm\text{ant}, \pm\text{cor}, \pm\text{back}, \pm\text{strid}, \pm\text{nasal}) \\
(0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0)
\]

\[
(2.62, 0.23, 1.62, 1.34, 1.01, 0.52, 0.03, -0.24, 0.62, 0.78, 0.92)
\]

If we think of the above in terms of an input activation function, we have an example of basic formula described above in (3).

(3) \( a_i^{t+1} = \sum_j w_j o_j \) where \( o_j \in \{0, 1\} \)

In addition to testing the network with SPE-style features, the network can process a wide variety of alternative featural systems, whether unary (cf. Schane 1984, Kaye, Lowenstamm & Vergnaud 1985), binary, ternary or scalar. In the discussion of learning in chapter six we will alternative feature systems where self-organizing networks can discover some or all of the features that specify a set of segments as well as determining their contribution to the sonority vector. In short, network analyses should be evaluated independently of the choice of some particular set of distinctive features. In order to provide a common point of reference, the analyses offered in subsequent chapters will use

---

In chapter three we will argue that derived sonority is contextually dependent. This contextual influence is not generally based on serial position, however, rather being dependent on the sonority of neighboring units (e.g. high sonority peaks). Goldsmith & Larson (1992) consider evidence that input sonority may be different for word-initial or word-final positions, but the effect is marginal. It would also be possible for a network to assign unique inputs to word-initial and word-final positions without changing the sonority vector associated with the segments that filled those positions.
the same feature sets as the competing alternatives. If no feature system is stipulated or assumed by a particular analysis, the network account will simply use a localist representation in order to insure that no TRIC's arise in the analysis (cf. Lachter & Bever 1988). It should be noted, however, that the ability of the network to encode (or induce) a wide variety of featural systems suggests that it may indeed offer a tool to evaluate the merits of alternative featural systems.

With any of the distributed alternatives discussed above we still have the problems created by a fixed mapping that posits a separate column of units for each skeletal position in the string. Hinton (1991) discusses a timesharing alternative to the fixed mapping architecture that solves both the redundancy and the learning problems. Since the syllable network requires the same sonority computation for each segment, we could create a single set of units (9) that would be serially accessed for each segment in the string.

(9) Autosegmental subnetwork

\[
\begin{array}{c}
\text{Inherent sonority} \\
\text{Features} \\
\end{array}
\]

While one might object to the serial access of the autosegmental network, the situation is no different for the network as a whole (or for most any network proposed in the literature). The same argument that might motivate us to adopt a fixed mapping architecture for the autosegmental network would have to apply to larger phonological units as well (e.g. how can network resources be reused for each word in an utterance?). Whether a network processes one segment at a time, one word at a time, one phrase at a time, or creates a dynamic window that continuously buffers and processes strings, the model has to have access to a time-sharing procedure that allows components of the network to be reused. The real challenge is to calibrate the different time slices that the various network modules encounter (features, segments, syllables, words, metrical phrases, etc.). While such a calibration presents a difficult computational challenge that requires further discussion, we will simulate the interfaces with serially ordered procedures.5 Each time the syllable network confronts a segment, it will compute its present inherent sonority value from a matrix such as Figure 10 (during learning it is supposed that the weights in the second row are subject to continuous updating).

---

5 In Marr's (1982) tripartite distinction between computational theory, algorithm, and implementation, the computation of sonority from autosegmental features is an implementational solution suggested by the seriality of the conventional computer architecture on which the network is simulated.
(10) Featural Weight Matrix

| feat. seg. | son. | cont | voiced | ant. | cor. | nasal | ...
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>weight</td>
<td>1.62</td>
<td>1.34</td>
<td>1.01</td>
<td>.03</td>
<td>-0.24</td>
<td>.92</td>
<td>...</td>
</tr>
<tr>
<td>b</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>t</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>d</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>m</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>l</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

While the computation of sonority can be accomplished by a timesharing procedure that does not require a separate representation for autosegmental features for each timing slot, other applications of the autosegmental network might require them to be permanently recorded and left manipulable. Chapter five will introduce a brief discussion of phonological rules such as assimilation, dissimilation, vowel harmony, and feature spreading that can potentially be represented by positing excitatory and/or inhibitory connections between neighboring units within the autosegmental network.

(11) Autosegmental Network (interconnections)

In the above network, assimilation, feature spreading and multiple associations would be represented by positive connection weights; dissimilation or blocking by negative connection weights and inert features by zero weights. Theories which implement feature
geometry could be modeled by further defining the interconnections between rows of units.

2.4 Syllable Network

The syllable network forms the core of the model. Stated simply, the network has three layers: input, processing, and recognition. As input, it takes from the featural network the inherent sonority values associated with each segment in the string. In the processing layer, the network computes a derived sonority value for each segment dependent on its own inherent sonority and contextual influences from neighboring segments in the string. As its primary output, the network provides a unique syllabification for the string by identifying maxima and minima of derived sonority.

**SYLLABLE NETWORK**

**Level-2: RECOGNITION**

*Local Maxima*

**Level-1: PROCESSING**

*Derived Sonority*

**Level-0: INPUT**

*Inherent Sonority*

![Diagram of Syllable Network](image)

**Figure 2.12**

2.4.1 Input - Inherent Sonority

Considering the input level first, the principal variable in the syllable network is the sonority coefficient associated with each segment. These inherent sonority values, the output from the autosegmental/featural network below (2.6), can either be stipulated or learned. In our early work with the network, sonority was simply stipulated in accordance with the "universal" sonority hierarchy (cf. chapter three; Larson 1990; Clements 1987; Dell & Elmedlaoui 1985; Jespersen 1904). One of the advantages of connectionist modeling, however, is the existence of powerful learning algorithms that make it possible to induce appropriate sonority values from a corpus of data (cf. chapter six). In addition to allowing the network to learn a potentially language-specific sonority hierarchy, the network also permits non-integer values. Traditional quantitizations of the sonority...
hierarchy simply assign an integer value to each successive major class, a procedure that eliminates potential distinctions within classes and that treats each successive class difference as equidistant (i.e. the Minimum Sonority Distance Principle, Harris 1983; Levin 1985).

2.4.2 Processing

If the network relied solely on the input sonority values of the immediately dominated units (a static property of segments), the syllabification predictions of the network would still be very similar to traditional sonority analyses. Chapter 3 will discuss several phenomena, however, where the traditional sonority hierarchy fails on either descriptive or explanatory grounds to appropriately account for the data (cf. Clements 1987). As a result, the network incorporates a dynamic that permits the derived sonority of a segment to be influenced by its environment. This contextual influence can come from three sources:

Lateral inhibition/excitation from neighboring processing units (left and right)
Feedforward inhibition/excitation from neighboring input units (bottom-up)
Feedback inhibition/excitation from neighboring recognition units (top-down)

2.4.2.1 Lateral inhibition/excitation

The influence that we have examined most thoroughly is lateral inhibition operating recursively between a unit and its left and right neighbors. Each unit in the processing layer is connected to its left neighbor with an alpha-weighted connection and with its right neighbor with a beta-weighted connection.

\[(13) \quad d_i^{t+1} = u_i + \alpha \cdot d_{i+1}^t + \beta \cdot d_{i-1}^t\]

In the above equation, \(d\) denotes the derived sonority of a given unit, the subscript indicates its positional index, the superscript denotes the time index, \(u\) denotes the inherent or underlying sonority, \(\alpha\) is the coefficient of leftward inhibition, and \(\beta\) is the coefficient of rightward inhibition. The network is considered to be at equilibrium when

\[(14) \quad d_i^{t+1} - d_i^t < \Delta\]

for any \(i\) in the string, where \(\Delta\) is an arbitrarily small coefficient of change. When \(|\alpha| + |\beta| < .50\) the network reaches equilibrium after a very small number of iterations (Prince 1991). The rapid convergence of the network is due to the fact that the magnitude of change for

---

6This default situation can be simulated within the network by setting alpha and beta at zero. As a result, the network can be used to simulate and test the descriptive adequacy of traditional sonority analyses with either universal or language-specific, stipulated or learned, integer or continuous sonority values.
each successive iteration is proportional to the standard polynomial expansion:

\[ \alpha = \alpha + \beta \quad \text{1st order approx.} \]

\[ \alpha = \alpha^2 + 2\alpha \beta + \beta^2 \quad \text{2nd order approx.} \]

\[ \alpha = \alpha^3 + \alpha^2 \beta + \alpha \beta^2 + \beta^3 \quad \text{3rd order approx.} \]

Generally speaking, a segment's immediate neighbors affect its sonority proportional to \( \alpha \) and \( \beta \). Segments which are two segments distant affect its sonority by \( \alpha^2 \) and \( \beta^2 \), three distant by \( \alpha^3 \) and \( \beta^3 \), ... \( n \) distant by \( \alpha^n \) and \( \beta^n \). Unless \( \alpha \) and \( \beta \) are significantly greater than .50, successive terms rapidly approach zero. The terms in the expansion which contain both \( \alpha \) and \( \beta \) (e.g. \( \alpha \beta \) in the 2nd order approximation) reflect the fact that if both \( \alpha \) and \( \beta \) are non-zero, a segment has an influence on itself. If \( \alpha \) and \( \beta \) have the same sign, the recurrent net will cause a unit to marginally increase its own activation. If they have different signs, a unit will naturally decrease its activation. While this marginal effect (2nd order approximation and smaller) typically has little measurable effect, it may, in fact, make a difference at the edges of a string. Edges are unique not only in the fact that they have only one segmental neighbor (if words are viewed in isolation) but also by the fact that potentially no structure whatsoever neighbors the unit. If no boundary units are present, the autologous effects of the network would be cut in half for these edge segments. While the implications of having potentially infinite or boundary structure (i.e. word boundaries - #) will be discussed in chapter five, we generally assume that no units precede or follow the string under analysis. In practical terms, the same edge effects usually occur whether or not additional boundary units are postulated.

To illustrate the effect of lateral inhibition in more concrete terms, we consider a network with only leftward inhibition (\( \alpha \)). Given a network where \( \beta = 0 \) and a nine segment string where the rightmost unit has an inherent activation of 1 and other segments have an inherent activation of zero, we can illustrate the effect of various values of \( \alpha \). In the following graphs, the shaded bars indicate the inherent or exogenous activation that is supplied to the network while the solid lines and numerical values illustrate the derived activation produced by inhibition/excitation of neighboring units.\(^7\)

\(^7\)While the derived activation values are plotted in a continuous graph, Jordan Pollack (p.c.) has correctly observed that the units are discrete. The intervening lines, whether plotted as a smooth curve or as straight lines are for illustration purposes only (similar to the way the sonority wave of a syllable is frequently represented as a continuous curve, c.f. Clements 1987).
Even with a only single activation pulse in the network, several interesting effects can be observed. With a negative \( \alpha \), an alternating waveform is produced. The alternating pattern of peaks and valleys produced by \( \alpha = -1 \) (2.16) models a pervasive phonological pattern exemplified in such diverse domains as the sonority sequencing principle, a widespread preference for CV syllables and alternating stress patterns (perfect grid). As \( \alpha \) moves from -1.0 to 0.0, the waveform is progressively damped to the point where no inhibitory effects are seen (4.16).

If \( \alpha \) is greater than zero, however, a single pulse is propagated the entire length of the string (2.18). When \( \alpha \) decreases from 1.00 to 0.00, the strength of the propagated pulse decreases proportionate to \( n^\text{th} \) power of \( \alpha \) as one moves \( n \) units distant from the original pulse (2.19).
Lateral inhibition also produces significant deformations in the sonority wave when applied to strings with a variety of sonority values. Given a hypothetical five-unit string comprised of units with the inherent sonority profile - 1 5 9 5 1, a sufficiently negative alpha would deform the left side of the waveform (onset) to the point where the derived activation of the first unit would exceed that of the second unit (2.20). Similarly, a sufficiently negative beta deforms the right side of the waveform (coda, cf. 2.20).

Smaller negative values for alpha and beta deform the wave but do not produce the anomaly between the edge unit and its neighbor (2.22, 2.23). In the five unit configuration the degree of deformation ultimately depends on the absolute magnitude of alpha/beta, the magnitude of the third unit and relative difference between the peripheral units.
While negative values of $\alpha$ or $\beta$ serve to decrease the number of units that can "fit" into a single wave, positive values increase the number. Given an inherent sonority profile 3 1 9 5 1, a sufficiently positive alpha can increase the derived activation of the second unit to the point that it exceeds the activation of the first unit (2.24). As alpha or beta take progressively greater positive values, their effect gradually overcomes the rise and fall of the inherent sonority values (2.25).  

When both alpha and beta are permitted to take non-zero values, a wide variety of complex deformations can result in the sonority waveform. Because of the interactions of the two within a recurrent network, it is important to determine the pairs of values which will permit the network to converge in a reasonable number of iterations. As noted above, Prince (1991) and Goldsmith (1991) have plotted ranges of convergent values for alpha and beta.

---

8While the cumulative effects of large positive values of alpha and beta do not seem to appear in syllabification phenomena, they will become important in the modeling of several stress systems which assign stress to a particular syllable independent of the weight of any of the syllables in the word (see 4.3).
More importantly, the complex interactions make it rather difficult to predict in advance which values of alpha, beta and sonority coefficients will correctly model the syllabification phenomena in a given language. This empirical issue will become the focus of discussion in chapter three where the descriptive adequacy of recurrent networks will be tested and even more so in chapter six where automatic learning procedures will be considered for the discovery of appropriate coefficients.

2.4.2.2 Feedforward inhibition/excitation

In the architecture described above, derived sonority is dynamically computed by permitting the activation of neighboring units to interact with a single exogenous input (inherent sonority from the autosegmental network). The effect of neighboring units can be incorporated in other ways as well. Rather than encoding a single input connection, the network could include an arbitrary number of connections from the left and right neighbors of the immediately dominated input unit.

(27) Recurrent Network
(28) Feedforward Network
It is generally the case that a complete set of connections between the input and processing layers could provide a single step feedforward alternative to the recurrent net described in 2.4.2.1 (essentially trading space for time). Such a complete set of pre-wired connections would be impractical, however, when one attempts to process strings of variable lengths. Fortunately, adequate performance can be demonstrated with a minimum number of additional input connections (see 6.3). In fact, positing input connections to a single left neighbor (\( \alpha_1 \)) and a single right neighbor (\( \beta_1 \)) provides an adequate first-order approximation to the output of the recurrent net. Since alpha and beta are typically rather small, second and third-order terms typically contribute little to derived activation. Moreover, the feedforward net processes information much more quickly, does not face the limitation of non-convergent values, and is more easily inspectable. Finally, it would be possible to account for longer distance effects by adding additional left and/or right connections that would be independent of the mediated effects of lateral inhibition in the recurrent net. For instance, the magnitude of an additional leftward connection (\( \alpha_2 \)) could be either greater than or less than the \( \alpha^2 \) effect of the same unit in the recurrent net. One could also posit a negative \( \alpha_2 \), an impossibility in the recurrent architecture described above where all units two distant have an excitatory effect regardless of the sign of alpha and beta. The choice between the two is primarily an empirical question that will be addressed in chapters three and four, but at some point the cost of additional connections in terms of processing, learnability and loss of constraint would outweigh the costs incurred by a recurrent net architecture. To date, the feedforward architecture appears to be a superior alternative for syllable network, though the choice is much less clear for the other two networks.\(^9\)

2.4.2.3 Feedback inhibition/excitation

As described thus far, the activity of the network is relatively homogenous across a target string. While distance and direction can influence the magnitude of an inhibitory/excitatory effect, all units are equally capable of influencing others and being influenced by them. The potential influence of a segment is directly proportional to its sonority without regard to its position in the string. High sonority segments (e.g. vowels) have significant inhibitory/excitatory potential while low sonority segments (e.g. obstruents) are relatively inert. But in any case the effect is linear. In an analysis of Spanish syllabification, Larson (1990) suggests, however, that some of the contextual effects in the network must be non-linear. It was argued that syllable peaks must exert a greater inhibitory effect on their neighbors than non-peaks.\(^{10}\) This can be accomplished by

\(^9\)The effects of perfect grid in the metrical network and unbounded harmony processes in the autosegmental network appear to provide strong evidence for a recurrent network whose single input activation can be propagated a theoretically unlimited distance (cf. 4.1.4, 5.2.2).

\(^{10}\)The justification for the non-linearity proposed in Larson (1990) now proves to be an artifact of the fact that that analysis used stipulated values for the sonority hierarchy (Clements' nine-class scale). Subsequent learning simulations with continuous sonority variables successfully model Spanish syllabification without recourse to peak-specific inhibition (see 6.4). It is an open question as to whether any syllabification phenomena require non-linearity.
positing feedback connections from the recognition layer. Since the recognition layer makes Boolean judgments concerning the location of syllable peaks and/or troughs (see below), connections from this layer can introduce non-linear influences into the network. Whether or not peak or trough-specific influences are germane to syllabification phenomena, feedback connections from the recognition layer can mediate important influences from higher prosodic levels (i.e. the metrical network). For instance, stressed syllables can perhaps differentially influence the derived sonority of segments within the syllable. More importantly, if derived sonority can influence not only syllabification but also the licensing of segments and/or features, feedback connections from the metrical network and from syllable peaks may prove to have crucial effects (cf. 5.2.4).

2.4.3 Recognition

Once the network computes derived sonority, it must be able to retranslate the quantized values into phonologically meaningful information. The most important function of the recognition layer within the syllable network is to provide an appropriate syllabification for the string. Whether one looks at a string's inherent sonority profile or its derived sonority profile, a pattern of rising and falling sonority is observed. Designing units that can identify maxima (peaks) and minima (troughs) in the sonority wave is a relatively straightforward process. Two kinds of computations can perform the task. Peaks and troughs can be identified by comparing the activation of a unit with that of its left and right neighbors (LOGICAL returns 1 if the formula is true, 0 otherwise).

$$\begin{align*}
(29) & \text{ Maxima (peak): } \text{LOGICAL}(d_i > d_{i+1} \text{ AND } d_i > d_{i-1}) \\
& \text{ Minima (trough): } \text{LOGICAL}(d_i < d_{i+1} \text{ AND } d_i < d_{i-1})
\end{align*}$$

While the above formulas involve relatively simple local computations, the introduction of Boolean operators could prove to be problematic in "neural-like" implementations (find references). Fortunately, the conjunctive inequalities above can be translated into linear threshold units that are less problematic. In general, a unit marks a transition in the sonority wave (i.e. peak or trough) if it generates a non-zero value in the following function:

$$\begin{align*}
(30) & \text{ } (|d_{i-1} - d_i| + |d_i - d_{i+1}| - |d_{i-1} - d_{i+1}| > 0)
\end{align*}$$

In a feedforward network, the task of identifying maxima and minima is even simpler. Assuming that we process the string from left to right (e.g. by using a sliding buffer, cf. Touretzky & Gupta, 1991), we can compare a unit to a single neighbor. If a unit has higher activation than its right neighbor, it is on the right side of the sonority wave; if it has lower activation, it is on the left side of the wave. Rather than having a tripartite division of the syllable (peak - H, trough - L, other - O), we would have a bipartite division (left/increasing - U, right/decreasing - D).
Translating from wave-speak to particle-speak or alternatively, from sonority-speak to constituency-speak is also a relatively straightforward process. In the HOL model (2.31), the peak (H) corresponds to the nucleus of the syllable, O's to the right of the H correspond to the coda (together comprising the rhyme), while O's to the left of the H comprise the onset. The only question concerns the appropriate constituency assignment for the non-initial L's. Traditionally, sonority troughs are assigned to the onset in accordance with the Maximal Onset Principle. The same assignment could well be provided in the current model though the wave metaphor lends at least some credence to the notion of ambisyllabicity (Kahn 1976, 1980). If the UD model (2.32) is chosen, troughs are automatically assigned to the onset while peaks are assigned to the rhyme. The crucial observation, however, is the fact that the recognition device allows both sonority and constituency analyses to be modeled within the network. While much of the current study will test how much work computational representations can perform, they are not fundamentally inconsistent with a theory that formally manipulates segments, strings, hierarchical constituents, etc.

As constructed, the recognition device provides a syllabification for any string presented to it. It is possible, however, that not all strings presented to the network are well-formed (cf. chapter three). Fortunately, the recognition device can be modified to provide either categorical or gradient judgments of well-formedness as well. In general, well-formedness failures result from putative peaks that are not licensed as syllable nuclei. For instance, the hypothetical English string /pnp/ conforms to the sonority sequencing principle and is parsed by the proposed recognition device as LHL. While the /n/ does qualify as a peak, it can not be licensed as a syllable nucleus in English. In terms of the syllable network, this condition can be represented by imposing a threshold on the peak recognition units. In order to be a syllable nucleus, a unit must not only be a peak, but also exceed some language-specific threshold. In theory, such a threshold could be imposed on other units in the network as well (e.g. codas). Since the network's recognition units allow us to translate numerical activation vectors into discreet constituents, if templatic conditions such as maximality (e.g. one coda slot) were demonstrated to be necessary, the network could also generate outputs that could implement these conditions. In general, the analyses of chapter three will attempt to
replace discreet constituency-based conditions with numerical sonority-sensitive conditions.

In addition to providing a unique syllabification of an input string and providing well-formedness judgments, the recognition layer can provide "quantitative" information such as syllable length as well. In addition to allowing us to derive moraic information by mapping units on to a constituent structure representation, the units can potentially provide direct information concerning quantity. For instance, if quantity were a function of the area enclosed by a syllable wave, syllable weight could be defined by a step function with that quantity as its independent variable.

2.5 Metrical Network

In addition to parsing syllables and computing derived sonority for segments, computational networks can be used to model the metrical structure of strings (cf. chapter four). As in the case of the syllable network, the metrical network is composed of three layers: input, processing and recognition.

**METRICAL NETWORK**

**Level-2: RECOGNITION**

*Maxima / Thresholds*

\[\square \ ]

\[\square \]

**Level-1: PROCESSING**

*Derived Stress*

\[\square \leftrightarrow \square \leftrightarrow \square\]

**Level-0: INPUT**

*Input Stresses*

\[\square \square \square\]

Figure 2.33

While the gross structure of the metrical network generally parallels that of the syllable network, the function of each of the units proves to be somewhat different. In the following sections, we will identify the ways in which the metrical network provides unique processing constraints, without reiterating the above discussion.
2.5.1 Input - Positional, Lexical, Moraic Activation

For the syllable network, segments define the vertical columns of units. As in most linguistic theories of metrical structure, the fundamental unit in the metrical network is the syllable (the mapping from the segment-based syllable network to the syllable-based metrical network will be discussed below). As a result, each vertical column in metrical network represents an individual syllable. Unlike the syllable network where most of the input comes from the inherent activation associated with each segment,\(^{11}\) input activation in the metrical network comes from a variety of sources. For quantity-insensitive languages, the most important variable is positional activation. Metrical analyses typically identify stress as being mapped from the left or right end of the string, with the possibility of an arbitrary number of syllables at the margin marked as extra-metrical. Rather than similarly suggesting that positional activation can be applied to a variety of units in the string, the network can achieve extra-metricality effects by allowing end-based positional activation to take both positive and negative values. For quantity-sensitive languages, we must also account for weight-based distinctions that generally assign higher input activations to heavy syllables in the string. An integrated network account must provide an interface with the syllable network that can communicate quantity distinctions to the input layer of the metrical network. Apart from this interface, the metrical network can be tested by simply marking heavy syllables.

Another unique characteristic of the metrical structure of a language is the fact that the lexicon can mark certain syllables as underlyingly stressed. The metrical network encodes this lexical stress as an exogenous input that can interact with positional and weight-based input activation. The metrical network is also different in that in any given string within a given language may possess several syllables which lack underlying positional (edge-sensitive) activation, which are not lexically stressed, and which are either light or not sensitive to weight distinctions. In other words, these syllables have no inherent activation. While it is possible that these syllables should all have an input activation of zero, it may be appropriate to assign these syllables some other input value or bias. The relative weight associated with each of the four variables can be determined experimentally (see chapter six).

2.5.2 Processing

In general, the processing alternatives for the metrical network are the same as those for the syllable network (see 2.4.2). Most of our analyses to date have used a recurrent network with lateral inhibition from the immediate left (β) and right (α) neighbors. Compared with a network with an arbitrary number of feedforward input connections (see 2.4.2.2), a recurrent net potentially offers a simpler analysis of perfect grid phenomena. Given a single pulse, the recurrent network can propagate the pulse an

\(^{11}\)Goldsmith & Larson (1992) identify syllabification phenomena in Polish and German that potentially require the introduction of positional activation. In addition to assigning an inherent sonority to each segment in the inventory, exogenous activation is assigned to morpheme-final positions.
indefinite length to create the appropriate pattern of primary and secondary stresses (chapter four).

![Graph showing stress patterns](image)

**Figure 2.34**

Prince (1992) also notes that the recurrent architecture makes valuable predictions concerning edge phenomena. On the other hand, a feedforward architecture for the metrical network can limit the distance an individual stress can be propagated. Prince (1992) argues that a weakness of the dynamic model is its ability to overgenerate stress systems. For instance, he argues that the network could model a language where the middle syllable in an arbitrarily long string could be selected as the single stress (if $\alpha = \beta$; $\alpha, \beta > 0$). In a feedforward network with a limited number of left and right connections, such a stress pattern could not be predicted (cf. 4.1.4; 6.5 for an response to Prince that does not require a revision of the network architecture).

### 2.5.3 Recognition

For the syllable recognition device, the principal task is to identify maxima in the derived sonority wave (with threshold units playing a lesser but sometimes crucial role). For the metrical recognition device the task is somewhat more complex. While successive segments can only rarely serve as nuclei of two separate syllables (cf. analysis of Spanish 3.2 and French 3.4.3), languages frequently permit some degree of stress on successive syllables (non-stress clash avoidant). As a result, identification of stress may require the determination of maxima and/or threshold units. In other words, a given language may stress all wave peaks, only those peaks which exceed a given threshold, or all units which exceed a given threshold (see 4.1.5). The metrical recognition device may also have to distinguish different degrees of stress (whether that distinction is analog or digital). Such a distinction has no direct corollary in the syllable network, unless a potential
discrimination of vowel length based on derived sonority is found to be derivable (cf. chapter five). Finally, the recognition device maintains feedback connections to the syllable network whereby stress can potentially influence syllabification and the licensing of segments and/or features.

2.6 Modular interfaces

Each of the three modules that have been discussed thus far can potentially be designed as self-standing and independently justifiable networks. A principal advantage of the proposed network architecture, however, is the possibility of combining the three into an integrated device for the representation and processing of phonological information. A fully integrated network would allow information to be passed top-down in feedback connections as well as bottom-up in feedforward connections. Just as the autosegmental network can pass sonority values to the syllable network to facilitate the appropriate syllabification of a string (3.1.2), the syllable network can potentially pass licensing information to the autosegmental network in order to facilitate appropriate phonetic realization (see ch. 5). Similarly, a metrical network that uses weight information supplied by the syllable network can, in turn, influence phonological processes that uniquely apply to stressed or unstressed syllables (e.g. vowel length, insertion, deletion, etc.; cf. 5.2).

Unfortunately, the co-registration of the various networks offers one of the most difficult computational challenges for the design of the model. As has been noted above, each of the network modules represents time as space by means of the vertical columns of units (cf. 2.7). The abstract unit of time represented by a column in the syllable network (i.e. segment) is, of course, significantly shorter than the abstract unit of time represented by a column in the metrical network (i.e. syllable). More importantly, there is no transparent process that permits one to predict how many segments are to be included in each syllable. While languages may indeed have a preference for CV syllables, most syllable templates are, in fact, much more complex. For instance, English syllables can arguably consist of one to seven or more segments (e.g. da skrajbd). The presence of variable-length syllables suggests that a fixed mapping architecture could not easily map from segments in the syllable network to syllables in the metrical network where the requirements of the latter require that syllables (or the segments marked as their head) be formally and computationally adjacent. The requirements of feedforward processing can be achieved by the buffering of input (particularly since the alpha/beta weights associated with each of the columns are identical). The maintenance of feedback connections such as those that can communicate additional activation from stressed syllables to their segmental heads requires somewhat more permanent links, however. Since the entire string may well have to be processed before stress is assigned, the links would have to be preserved throughout in order to have the stress information available to the segmental network. In the current serial simulation of a connection machine we simply use linked lists to maintain

\[\text{We ignore for the sake of discussion the debate concerning the status of onsets beginning with } /s/ \text{ and the putative appendix containing the } /d/ \text{ (cf. Selkirk 1982). Regardless of the status of either end of the string, it is clear that English permits syllables of variable length.}\]
the appropriate connections. More importantly, the analyses of the following three chapters will consider the three subnetworks independently so that the interface between the syllable and metrical networks will not come into play. It should also be noted that the current concerns are implementational rather than theoretical. Hinton (1990) discusses a variety of computational processes that can perform the necessary interface operations.

2.7 Representation of temporal structure - Reprise

The representation of temporal structure proves to be one of the most vexing problems for connectionist nets (Hinton 1990). The proposed network architecture demonstrates several advantages of the semi-local mode of representation where time is represented as is traditionally done within linguistics, namely, as a linearly ordered string. Each segment is assigned a column of units in the network. While the updating of units in the network can be proceed in any order (left-to-right, right-to-left, random, etc.) without any difference in performance, the network does preserve the temporal order of the string by the pattern of connections. Within the processing layer, each unit is only connected to its left (β) and right (α) neighbor. Between the input and processing layers, each processing unit is connected only to the input unit in same column and to an arbitrary number of its left and/or right neighbors (0-2). It is possible to generally reconstruct a string from an unordered set of paired inherent and derived sonority values. Since the derived sonority value is contextually dependent, a unique derived sonority value for each segment is determined by the environment in which it is found. It can also demonstrated the fact that even within a recurrent network, inherent sonority values can be reconstructed from derived values in a single step computation where the derived values are supplied as input to an identically configured network as the one used for production. In other words, within a single processing network, production (inherent → derived sonority) can be accomplished in a small number of iterations while parsing (derived → inherent sonority) can be accomplished in a single step using the same architecture.

The notion of time that is represented in the network as currently devised is also abstract rather than concrete. While the units are temporally ordered, they do not reflect constant units in real time. Each segment, whether consonant or vowel, takes up the same "space". While this has little or no impact on the syllabification and stress assignment tasks that are discussed in chapters three and four it may well come into play in the process of phonetic realization described in chapter five. One of the most significant difficulties confronting the proposed network architecture is the difficulty in representing phonological operations that ostensibly change the length of the string (insertion, deletion, reduplication). If these are viewed as dynamic processes that insert or delete segments after the string is mapped onto the network, the brittleness of a fixed mapping architecture would prove to be problematic (see extended discussion in chapter five). As was noted in (9) the current model uses timesharing rather than adopting a fixed mapping architecture. One of the critical differences between the proposed model and traditional connectionist architectures is the fact that each of these columns of units has the same pattern of connections and, more importantly, the same connection weights as every other column. Alpha and beta are system-wide parameters rather than being positionally dependent.
While postulating a network with multiple connections having the same weight might pose a unique learning problem (i.e. how could each similar connection learn to have precisely the same weight), it offers several advantages. First, the model has fewer variables, thereby reducing its solution space and offering a highly constrained architecture. Beyond the assignment of sonority coefficients to each segment, syllabification is accounted for with a very small number of additional variables (e.g. as few as two). Second, the homogeneity of the columns of units permits us to conveniently deal with strings of different lengths, a perennial problem for connectionist modeling. At least two different procedures are available. A permanent network could constructed with a number of columns greater than the longest string to be processed. Whether a given string is anchored to the left edge of the network or some other arbitrary (or random) position, the outcome would not be affected. Since the lateral inhibition in the network is constant rather than positionally dependent, the same syllabification would result regardless of length, beginning position, or ending position. Alternatively, a network could be dynamically created to accommodate the length of a presented string. In our simulations, we have adopted the latter strategy. In a spreadsheet implementation, it is easy to replicate a single column of units an arbitrary number of times (though it is more difficult to imagine how this would be accomplished in neural-like wetware). If we use a feedforward rather than recurrent net, the need to replicate columns of units becomes unnecessary altogether. With an appropriately sized buffer (e.g. three columns wide), a consistent alpha and beta make it possible to perform all necessary computations locally. With access to the inherent sonority value of the preceding and the following segment, any arbitrary segment in a string can compute its own derived sonority.
Chapter 3

Syllabification

In order to demonstrate the utility of the dynamic computational network described in chapter two, three phonological domains will be examined—syllabic, metrical and segmental. The syllable will receive the initial and greatest attention due to the recent proliferation of phonological analyses that rely on the syllable as the domain for phonological constraints/processes and, more importantly, because syllable-sensitive constraints often depend on information that is most successfully processed within the computational framework. Rather than simply testing the network against a variety of languages and syllabification problems (i.e. putting numbers into a computer and hoping that phonology falls out), however, we will examine well-established phonological analysis that utilize the kind of information that makes them amenable to a computational analysis. In the context of the evaluation guidelines introduced in chapter one, the goal will be primarily to demonstrate that a perspicuous analysis of syllabification phenomena requires the computational manipulation of sonority-sensitive energy waves and only secondarily that constituency analyses of syllabification may become superfluous or unnecessary.

3.1 The syllable as a phonological construct

While the syllable has long been recognized as a critically important unit in the organization of phonological information (cf. Whitney's 1865 description of Sanskrit grammar), it has suffered a rather checkered history in linguistic theory, being a construct that Haugen bemoaned for having "baffled the best linguistic minds" (cited in Levin 1985). While the syllable appeared in a variety of pre-generative analyses such as Firth 1948, Kurylowicz 1948, Pike & Pike 1947, Hockett 1955, and others, it was conspicuously absent from Chomsky and Halle's initial generative program in *The Sound Pattern of English* (1968). Kenstowicz and Kisseberth offer one potential rationale for its absence when they observe in 1979:

> It is generally accepted that at the phonetic level of representation the sounds of any utterance are organized into larger units called syllables. *However*, the syllable is *probably the most elusive of all phonological/phonetic notions*. . . . Until very recently the syllable has been largely ignored in generative phonology on the assumption that, though it may make sense to talk of sounds being organized into larger units phonetically, all phonological generalizations could be *satisfactorily stated* in terms of the individual sounds themselves *without invoking the notion of* the syllable *(emphasis mine).*

While Kahn (1976), Selkirk (1982), Halle & Vergnaud (1978), McCarthy (1979), Prince (1980) and others have demonstrated that phonological generalizations can indeed be
stated more satisfactorily by reference to the syllable as a phonological constituent, Itô (1986) argues that it still had not received due consideration.

While the syllable has enjoyed its status as a structural unit for some time, I believe it has not yet faced the full responsibilities which such a status entails.

Even when the syllable receives a central place in phonological description in studies such as Selkirk (1982), Levin (1985) and Itô (1986), it has a rather chameleon-like appearance due to the marriage of two seemingly incompatible theoretical constructs. Goldsmith (1990) contrasts two approaches to the representation of syllables, a syntactic phrase-structure model that he characterizes as the external representation of the syllable, and a sonority approach that he characterizes as an internal representation of the syllable. In the next two sections we will similarly describe two alternative conceptions of syllable structure: a constituency model in which syllable representations and the rules that modify them are described in terms of formal symbolic atoms governed by a combinatorial syntax, and a sonority model in which the energy relationships present within and between syllables are computed.

While the two approaches may appear antithetical to each other, modern phonological theories, in fact, usually incorporate elements of both approaches. Even though Selkirk, Levin and Itô propose constituent structure models for describing syllables, each of their accounts rely crucially on the sonority hierarchy to construct and appropriately order segments within a syllable. As was noted in chapter one, the current tension, along with a potential resolution, is, in fact, presaged by Pike's alternative conceptions of linguistic phenomena. As a result, we can perhaps label the two approaches, particle phonology (with apologies to Sandy Schane 1984) and wave phonology. Although there may be a strong temptation to embrace one and criticize the other, both may, in fact, serve as complementary metaphors that are each capable of explaining certain phonological phenomena. While the constructs are not strictly equivalent in either extensional or intensional terms, during the course of our discussion we will alternately refer to the sonority model as a computational model or a wave model of the syllable. In the same way, we will somewhat interchangeably refer to the constituency model as a syntactic or particle model.

3.1.1 Constituency Models of Syllable Structure

Although the majority of contemporary phonological theories agree on the importance of the syllable as a phonological construct, less consensus exists concerning its appropriate representation. Since the advent of generative phonology, there has been a great deal of discussion concerning the nature and complexity of the constituents dominated by the syllable node in a phonological tree (or graph).\footnote{Pre-generative discussions concerning the internal structure of syllables include Kurylowicz (1948), Pike and Pike (1947) and particularly Hockett (1955). For excellent survey of early approaches to syllable} In early string-based
approaches, the syllable constituent took a flat structure, consisting of a string of atomic segments delimited by syllable boundaries (dot boundaries) or left and right brackets (Kisseberth & Kenstowicz 1979, Fujimura & Lovins 1978). As the geometry of segmental representations became more complex (e.g. autosegmental phonology), the syllable node came to dominate a skeletal tier which is composed alternatively by X slots (Levin 1985), weight units or moras (Hyman 1986), or a string of Cs and Vs (McCarthy 1979; Clements & Keyser 1983, Itô 1986). Since several segmental terminals can appear in a given syllable, a variety of hierarchical constituents have been proposed to mediate between the syllable node and the segment (see Fudge 1968; 1987 for a survey). In most current analyses the syllable node dominates at least an onset and a rime (Selkirk 1982; Halle & Vergnaud 1980, Harris 1983). In many cases, the rime is further divided into a nucleus and a coda (Cairns & Feinstein 1982, Goldsmith 1990). In addition to the differences between proposed constituent structures, the assignment of individual segments to a constituent varies between different approaches. For instance, pre- and post-nuclear glides have an ambiguous status. While most analysis include on-glides as part of the onset, Harris' (1983) analysis of Spanish includes them in the rime. In models that contain both nuclear and coda slots, off-glides are alternatively placed in either one or the other (Selkirk 1982; liquids are also occasionally included in the nucleus). Within Kaye, Lowenstamm & Vergnaud's model, a distinction is made between long vowels where both moras are found in the nucleus (contra Goldsmith 1990 where a nucleus can contain only one slot) and off-glides which are assigned to the coda (Charette 1988).

While there is little consensus as to how the syllable is to be divided into constituents and what information those constituents encode, what all of the proposed structures share in common is the fact that phonological information is represented by grammatical constituents that are organized into a hierarchical structure and governed by categorial properties. In this regard, phonological representations have been constructed on analogy to, or perhaps even as an extension of syntactic representations (see Levin 1985 for an explicit extension of X-bar syntax into phonology).

The constituency-sensitive categorial approach to syllable structure representations has been successful in accounting for a wide range of phonological phenomena. Templatic approaches such as Prosodic Phonology (Itô 1986) employ syllable constituency to account for most phonotactic constraints. For instance, maximality controls how many segments can be assigned to a given constituent (Itô 1986) or the number of branches that a constituent can dominate (Harris 1983). Prosodic licensing determines which segments can appear (cf. autosegmental licensing, Goldsmith 1989; Bosch 1988). Membership in or assignment to a particular constituent has been adduced as an explanation for a variety of constituency-sensitive rules such as aspiration and intervocalic flapping in English (Selkirk 1982), vowel length in Icelandic (Vennemann 1972, 1974), as well as lenition and fortition rules (Foley 1977).

organization, see Levin (1985). For more a discussion of more recent approaches, see Bosch (1988) or Itô (1986).
3.1.2 Sonority as a Phonological Variable

At the same time that constituency-dependent rules and representations have been developed, an alternative conception of phonological information has been gradually integrated into phonological theories with much less fanfare and unfortunately, less theoretical rigor as well. Beginning with Sievers (1893) and Jespersen (1904) it has been recognized that syllables generally conform to a distinctive bell-shaped distribution whereby vowels form the core (nucleus) of the syllable while segments tail off to either margin consistent with the following hierarchy (Jespersen 1904):

Sonority Hierarchy

- Low vowels
- High vowels
- Glides
- Liquids
- Nasals
- Voiced continuants
- Voiceless continuants
- Voiced stops
- Voiceless stops

Figure 3.1

This sonority hierarchy, while falling out of vogue during the descriptivist and structuralist periods of the mid-twentieth century, has been reintroduced into the phonological discussion by early critics of standard generative phonology (Hooper 1972; Venneman 1974; Foley 1977). In the subsequent twenty-year period, it has come to play at least a descriptive role in nearly every discussion concerning the syllable (cf. Selkirk 1978, 1982; Harris 1983; Dell & Elmedlaoui 1985; Levin 1985; Itô 1986; Clements 1987). Constraints such as the Sonority Sequencing Principle (Selkirk 1982), formalize the role of sonority in syllable construction by requiring that a syllabification algorithm build syllables from segments that continuously increase in sonority on the left side of the nucleus and continuously decrease on the right side of nucleus. In recent theories, the Sonority Sequencing Principle has been augmented by several additional principles and stipulations in order to achieve greater descriptive adequacy. Notable among these are the
Minimum Sonority Distance Principle (Harris 1983), the Maximal Onset Principle, the Syllable Contact Law (Murray and Venneman 1983), the Inertial Development Principle (Foley 1977), the Feature Contrast Principle (Clements 1987) and the Core Syllabification Principle (Clements 1987).

While sonority has been demonstrated to be a useful phonological construct, it has been notoriously difficult to give an adequate characterization of what sonority encodes or to formalize the constraints that make reference to it. While sonority appears to represent a scalar value associated with each segment, the sonority hierarchy is typically represented formally as an ordered list of natural classes. Whether syllable-sensitive phonotactics are due to sonority or constituency, a phonological grammar must account for the parsing and generation of syllables. While several explicit syllabification algorithms have been proposed (e.g. Selkirk 1982; Harris 1983; Walli-Sagey 1985; Dell & Elmedlaoui 1985; Itô 1986; Clements 1987), it is generally very difficult to construct derivational algorithms within the framework of generative phonology. Due to the lack of adequate computational machinery to directly process sonority relationships within and between syllables (cf. wave metaphor discussion in 1.1), the syllabification algorithms are limited to references to adjacency or non-adjacency on the sonority hierarchy which is characterized as a ordered list or to templates that describe the various positions in the syllable in terms of sets of features which only incidentally make the positions correspond to the Sonority Sequencing Principle.

By allowing the direct manipulation of sonority as a phonological variable, the dynamic computational network can efficiently generate (or diagnose) the periodic waves that comprise Pike's wave description of the syllable. In addition to providing a straightforward parser, the dynamic computational network is uniquely warranted in situations where sonority-sensitive constraints are invoked. Constraints such as the Sonority Sequencing Principle and the Minimum Sonority Distance Principle require that sonority be treated as a scalar property of a segment. They also require the manipulation of this scalar variable in several context-dependent computations. It is necessary to be able to compare the sonority values of adjacent segments, to be able to scan for increasing and decreasing sonority slopes, and to compute local maxima and/or minima. It may also be necessary to test values against thresholds and to calculate whether the absolute difference between adjacent units exceeds some established coefficient. While these computations are notoriously to perform difficult within generative rule formalisms, they can be calculated efficiently (locally and in parallel) in the computational network proposed in chapter two. As a result, syllabification analyses which take advantage of sonority-sensitive constraints are much more perspicuous than constituency-based alternatives.

Three such analyses will be considered in the next section: Spanish (Harris 1983), Icelandic (Itô 1986), and Imldawn Tashhiyt Berber (Dell & Elmedlaoui 1985). The final section of the chapter will address Clements' (1987) concerns relating to the descriptive adequacy of sonority-based syllabification theories by considering unique syllabification problems that appear in English, Polish, Russian, French and Spanish.
3.2 Spanish

Spanish is a language that, according to Harris (1983; 1989), requires the use of both sonority-sensitive and constituency-sensitive constraints in order to account for the distribution of segments in well-formed syllables. While Spanish syllables must conform to the Sonority Sequencing Principle and the Minimum Sonority Distance Principle (MSD > 1), Harris also uses such constituency-sensitive constraints as positing separate onset and rime construction rules and specifying a maximal binary branching constraint on constituents. To the extent that Harris' account necessarily makes reference to sonority-sensitive constraints while retaining constituency-sensitive conditions it satisfies the second level of justification for a computational model (cf. 1.2). On closer inspection, however, Harris' constituency constraints prove to be unnecessary, thereby providing an even stronger (level one) justification.

3.2.1 Spanish Onsets - Harris' Account

Of the set of possible onsets generated by concatenating segments from the five major classes O(bstruent), N(asal), L(iquid), G(lide) and V(owel) only a very small subset is actually possible in Spanish. For example, of the 25 possible two-segment sequences, only OL is well-formed. Harris (1983) accounts for the distribution by positing the following rule:

(2) Onset Rule (Final Version): Construct a maximally binary branching tree of category O(nset) whose branches dominate [+consonantal] segments that are not adjacent on the universal sonority scale (p. 21).

This rule assumes the following sonority scale (with no discrimination within major classes):

(3) OBSTRUENT NASAL LIQUID GLIDE VOWEL

Four different constraints are subsumed in the statement of this rule:

(4a) Licensing: The Onset constituent can only dominate [+consonantal] segments.

(4b) Maximalilty: The Onset template consists of a tree with a single binary branching.

---

2A version of the following discussion appeared in CLS 26:2, The Parasession on the Syllable in Phonetics & Phonology.
(4c) Sonority Sequencing Principle: The order of the segments in the Onset constituent must increase in sonority from left to right (cf. Selkirk 1983; Clements 1987).

(4d) Minimum Sonority Distance Principle: Segments can not be adjacent on the universal sonority scale.

While four constraints are encoded in the rule, its "work" is principally done by the two sonority-based constraints, however. Once vowels and glides are excluded from onsets by the requirement that segments be [+cons], only three classes of segments remain eligible for inclusion. Harris' rule appears to limit onsets to two segments by the Maximality Condition, but the same effect is achieved redundantly by requiring non-adjacency on the sonority scale (ONL would violate non-adjacency twice). Though it is not stated explicitly in the rule (since it is assumed to be part of universal grammar), the Sonority Sequencing Principle is adduced to filter the following sequences: *LL, *LN, *LO, *NN, *NO, *OO (the exclusion of *LL, *NN and *OO could have been attributed to non-adjacency as well). Finally, the Minimum Sonority Distance Principle filters *ON and *NL.

The reliance on sonority-sensitive constraints in Harris' analysis provides a warrant for a computational network that instanciates these constraints, whether or not a Spanish grammar also requires the two constituency-sensitive constraints. Because constituents can be mapped from the recognition layer of the syllable network, constituency-sensitive constraints do not necessarily invalidate computational models. A much more compelling argument for a dynamic computational network can be made, however, if it can successfully filter the ill-formed clusters without recourse to the syntactic conditions. Before testing the network on the complete set of logically possible onset clusters, it is possible to reanalyze Harris' set of constraints to see if they are actually necessary.

If we consider all possible strings from the set {ONLGV} that consist of up to three segments, we can compute the functional load of each of the four constraints. The following table lists the number of ill-formed clusters that are correctly filtered by each constraint. Since several ill-formed clusters are excluded by more than one of the constraints, the parenthesized value indicates the number of clusters which are uniquely filtered by a given constraint.

---

3The reference to a universal sonority hierarchy may well be problematic (cf. Clements 1987). Syllable structure constraints in several languages appear to presuppose different orderings of segments (/r/ vs /l/), different categorizations of natural classes and relationship between voiced stops and voiceless continuants (cf. Levin 1985). Additionally, the universal Sonority Sequencing Principle has both a strong and a weak version. In the former, Sievers (1893), onset segments must exceed the sonority of their left neighbor. In the latter, Jespersen (1904), onset segments must not have lower sonority than their left neighbor. The constraint against *OO, *NN and *LL in Spanish represents the strong version which is not universal.
(5) | Permutations: 5 | 25 | 125 |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceptable Onsets: 3</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Licensing: 2 (2)</td>
<td>16 (6)</td>
<td>98 (0)</td>
<td></td>
</tr>
<tr>
<td>Maximaliy: 0 (0)</td>
<td>0 (0)</td>
<td>125 (0)</td>
<td></td>
</tr>
<tr>
<td>Sonority Sequencing: 0 (0)</td>
<td>15 (1)</td>
<td>115 (0)</td>
<td></td>
</tr>
<tr>
<td>Minimum Distance: 0 (0)</td>
<td>12 (2)</td>
<td>95 (0)</td>
<td></td>
</tr>
</tbody>
</table>

Given that the Minimum Sonority Distance Principle prohibits adjacency on the sonority scale and the Licensing Constraint excludes glides and vowels, the maximality constraint does not uniquely filter out any of the ill-formed strings and, as a result, by Occam's razor it is not necessary. While each of the other three remaining constraints do appear to be independently necessary, a closer examination allows even more simplification of the onset rule.

Within the context of Harris' analysis, each of the constraints subsumed under the onset rule operates within rather than between constituents (onset and rime). The Sonority Sequencing Principle (SSP), which is assumed to be a constraint of universal grammar, has a broader domain of application, however. It crucially makes reference to the peak of a syllable, requiring that segments to the left of the peak continuously increase in sonority and elements to the left decrease in sonority (Clements 1987). If Harris' application of the Sonority Sequencing Principle were appropriately extended to include all segments up to and including the peak of the syllable, the Licensing Condition that onsets contain only [+consonantal] segments would become unnecessary. The strings uniquely filtered by the licensing requirement in Harris' account would all violate the strong version of the Sonority Sequencing Principle that is independently required for Spanish. For Harris, on-glides are considered part of the rime so that a rime can begin either with a glide (G) or a vowel (V). Any hypothesized onset that included a vowel would create either a VGV or a VV sequence, where the final V would be the peak of the syllable. Either configuration would violate the strong version of the SSP. So, if the syllable peak is included in the domain of the SSP, vowels would be excluded from onsets even without a distinct licensing restriction. The situation for glides is slightly more complex. G+GV sequences violate the SSP in the same way that V+V violates it. If a glide precedes a vowel-initial rime (G+V), the SSP predicts that the configuration would be well-formed, but the resulting GV sequence would simply be (re-)parsed as a well-formed rime (by Harris' analysis). This would be precisely the correct result. Again the licensing condition that excludes vowels and glides from the onset constituent would prove redundant. In tabular form, the facts would be as follows:

(6) | ONSET | ONSET + initial segments of RIME |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td>*V+V (SSP strong version)</td>
</tr>
<tr>
<td>*V+GV</td>
<td>(SSP)</td>
</tr>
<tr>
<td>G</td>
<td>G+V (parsed as a GV rime)</td>
</tr>
<tr>
<td>*G+GV</td>
<td>(SSP strong version)</td>
</tr>
</tbody>
</table>
At this point, we have demonstrated that neither of the constituency-sensitive constraints (licensing, maximality) are actually necessary for the correct characterization of the Spanish onset. Fortunately, even more simplification in the onset rule can be made with an appropriate reformulation of the Minimum Sonority Distance Principle. For Harris, the sonority hierarchy is represented as an ordered list rather than as a scalar variable. As a result, Harris' onset rule simply refers to adjacency. If the sonority hierarchy is viewed in computational terms, the non-adjacency requirement is actually a special case of a more general constraint that requires that the sonority of a segment in an onset cluster must exceed the sonority of its left neighbor by more than a specified amount $\delta_O$. A parallel constraint can operate in the coda requiring that the sonority of a segment in a coda cluster must exceed the sonority of its right neighbor by more than a specified amount $\delta_C$. For the Spanish onset a Minimum Sonority Distance Principle with $\delta_O=1$ encodes the non-adjacency constraint proposed by Harris.

The generalized version of the Minimum Sonority Distance Principle (MSD) offers several advantages. First, it allows a parametric variation between languages relating to the degree of difference required between segments in an onset cluster (Levin 1985). Second, Harris' non-adjacency analysis for Spanish onsets requires a carefully defined sonority scale (ONLGV) in order to produce the correct generalizations. While Harris claims this scale is universal, some languages require additional discrimination by utilizing oppositions such as $+/-$ voice, $+/-$ continuant or $/l/$ vs. $/ll/$ (Levin 1985; Clements 1987). If the sonority hierarchy is quantized, it is also possible to posit a difference between successive categories that is greater than or less than one. For instance, it is not absolutely necessary to consider the difference between O and N to be the same as the difference between L and G. Finally and most critically, if the MSD is allowed to encode direction (i.e. positive or negative numbers) as well as magnitude, the Sonority Sequencing Principle proves to be just a special case of the Minimum Sonority Distance Principle where $\delta_O=0$ and $\delta_C=0$. In all cases where the Sonority Sequencing Principle is violated, the appropriate Minimum Sonority Distance Principle would also be violated. As a result, all of the constraints encoded by Harris' onset rule can be accounted for by one constraint, the Minimum Sonority Distance Principle.

Before examining how these sonority-sensitive constraints can be captured by a computational network, one final issue needs to be addressed. Of the five major classes of segments discussed by Harris, glides have a rather anomalous status. In general, the glides /y/ and /w/ in Spanish are simply [-syllabic] variants of underlying /i/ and /u/.

---

4Harris (1989) recognizes the possibility of language variation in the degree of discrimination utilized in the sonority hierarchy. He retains the adjacency of the five major classes by using nested distinctions.

5The strong version of the Sonority Sequencing Principle is encoded in the generalized Minimum Sonority Distance Principle by requiring that the difference by "greater than" $\delta$. The weak version requires that the difference be "greater than or equal to" $\delta$.

6There may be some motivation for retaining underlying glides in words such as yugo (realized as zugo in some dialects). That issue does not bear on the current analysis however.
the feature [+syll], a network account improves on Harris’ rime rule and eliminates much of the motivation for GLIDE as an independent class.

Glides also present a constituency problem. For a variety of descriptive as well as theoretical reasons, Harris includes both pre-nuclear and post-nuclear glides as part of the rime constituent. By including pre-nuclear glides in the rime Harris is able to identify an appropriate maximality constraint on rimes (3 segments maximum). The computational network proposed in this paper would appear to make such a maximality condition as redundant in the rime as it is in the onset (see 3.2.3 below). Second, Harris notes that pre-nuclear glides make syllables heavy with respect to quantity-sensitive stress rules in the same fashion as post-nuclear elements (codas). Although such a discussion is outside of the scope of the current topic, the proposed network will attempt to account for such moraic phenomena without explicit reference to constituency (cf. chapter five). As noted earlier, the Sonority Sequencing Principle divides the syllable into pre-nuclear and post-nuclear sonority slopes, thereby creating an incongruity between the two representations. While constituent structure labels such as ONSET, NUCLEUS, RIME and CODA bear little theoretical significance within the proposed computational network, such labels will be retained for referential clarity. The NUCLEUS of the syllable will be identified with the single segment that appears at the peak of the sonority wave. The label ONSET will refer to all segments to the left of the peak while the label CODA will be used to refer to segments to the right of the peak. The segment appearing at the trough between two syllable peaks will be classified as part of the onset (consistent with the Maximal Onset Principle; cf. Bosch 1988).

By considering glides to be reflexes of non-nuclear vowels and by considering pre-nuclear glides to be part of the onset rather than the rime, the statement of constraints governing the onset becomes maximally simple. The Licensing Condition that limits onsets to [+consonantal] segments was earlier shown to be redundant within Harris’ framework. In fact, all classes of segments can appear in onsets (ONLGV) with onset vowels being realized as glides. The only constraint on onset segments is the Minimum Sonority Distance Principle where Spanish requires that a segment within an onset cluster exceed the sonority of its left neighbor by more than one degree. The following therefore comprise the set of well-formed onsets in Spanish:

(7) \[
\begin{array}{ll}
\text{ONSET} & \text{ONSET + NUCLEUS} \\
O & O+V \rightarrow OV \quad (padre) \\
N & N+V \rightarrow NV \quad (madre) \\
L & L+V \rightarrow LV \quad (libre) \\
G (?) & G+V \rightarrow GV \quad (yugo) \\
V & V+V \rightarrow GV \quad (lago) \\
\text{OL} & \text{OL}+V \rightarrow \text{OLV} \quad (plata) \\
\text{OV} & \text{OV}+V \rightarrow \text{OVG} \quad (tiempo) \\
\text{NV} & \text{NV}+V \rightarrow \text{NGV} \quad (miedo) \\
\text{LV} & \text{LV}+V \rightarrow \text{LGV} \quad (lienzo) \\
\text{OLV} & \text{OLV}+V \rightarrow \text{OLGV} \quad (cliente) \\
\end{array}
\]
3.2.2 Spanish Onsets - Network Account

Once the Spanish onset rule is reduced to the application of the Minimum Sonority Distance Principle and the Sonority Sequencing Principle, the computational network described in chapter two becomes an ideal device for encoding its constraints. Lateral inhibition models the effects of the Minimum Sonority Distance Principle directly. For example, in a string consisting of a two-segment onset cluster and a nucleus, the high sonority nucleus will reduce the sonority of its lower sonority left-hand neighbor by a greater degree than that segment will inhibit its neighbor. As was noted in chapter two, the first "5" in the sequence "15951" will be inhibited by the following "9" to such an extent that when $\alpha = -.776$, the "5" will have a lower derived value than the preceding "1". In general, as the sonority of the peak increases and/or the magnitude of $\alpha$ increases, the difference between the segments in an onset cluster must also increase in order for the sequence to have the continuously rising sonority profile required by the Sonority Sequencing Principle. In applying the current model to the Spanish data noted by Harris, for each of the 155 possible strings of 1-3 segments, a value of -.40 for $\alpha$, 6.00 for the threshold and 0.00 for the other coefficients will generate the correct well-formedness predictions. Several of the critical cases deserve discussion. Harris identifies *ON and *NL as onset clusters which uniquely require the Minimum Sonority Distance Principle for exclusion. Both are appropriately marked as ill-formed by the syllabicity network.

The presence of a sub-threshold derived sonority peak, marked in the above graphs by an asterisk and the lack of a peak on the syllable tier, diagnoses the string as ill-formed. For the remainder of this discussion, the results of the network will be presented in a textual format where the numbers represent the derived sonority of each segment in the string. Sonority peaks will be bold-faced with sub-threshold peaks (false peaks) underlined and marked with an asterisk.

\[(10) \ *ml\eta \ -1.44 \ 3.20 \ 1.78 \ 9.37 \ -3.05 \ 9.00 \ 0.00 \ # \ *

\text{Figure 3.8} \quad \text{Figure 3.9} \]
While *ON and *NL strings are ill-formed, OL strings prove to be acceptable.

(11) \textit{plata} \begin{tabular}{cccccccc}
0.08 & 0.21 & 1.98 & 10.04 & -2.60 & 9.00 & 0.00 & \\
\# & \(p\) & \(l\) & \(a\) & \(t\) & \(a\) & \# \\
\end{tabular}

The following examples demonstrate that appropriate CV+V strings are deemed to be well-formed. The network automatically selects the appropriate syllable nucleus and accounts for glide formation as well.

(12) \textit{tiempo} \begin{tabular}{cccccccccccccccc}
0.33 & -0.83 & 4.58 & 6.05 & 4.88 & -2.20 & 8.00 & 0.00 & \\
OV+V & \# & \(t\) & \(i/y\) & \(e\) & \(m\) & \(p\) & \(o\) & \# \\
\end{tabular}

\textit{lienzo} \begin{tabular}{cccccccccccccccc}
-1.68 & 4.22 & 4.44 & 6.39 & 4.05 & -0.18 & 7.99 & 0.00 & \\
LV+V & \# & \(l\) & \(i/y\) & \(e\) & \(n\) & \(s\) & \(o\) & \# \\
\end{tabular}

The underlying \(i/y\) in the above words is forced out of the nucleus as a result of the inhibition produced by the following \(l/e\). Not only is its status as a glide predicted by its non-nuclear position but also its relatively low derived sonority. Finally, the network also correctly predicts that the three segment onset OLV+V in words like \textit{cliente} should be well-formed:

(13) \textit{cliente} \begin{tabular}{cccccccccccccccc}
0.07 & -0.19 & 2.69 & 5.14 & 6.27 & 4.98 & -2.59 & 8.36 & 0.00 & \\
OLV+V & \# & \(k\) & \(l\) & \(i/y\) & \(e\) & \(n\) & \(t\) & \(e\) & \# \\
\end{tabular}

Before leaving the subject of Spanish onsets, it should be noted that not all onset clusters that are predicted to be acceptable by reference to major classes are in fact well-formed. In addition to the onset rule, Harris posits language particular filters for Spanish that explain the following:

(14) \textit{tl} (* some dialects) \textit{fl} *\textit{s}l *\textit{c}l *\textit{z}l *\textit{x}l

*\textit{dl} \textit{fr} *\textit{s}r *\textit{c}r *\textit{z}r *\textit{x}r

The need for language-specific filters results in part from the fact that the Minimum Sonority Distance Principle, as formulated by Harris, does not permit discrimination between the various obstruents with respect to sonority values. The proposed network offers significantly more flexibility in handling these clusters, however. It was noted above that many languages require finer discrimination in sonority values than would be available if we were limited to the major classes ONLG. The values used in this simulation have in fact been those of the fully articulated sonority hierarchy (the network performs equally well with a restricted sonority scale such as Harris').
Within the class of obstruents, the sonority values underlying this simulation are as follows:

(15)  Voiceless stops  1.00  
      Voiced stops  2.00  
      Voiceless continuants  2.00  
      Voiced continuants  3.00  
      Affricates  3.00  
      /s/  3.00

As a result, *sl, *sr, *zl, *zr, *zl, *cz are all predicted to be ill-formed without any special stipulations:

(16)  *sleta  -0.99  2.20  1.78  9.37  -3.05  9.00  0.00
      #  *s  l  e  t  a  #

The sonority values which underlie the network also generate the prediction that xr and x/ should be considered well-formed though accidentally missing, a judgment shared by Harris (Jrusche).

The fact that the network can distinguish between various types of obstruents does not solve the problem of t/l and *dl however. The values stipulated for these segments would suggest that both should form acceptable onset clusters. While a full discussion lies beyond the scope of this paper, the most promising solution relies on the observation that ill-formed clusters share significant features (e.g. voicing, point of articulation). If the right-hand member of the pair that shares an autosegment bears the "cost" in terms of sonority, the left-hand member would have significantly higher sonority than it would have in isolation. Such a move would account for several assimilation phenomena, the behavior of geminates, homorganic nasals and "coda licensing" in Prince-languages (see 5.2.1, fn. 15).

3.2.3 Spanish Rimes - Harris' Account

The description of the segmental distribution in Spanish rimes is significantly more complex than that required for onsets. Harris offers three modes of description: a tabular representation of the data, a list of observations concerning the table and finally, three rules that account for the distribution.

The choice of sonority values in the simulation is for the sake of argument (consistent with the sonority hierarchy presupposed in Harris 1983). The learning algorithm discussed in chapter six permits the discovery of more appropriate coefficients that simplify the analysis. The point of the argument here is that even with traditional sonority values, the introduction of lateral inhibition permits the correct characterization of the syllable structure constraints in Spanish.
(17) SPANISH RIMES - Harris 1983

<table>
<thead>
<tr>
<th>MEDIAL</th>
<th>FINAL</th>
<th>MEDIAL</th>
<th>FINAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. 1.</td>
<td>V pa-ta tapa</td>
<td>b. 1.</td>
<td>Vs pas-ta res</td>
</tr>
<tr>
<td>2. VG</td>
<td>au-tor lei</td>
<td>2. VGs</td>
<td>claus-tro seis</td>
</tr>
<tr>
<td>3. VL</td>
<td>sal-ta mar</td>
<td>3. VLS</td>
<td>pers-piczaz vals</td>
</tr>
<tr>
<td>4. VN</td>
<td>com-pra sarten</td>
<td>4. VNs</td>
<td>mons-truo Mayans</td>
</tr>
<tr>
<td>5. VO</td>
<td>seg-mento red</td>
<td>5. VOs</td>
<td>abs-tracto Felix</td>
</tr>
<tr>
<td>6. VGL</td>
<td>* *</td>
<td>6. VGLs</td>
<td>* *</td>
</tr>
<tr>
<td>7. VGN</td>
<td>(*) *</td>
<td>7. VGNs</td>
<td>* *</td>
</tr>
<tr>
<td>8. VGO</td>
<td>(*) *</td>
<td>8. VGOs</td>
<td>* *</td>
</tr>
</tbody>
</table>

c. 1. GV | nue-vo apio |
| 2. GVG | (*) buey |
| 3. GVL | fuer-tem fiel |
| 4. GVN | siem-pre Juan |
| 5. GVO | diag-nosis Goliat |
| 6. GVGL | * * |
| 7. GVGN | * * |
| 8. GVGO | * * |

RIME RULE R1: Construct a maximally binary branching tree of category R(ime) whose obligatory left branch dominates [+syll,-cons] and whose optional right branch dominates [-syll].

RIME RULE R2: Adjoin a [-cons] segment to a rime.

RIME RULE R3: Adjoin the segment /s/ to the right of an existing rime.

Harris’ rules refer either explicitly or implicitly to a variety of constraints: a Maximal Condition that limits "core" rimes to two segments (binary branching), a Licensing Condition that restricts which segments can appear in certain positions (+/-syllabic, -consonantal), obligatory and optional elements, adjunction (adjoin either a [-cons] to the left or /s/ to the right), and the Sonority Sequencing Principle.

As in the case of the onset, most of the "work" of the rules can be accomplished independently through sonority sensitive constraints. For instance, Harris observes that in the table of potential rimes, most rimes with three segments are well-formed while all four-member rimes are ill-formed. While it is true that 9 of the 12 three-segment rimes that are listed are well-formed, only a small percentage of all possible three-segment rimes are listed (including surprising omissions like *VLN, *VLO, and *VNO). Excluding /s/ for a moment, 125 three-segment strings can be generated. Of these only four are well-formed (GVC, GVL, GVN, GVO). The acceptability of glide-initial rimes has already been accounted for in the proposed model by including pre-nuclear glides in the onset (V+V →
GV). No other three-segment permutation of ONLGV proves to be well-formed, with nearly all eliminated by the Sonority Sequencing Principle and the Minimum Sonority Distance Principle.

The real issue concerning the rime has to do with the status of /sl/, a problem that plagues analyses of most European languages. An analysis must account for the fact that /sl/ can be adjoined to well-formed two-segment rimes whether or not its inclusion violates other constraints such as the Sonority Sequencing Principle (e.g. abstracto). It must simultaneously rule out adjunction to rimes that are preceded by a pre-nuclear glide. Indeed, this was one of the principal motivations for Harris including the pre-nuclear glide in the rime since it generalizes the three-segment limitation on rimes. Even with this maximality condition, Harris is forced to posit a rule which he admits is highly marked. The analysis of fiesta as GVVs rather than GVO also raises questions concerning the explanatory adequacy of the account. While GVVs sequences do not violate the three-segment maximum, a derivation where V is generated by R1, G is generated by R2 and /sl/ is generated by R3 should be impossible. If fact, the principle of maximal rule application (Harris 1983, p. 26) suggests that R1 would interpret GVVs as GVO and Vs as VO, thereby correctly predicting their acceptability but not considering them an instantiation of R3 despite the fact the fiesta seems intuitively to pattern with monstruo with respect to the nature of the /sl/.

3.2.4 Spanish Rimes - Network Account

A computational network is generally capable of correctly predicting the acceptability of possible rimes with significantly less theoretical machinery than required in Harris' account. To adequately account for the phonotactic restrictions on the Spanish rime, left-to-right inhibition must be introduced into the network (a Minimum Sonority Distance Principle operates on both sides of the nucleus, cf. Levin 1985; Clements 1987).8

Temporarily setting aside the question of /sl/-adjunction rimes, the network succeeds in predicting the acceptability of V, VG, VL, VN and VO rimes. Most illuminating is the network's processing of X+VG strings (particularly G+VG).

\[
\begin{array}{c|cccccc}
\text{(18) autor} & -3.43 & 8.58 & 3.79 & -2.07 & 7.68 & 2.93 & 0.00 \\
VG & \# & a & u/w & l & o & r & \# \\
\hline
\text{buey} & -0.17 & 1.56 & 3.98 & 7.56 & 3.98 & 0.00 \\
G+VG & \# & b & u/w & e & i/y & \#
\end{array}
\]

The network correctly identifies syllable peaks and generates both on-glides and off-glides as a result of the effects of lateral inhibition and the fact that high vowels have lower

8In the initial simulations with fixed integer sonority values, the correct parsing of Spanish rimes required peak-specific feedback inhibition (\(\alpha_2/\beta_2\)): \(\alpha = -.11 \quad \alpha_2 = -.29 \quad \beta = .00 \quad \beta_2 = -.40\). When segments are permitted to take non-integer values, the entire data set can be accommodated without the feedback loop (cf. 6.4; Appendix II.10).
underlying sonority than other vowels. The syllable peak usually corresponds to the vowel that has the higher underlying sonority unless the sonority values are equal. If the two underlying vowels are identical (ii, uu) the output is ill-formed (not yi, iy, wu or uw; Spanish does not license long vowels). The dominance of right-to-left inhibition predicts that iu → yu and ui → wi should be more likely than iu → iw and ui → uy (all things being equal - stress on the first vowel would produce the opposite pattern; cf. Clements and Keyser 1983).

The network performs correctly in generally excluding *VGL, *?VGN, *?VGO as well as *VLO, *VNO and *VLN. In certain configurations, VGO is predicted to be well-formed, however. Harris notes that auxilio is acceptable while other VGO strings are not. The network makes a similar prediction, accepting auxilio and considering the general class very marginal.

\begin{align*}
(19) \text{auxilio} & \quad -3.45 \quad 8.62 \quad 3.45 \quad 0.96 \quad 0.33 \quad 6.68 \quad 2.93 \quad 3.65 \quad 8.37 \quad -3.35 \\
& \quad \# \quad a \quad u/w \quad k \quad s \quad i \quad l \quad i/y \quad o \quad \#
\end{align*}

\begin{align*}
(?) \text{auktilio} & \quad -3.46 \quad 8.65 \quad 3.18 \quad 3.19 \quad -1.67 \quad 6.68 \quad 2.93 \quad 3.65 \quad 8.37 \quad -3.35 \\
& \quad \# \quad a \quad u/w \quad k \quad t \quad i \quad l \quad i/y \quad o \quad \#
\end{align*}

Under current assumptions, the network correctly filters *GVGs, *GVLS, *GVNs, *GVOs, *GVGLs, *GVGNs and *GVGOs without imposing a separate maximality constraint. The network also correctly accepts GVVs, VVs, and VGs. The network has difficulties, however, with VLs, VN, and VO, marking them as ill-formed despite their acceptability. Since the assignment of sonority values patterns /s/ with other continuants (sonority = 3.00), the lateral inhibition that excludes VLO also excludes VLs. The worst case scenario is VOs+O (abstracto) where /s/ has higher underlying sonority than either of its neighbors. No simple twiddling with inhibition coefficients could salvage VOs+O as well-formed. While a variety of theoretical moves could be implemented to account for the special status of /s/, none flow uniquely from the computational potential of the network. At this point the status of /s/ must be left as a question for future investigation.

3.3 Icelandic⁹

In the preceding discussion of Spanish syllabification it was demonstrated that basically all of the syllable structure constraints in Spanish can be reduced to the application of the Minimum Sonority Distance Principle and that the necessary sonority-sensitive computations can be processed by means of a dynamic computational network. The Spanish data does not, however, demonstrate that only a network account can adequately describe the data. The distribution of intervocalic as opposed to word-initial clusters in Icelandic allows us to take the argument a step farther by demonstrating that the network does a superior job of processing sonority-sensitive constraints.

⁹A version of the following discussion was presented at the Chicago Linguistic Society parasession, The Cycle in Linguistic Theory, April 1992 and will appear in CLS 28:2.
For the past twenty-five years, Icelandic has proven to be one of the most
intransigent data sets for phonological theory. The degree of apparent variability in
nominal and verbal paradigms has challenged phonologists to find descriptively adequate
solutions that simultaneously serve as satisfying explanations of the phenomena. As a
result, as phonological theory has developed, Icelandic has served as a crucial test case for
nearly every major new theoretical innovation. Stephen Anderson's 1969 MIT dissertation
examined Icelandic within the developing paradigm of generative phonology (see also
Orešnik 1972). Kiparsky argued in 1984 that Icelandic demonstrates the superiority of
the lexical phonology framework by requiring the entirety of its theoretical arsenal to
account for the data. More recently, Itō (1986) has used the data as a case study for
prosodic phonology, again because of the fact that all of the major resources of the theory
are required for an adequate characterization of the data. Because of the complexity of
the data, Icelandic serves as an important test case for computational theories as well (cf.

3.3.1 The problem: Predicting vowel length in Icelandic

In an observation that has been adopted by nearly all subsequent investigators,
Venneman (1972) noted that vowel length in Icelandic is predictable from syllable
Stressed vowels and diphthongs are long in open syllables (20) and in closed
monosyllables with a single final consonant (21) and short elsewhere (22).

(20) Long vowels in stressed open syllables

\[
\begin{align*}
hi: . ti & \quad \text{'heat'} \quad \text{(data in from Itō 1986)} \\
o: . kur & \quad \text{'usury'} \\
su: . pa & \quad \text{'to sip'} \\
ho: . fuod & \quad \text{'head'} \\
sko: & \quad \text{'shoe'}
\end{align*}
\]

(21) Long vowels in monosyllables with single final consonant

\[
\begin{align*}
ha:s & \quad \text{'hoarse'} \\
ny:r & \quad \text{'new'} \\
spi:p & \quad \text{'ship'}
\end{align*}
\]

(22) Short vowels elsewhere

\[
\begin{align*}
hi:.ti & \quad \text{'(he) hit'} \\
sup:.tu & \quad \text{'sip!'} \\
har:.dur & \quad \text{'hard'} \\
el:.eka? & \quad \text{'love'} \\
björn & \quad \text{'bear'}
\end{align*}
\]
The key is to develop a syllabification procedure that correctly parses intervocalic clusters. Icelandic differs from neighboring languages by parsing the majority of VCCV sequences as VC.CV, whether or not the consonants conform to the sonority profile of a legitimate onset.

(23a)  
\[
\begin{array}{l}
\text{ep} . \text{li} & \text{'apple'} \\
\text{es} . \text{ki} & \text{'ash'} \\
\text{sig} . \text{la} & \text{'sail'} \\
\text{haeg} . \text{ri} & \text{'right'} \\
\text{af} . \text{laga} & \text{'out of order'} \\
\text{vel} . \text{ja} & \text{'choose'} \\
\text{tem} . \text{ja} & \text{'domesticate'}
\end{array}
\]

The preference for VC.CV syllables does not represent a universal syllable structure condition, however. Clusters formed from the union of the set \{p,t,k,s\} and \{r,v,j\} do prove to be tautosyllabic (V.CCV).

(23b)  
\[
\begin{array}{l}
\text{snu} . \text{pra} & \text{'chide'} \\
\text{so} . \text{tra} & \text{'sweet gen. pl'} \\
\text{vo} . \text{kva} & \text{'water'} \\
\text{tvi} . \text{svar} & \text{'twice'} \\
\text{e} . \text{sja} & \text{'a mountain'} \\
\text{ve} . \text{kja} & \text{'awaken'}
\end{array}
\]

Unfortunately, the problem is not so simple as to identify a segmental distribution for acceptable Icelandic onsets. The heterosyllabic clusters in (23a) prove to be well-formed tautosyllabic clusters in other environments.

(24a) Word-initial (#CCV)  
\[
\begin{array}{l}
\text{kli} . \text{fa} & \text{'climb'} \\
\text{pla} . \text{ta} & \text{'plate'} \\
\text{blað} & \text{'leaf'} \\
\text{brek} . \text{ka} & \text{'slope'} \\
\text{dra} . \text{ga} & \text{'to draw'} \\
\text{dver} . \text{gur} & \text{'dwarf'} \\
\text{djö} . \text{full} & \text{'devil'} \\
\text{skap} & \text{'temper'} \\
\text{flas} . \text{ka} & \text{'bottle'} \\
\text{fjos} & \text{'cattle'} \\
\text{fru} & \text{'Mrs.'} \\
\text{rju} . \text{ka} & \text{'smoke'} \\
\text{njo} . \text{ta} & \text{'enjoy'} \\
\text{mjolk} & \text{'milk'} \\
\text{ljo} . \text{tur} & \text{'ugly'} \\
\text{sta} . \text{fa} & \text{'spell'}
\end{array}
\]
(24b) Tri-consonantal intervocalic clusters (VC.CV)

\begin{itemize}
\item \textit{gil.dra} 'trap'
\item \textit{hel.dri} 'notable (compar.)'
\item \textit{tim.bri} 'timber (dat.)'
\item \textit{an.dvaka} 'sleepless'
\item \textit{af.griða} 'help, dispatch'
\end{itemize}

3.3.2 Itô's solution

Itô accounts for the distribution of intervocalic clusters (V.CV vs. VC.CV) by identifying four conditions on the syllable template of Icelandic (one universal, two set by parameter and one language-specific). The Universal Core Syllable Condition insures that VCV sequences will be parsed as V.CV by mandating that CV sequences be tautosyllabic. The Icelandic template contains both an "optional" coda position and two (or more) onset positions.\footnote{Strictly speaking, the coda position is not optional. If it is not filled by a consonant, the vowel is lengthened to accommodate the slot. Indeed, this forms the rationale for Itô's account of vowel lengthening.} After assigning the obligatory onset, the assignment of remaining intervocalic consonants to onset or coda is achieved initially by setting the \textit{directionality} parameter at Left-to-Right (i.e. coda slots are filled before onset slots). This results in the default assignment of VC.CV. The Icelandic Tautosyllabicity Condition overrides this default assignment when \{p,t,k,s\} are followed by \{r,j,v\}. Additionally, while L-R directionality maximizes codas, Icelandic fails to have VCC.CV because of a coda condition that limits the template to a single coda slot. Incidentally, VC.CV and #CCV are further limited by the general Sonority Sequencing Principle that requires that all onsets conform to the sonority hierarchy and a Minimum Sonority Distance Principle that rules out sequences such \*bn or \*tm that conform to the sonority hierarchy but are not a sufficient distance apart.

(25) Summary of Syllable Types / Conditions that generate syllables

\begin{itemize}
\item V.CV Universal Core Syllable Condition
\item VC.CV Universal Core Syllable Condition
L-R Directionality (maximizes coda)
\item V.trV Universal Core Syllable Condition
Icelandic Tautosyllabicity Cond. (overrides L-R Directionality)
\item VC.CCV Universal Core Syllable Condition
L-R Directionality (maximizes coda)
Coda condition (single coda slot)
(Sonority Sequencing Principle / Minimum Sonority Distance)
\end{itemize}
3.3.3 Critique -- Descriptively adequate but "implausible"

While Itô's set of templatic constraints does correctly parse intervocalic clusters and thereby generally accounts for the distribution of long and short vowels, it accomplishes this with a significant theoretical cost and, we would argue, with a high degree of implausibility (in terms of explanatory adequacy). Several considerations argue against Itô's analysis. First, Itô argues that Icelandic generally conforms to universal principles, requiring only one language-specific constraint, the tautosyllabic condition. This would suggest that the "oddity" in Icelandic is -

\[
\begin{align*}
(26a) & \quad V.prV & *Vp.rV \\
& \quad V.kjV & *Vk.jV
\end{align*}
\]

rather than -

\[
\begin{align*}
(26b) & \quad *V.brV & Vb.rV & \text{Conforms to sonority hierarchy} \\
& \quad *V.plV & Vp.lV & \text{Maximal Onset Principle}
\end{align*}
\]

The latter syllabifications conform to the sonority hierarchy, the maximal onset principle and are the ones found in related languages. If an idiosyncratic condition is to be required for Icelandic, it seems more likely that it should account for those forms which are unexpected rather than those which are not. The postulation of a tautosyllabic condition is required by the fact that the more radical proposal of left-to-right directionality overgenerates codas.

The observation that Icelandic has left-to-right rather than right-to-left directionality (scanning) is also problematic. It seems an overly powerful device if the goal is simply to mark certain otherwise well-formed clusters as heterosyllabic in the VCCV environment (contra #CCV and VC.CCV environments as well as the twelve cases of V.CCV).\(^{11}\) It is also rather odd (though not impossible) to have a coda maximizing language where the coda is maximally constrained (one and only one segment word internally - VC.CCV rather than VCC.CV) while the onset can be significantly more complex. Historically, the change from right-to-left to left-to-right directionality with the concomitant addition of the tautosyllabic condition creates problems both for a model of historical change and for any realistic learning theory.

The most important difficulty created by Itô's analysis is that it appears to miss the crucial generalization concerning the relationship between the abstractly defined sets \{p,t,k,s\} and \{r,v,j\}. The former set represents the four lowest sonority consonants in Icelandic while the latter set represents the three highest sonority consonants (v is the Icelandic reflex of the front glide). Icelandic seems to place a condition on intervocalic

\[\text{In fairness to Itô, she also cites cyclic interactions between epenthesis and deletion as evidence for left-to-right directionality in Icelandic (Itô 1985). While this evidence will not be analyzed in detail in this article, alternative analyses are available for this data as well.}\]
onsets that requires them to be maximally separated with respect to sonority. Harris (1983), Levin (1985) and others have noted that onsets are frequently subject to a Minimum Sonority Distance Principle that excludes onsets such as English *pn and *ml. Levin (1985) further observes that this MSD varies between languages and frequently between onsets and codas as well. What appears to be unique to the Icelandic case is that even within onsets the MSD is contextually dependent, being greater in biconsonantal intervocalic clusters than in other environments (VCCV vs. #CCV or CCCV).

3.3.4 Computational alternative

In 3.2.2 and 3.2.4, we provided a computational alternative that accounts for the Minimum Sonority Distance Principle in Spanish using lateral inhibition within a dynamic computational network (cf. Larson 1990). While the function of lateral inhibition within the network is similar to that of the Minimum Sonority Distance Principle both in process and effect, it differs in one crucial respect. The MSD serves as a static condition between two segments (i.e. the sonority of the second segment in an onset cluster must exceed the sonority of the second by a fixed amount). In the network account, however, the "distance" constraint depends not only on the sonority difference between the two segments but also on their context. In the paradigm MSD case (e.g. English *mla, negative α), the cluster is deviant because the high sonority vowel inhibits the segment to its left to such a degree that the second consonant is left with lower derived sonority than the first. In other words, this MSD constraint is due to the fact that the onset consonants are too close in sonority in the environment where they precede a vowel (or other high sonority segment). As a result, the dynamic model predicts that the environment of preceding and following segments should potentially affect the well-formedness of onset clusters. Icelandic provides precisely the phenomena that should allow us to choose between the static and the dynamic sonority models. The well-formedness of onset clusters in Icelandic is dependent on the preceding environment (V vs. #/C). Just as a negative α can dictate the permissible distance between the consonants that precede a vowel, a positive β can determine the degree of affinity the initial consonant has to its preceding vowel. As β becomes more positive the post-vocalic consonant increases in sonority, thereby becoming more likely to be realized as a coda rather than as initial onset segment (through in the sonority wave). Since α, β, and the sonority coefficients for each segment interact dynamically, the empirical question is whether or not a descriptively adequate set of coefficients can be discovered.

While the present goal is to simply stipulate descriptively adequate values that can account for the distribution of clusters rather than test any inductive learning algorithm, the learning procedure discussed in chapter six is, in fact, the most efficient mechanism for finding such values.12 We therefore used the following learning simulation to discover

---

12 As noted above, the network architecture is consistent with a wide variety of learning theories, including stipulation or hard-wiring. The use of an inductive learning procedure to determine descriptive adequacy does not presuppose that the language acquisition device functions in the same way any more than the
appropriate coefficients: For a test corpus we used the 65 Icelandic forms cited in Itô (1986), coded both with the broad phonetic transcription provided by Itô and the syllabification predicted by her analysis (see Appendix II.7). At the beginning of the simulation, the network was seeded with random sonority values (0.00 to 10.00) for each of the segments in the phonological inventory and random values (-.25 to +.25) for $\alpha$ and $\beta$. The words in the corpus were presented in random order to the network with the coefficients randomly modified in the event of an error (an error being defined as any discrepancy between the experimental and the target syllabification) using a modified simulated annealing algorithm (cf. 6.3; Larson, CLS 1992a). After each of the 65 words were presented they were re-randomized, thereby insuring that the order of presentation did not influence performance. Once the network either successfully learned to correctly parse all 65 forms simultaneously or when the network froze (temperature equal to 0), the results were tabulated, and the ending values of each of the coefficients were recorded.

The network proved to be very successful in correctly parsing all of the Icelandic data, achieving 100% performance in 172 of the 175 replications of the above procedure (in the other three tests the network correctly syllabified 64 of the 65 words). Over the course of the entire experiment, the network achieved .9997 performance. If the final values of alpha and beta for each replication are plotted in a two-dimensional space, the scatter plot in Figure 3.27 is generated.

Figure 3.27

---

paradigm building process of descriptive linguistics presupposes that children develop their paradigms through induction.

13 As will be noted below, the syllabifications adopted by Venneman (1972) and following may simply be theory-internal judgments not supported by phonetic measurements (cf. Orešnik & Pétursson 1978). The decision to follow Itô has no effect on the ability of the network to find descriptively adequate values, however.

14 To insure the replicability of the results with a variety of initial weights, the entire procedure was repeated 175 times, beginning each time with randomized values and ending when the network froze in an energy minimum/harmony maximum (cf. Touretzky, PDP 1986; Smolensky 1988).
A more precise plotting of the solution space can be generated by saving a trace of all presentations made during the entire experiment (1.1 million words). For each presentation, the system records the success or failure of the syllable network in a matrix with alpha and beta as its x- and y- coordinates (Figure 3.28). The elevation contours represent the overall percentage of success.\textsuperscript{15}

The plots of the alpha/beta solution space reveal a fairly broad range of values for alpha and beta where the network performs well. More importantly, as noted above, the crucial predictor of success is the positive beta. Alpha appears to be able to take a range of both positive and negative values as long as beta is sufficiently positive for the vowel to attract the appropriate coda consonants (i.e. all but the lowest sonority $p,t,k,s$).

Figure 3.28

It is also instructive to compare the plot of the alpha/beta solution space for Icelandic with the solution space for English (Figure 3.29).

\textsuperscript{15}The actual values represented in the right column are deflated by the fact that in the initial stages of each replication, the network makes errors because it is not yet learned the appropriate sonority vector. The actual performance of the network needs to be represented in 44-dimensional rather than 2-dimensional space. These early errors don't affect the relative mapping of the data, however, since they are distributed randomly throughout alpha/beta space.
While the two plots are different, they are not nearly so different as Itô's characterization of the syllable template for the two languages would suggest. The computational network more appropriately accounts for the evolutionary differences between historically related languages. Relatively small changes in alpha, beta and/or the sonority coefficients can result in apparently significant paradigmatic differences. Viewing the differences between Germanic languages in terms of a variable leftward affinity to vowel (positive β), potentially offers a valuable insight into historical linguistics as well as language acquisition theory.

3.4 Imdlawn Tashlihyyt Berber\textsuperscript{16}

In addition to being able to characterize cross-language differences in the parsing of intervocalic clusters (Icelandic vs. West Germanic) and more generally, the distribution of acceptable onset clusters (Spanish, cf. Icelandic), the dynamic computational network allows us to characterize more complex syllabification systems as well. Dell & Elmedlaoui's (1985) analysis of syllabification in Imdlawn Tashlihyyt Berber (subsequently Berber) is a classic description of a language where even the syllable peak cannot be directly predicted by a featural description of the segments in the language (i.e. +syllabic vs. -syllabic). Given an appropriate environment, it appears any segment in Berber can potentially appear as the syllable nucleus (we designate the syllable nucleus with a capital letter; a complete list of syllabified forms from Dell & Elmedlaoui appears in Appendix II.7; the parsing of intervocalic clusters is determined by Dell & Elmedlaoui's syllabification algorithm, native speaker judgments are inconclusive).

\textsuperscript{16}A earlier version of the discussion concerning Berber appeared in a article co-authored with John Goldsmith, "Local Modeling and Syllabification," CLS 26:2, The Parasession on the Syllable in Phonetics & Phonology.
In addition to allowing any segment to be licensed as a syllable peak, Berber syllabifications are sensitive to its inflectional morphology. The forms in (31) vary both with respect to syllable boundaries and the location of the peak in the second syllable.

### 3.4.1 Dell & Elmedlaoui's analysis

Dell & Elmedlaoui explain Berber's complex syllabification facts through an elaborate parsing algorithm. Rather than searching for syllable peaks directionally, Berber searches for peaks based on the sonority of the segments in the string, using the following sonority hierarchy:

1. **low vowel /a/**
2. **voiceless fricatives**
3. **voiced fricatives**
4. **nasals**
5. **liquids**
6. **high vocoids /i,u/**
7. **segments 1-6**
8. **segments 1-7**
The syllable parser scans left-to-right through the string to find the segment with the highest sonority. Once the highest sonority segment is identified, it is marked as a syllable peak if it word-initial or has an unsyllabified segment to its left. Once the syllable peak is marked, the segment to its left is incorporated into its syllable as an onset. If the incorporated onset segment happens to be a high vocoid it is realized as a glide. Once the procedure successfully creates a core syllable (CV), the process is repeated with the parser searching for the unsyllabified segment with the next highest sonority. Once all of the possible CV-syllables are created, the remaining segments are adjoined to the preceding syllable as a coda (coda adjunction). For example, the form /tluat/ would be syllabified as follows:

(34) Berber syllabification algorithm

1. Identify highest sonority segment: \( tu \ A \ t \)
2. Incorporate preceding segment as onset: \( tl \ uA \ t \)
3. Glide formation: \( tl \ wA \ t \)
4. Identify highest sonority segment: \( t \ L \ wA \ t \)
5. Incorporate preceding segment as onset: \( tL \ wA \ t \)
6. Identify highest sonority segment: \( tL \ wA \ T \)
7. *Incorporate preceding segment as onset: \( tL \ wA t \)
8. Adjoin unsyllabified segment to preceding syllable \( tL \ wAT \)

While Dell & Elmedlaoui's syllabification algorithm generally accounts for the data (phrase final syllables are permitted to be more complex), it is rather implausible. The parsing procedure is doubly iterative. Each time the parser attempts to produce a CV syllable, it must potentially go from the top of the sonority hierarchy to the bottom. The parser must then be able to determine if the preceding segment is syllabified and, if not, it creates a CV syllable. The parser must then return to the top of the sonority hierarchy and continue to create syllables until no eligible segments remain. It is only at that point that the coda adjunction procedure would be accessed. Such a procedure could not be completed in real time, requiring a number of buffer operations and not permitting local computations.

3.4.2 Network alternative

Goldsmith & Larson (1990) demonstrate that a dynamic computation can successfully model the syllabification phenomena of Berber. Using the same stipulated sonority values for each of the segments in the Berber inventory (32) and stipulated values for alpha (-0.60) and beta (-0.10),\(^{17}\) we produce sample analyses that demonstrate how the network successfully syllabified those words with consonantal syllable peaks (see Appendix II.8 for database).

\(^{17}\)In the notation used in Goldsmith & Larson (1990), inhibitory values for alpha and beta were represented by positive numbers. Since all alpha/beta influences were presumed at that time to be inhibitory, the activation function used minus signs where now they have plus signs, a simpler notation when one permits alpha and beta to take both positive (excitatory) and negative (inhibitory) values.
(35) (α / β) t l u a t
a. 0.0 / 0.0 0 5 7 8 0 (inherent values)
b. -0.3 / 0.0 -1.09 3.62 4.60 8.00 0.00
c. -0.5 / 0.0 -1.75 3.50 3.00 8.00 0.00
d. -0.5 / -0.15 -2.11 4.22 2.19 8.35 -1.36
e. -0.6 / -0.10 -2.60 4.34 1.54 8.38 -0.89

Figure 3.36

In the above figure it should be noted that the inherent sonority values would predict that /tuat/ is monosyllabic. The inhibitory effect of α results in the high-sonority /a/ inhibiting the /u/ so that its derived sonority is less than that of the /l/ which precedes it. Since Berber potentially licenses all segments as syllable peaks (i.e. all local maxima are syllable peaks -- no threshold), the /l/ becomes the nucleus of the initial syllable. In 3.37 and 3.38, we observe that the network correctly syllabifies a wide range of data.
More important than the fact that values for alpha, beta and the sonority hierarchy can be stipulated such that Berber words can be appropriately syllabified is an understanding of why certain values should work. In the discussion of Spanish and Icelandic, alpha proved to be the determinant of the Minimum Sonority Distance for an onset. One way that we can look at the Berber data is that Berber has a sufficiently high Minimum Sonority Distance to insure that no onset clusters are permissible. Peaks have sufficient influence over their left neighbors to give them lower sonority than the next segment to the left. While the results of such leftward inhibition would make the resultant peak un licensable in Spanish and Icelandic, Berber permits the resulting local maxima to serve as syllable nuclei. Rather than representing a radically different syllabification system, Berber simply varies with respect to the threshold parameter for well-formed syllable peaks. Otherwise, Berber functions like Spanish and Icelandic in that it utilizes the Minimum Sonority Difference Principle to limit the complexity of onsets.

Before leaving the discussion of Berber it is important to note that the issue is more complex than either Dell & Elmedlaoui's account or the stipulative network account suggests. Both models seem to suggest that Berber prohibits complex onsets, a generalization which also could have been captured by a templatic constraint that prohibits branching in the onset constituent. But Berber does permit onset clusters in certain environments. Dell & Elmedlaoui report the following:

(39)  zwI 'beat down'
    twR . tAt 'kind of feline'.
    tlUr . tNt 'she gave them (f.) back'
    contrast tL . wAt (Figure 3.34)
    txZ . nAkkw 'she even stockpiled'
    thX . lAkkw 'she even behaved as a miser'

As described above, none of the preceding forms would be syllabified in accordance to Dell & Elmedlaoui's syllabification algorithm. But the exceptions are not unprincipled. In
each case, a strict application of the CV syllable rules would leave the first segment un syllabified. The contrast between *tlwAt* and *tlUr.tNi* illustrates the difference. While both words have the same first three underlying segments (*tlu*), the crucial contrast is between the fourth segment (*la* vs. *r/)*. In the first word, the *la* has the highest sonority. In the second, the *lu* has the highest sonority. By the ordered procedure of Dell & Elmedlaoui, the selection of the highest sonority segment produces a chain reaction that leaves *lu* un syllabified in the second form but not the first. It should not be surprising that such an un syllabified obstruent should not serve as a syllable peak even if they can do so in word internal positions. What is left unexplained, however, is the procedure by which the word-initial un syllabified obstruent is adjoined to the onset.

In the network account, what we observe is a competition for syllable affiliation. Since high-sonority segments possess higher energy that both protects their status as peaks and makes them potentially more active in influencing their neighbors through inhibitory connections, the network models the same effect as Dell & Elmedlaoui. The network can perform all the calculations locally and, in a feedforward architecture, in a single pass. The network also has the ability to make gradient rather than categorical decisions. The un syllabified word-initial obstruents which must be adjoined to onsets by a stipulative procedure in Dell & Elmedlaoui's account can be made to fit into well-formed obstruents in the network (i.e. the derived sonority values of the onset segments conform to the Sonority Sequencing Principle). With appropriate values of alpha, beta and the sonority vector, each of the onsets in (39) can be predicted as well-formed in the word-initial environment (particularly if the network includes initial and/or final activation, see 2.5.1, fn. 11).

It should also be noted that the sample analyses offered in Goldsmith & Larson (1990) are not descriptively adequate for Dell & Elmedlaoui's entire corpus. The stipulated values worked adequately for the illustrated forms in (31) and (32). If applied to all 81 forms cited in the article, the stipulated values fail at several points. Attempting to stipulate any set of descriptively adequate values proves to be an exceedingly difficult task (i.e. attempting to simultaneously account for regular CV syllables as well as coda adjunction and the exceptional onset adjunction forms). Fortunately, the learning device proposed in chapter six provides an automatic mechanism by which the appropriate values can be induced (α = -.35; β = -.10; 96.4% average performance). One of the principal advantages of the network architecture is the opportunity to rapidly modify and test a variety of values for each of the parameters, a procedure that would be very difficult using paper and pencil.

3.5 Problem cases for sonority hierarchy

Before completing the study of syllabification within a dynamic computational network, several problem cases demand attention. Clements (1987) offers three classes of exceptions to the traditional sonority hierarchy.
(40a) Sonority plateaus within a syllable

English: *apt, act, sphere
Russian: mnu, ikut, kto, gd’e
Marshallese: qqin, kken, lliw

(40b) Sonority reversals within a syllable

English: spy, sty, sky, axe, apse, adze
German: Spiel, Stein, Sklave, Obst, hatst
Russian: rta, lba, mga
French: table [tabl], autre [otr]

(40c) Syllable peaks that are not sonority peaks

English: you vs. we, rural [rr-a]
French: troua [tru-a] vs. trois [trwa], Louis [lu-i] vs. lui [l i]
Spanish: pais [pa-is], rio [ri-o], bahia [ba-i-a]
Turkish: daa [da-a], cii [ci-i]²

3.5.1 Sonority plateaus

The first class of problem cases concern forms where adjacent segments within a single syllable have the same underlying sonority value. As Clements notes, this would only appear to be a problem if one adopts the strong version of the Sonority Sequencing Principle where adjacent segments must have a Minimum Sonority Distance greater than zero (cf. Sievers 1893). In the less restrictive version defended by Jespersen (1904), sonority plateaus do not violate any well-formedness condition. Unfortunately, the issue is not quite that simple. The data in (40a) represent three different conditions. In English, apt, act are well-formed but *atp and *atk are not. When /s/ is introduced into word-final clusters, an interesting pattern appears (the distribution of /s/ will be discussed in 3.4.2).

(41)

\begin{align*}
\text{acts} \quad [\text{akts}] & \quad *\text{atks} \\
\text{axed} \quad [\text{akst}] & \quad *\text{atsk} \\
\text{asked} \quad [\text{askt}] & \quad *\text{astk}
\end{align*}

While the /s/ can appear in any position with respect to the /k/ and /l/, the latter are crucially ordered with respect to each other. In general, coronals must be peripheral to non-coronals in English. As a result, it would incorrect to suggest that coronal and non-coronal stops have the same sonority value and that English accepts sonority plateaus. It

²The putative incongruity between sonority peaks and syllable peaks is accounted for by lateral inhibition (see 3.3.2).
would rather appear that English distinguishes between different points-of-articulation in
determining the sonority of obstruents.

Although English might not prove to be an example of a language that accepts
sonority plateaus, its discrimination among voiceless stops raises an important objection to
argues that the ordering of classes of segments within the sonority hierarchy is universal
but that languages may vary with respect to how many distinctions are recognized.

(42) *Universal sonority scale* (Harris 1989)

<table>
<thead>
<tr>
<th>← less sonority</th>
<th>more sonority →</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obstruents = [-sonorant]</td>
<td>Resonants = [+sonorant]</td>
</tr>
<tr>
<td>Consonants = [+consonantal]</td>
<td>Vowels = [-cons]</td>
</tr>
<tr>
<td>Stops = [-cont]</td>
<td>Continuants = [+cont]</td>
</tr>
<tr>
<td>[-voice]</td>
<td>[+voice]</td>
</tr>
<tr>
<td>p t b d f s v z m n l r i u e o a</td>
<td></td>
</tr>
</tbody>
</table>

Unfortunately, even the order of the segments is arguably non-universal. Jespersen's
sonority hierarchy conflates voiceless stops and voiceless continuants and orders both as
less sonorant than voiced stops (contra Harris). Levin (1985) argues that languages vary
with respect to the relative ordering of /r/ and /l/, the relative ordering within the class of
voiceless stops and within the class of nasals. Rather than disconfirming the present
analysis, however, the non-universality of the sonority hierarchy offers evidence for an
approach where sonority can be induced by a learning device from the distribution of
segments within syllables in a given language (see 6.3). The learning device is capable of
not only increasing or decreasing the number of distinctions made with the sonority
hierarchy (including the ability to make distinctions greater than or less than 1), but also
determining language-specific ordering of individual segments.

The Russian evidence presents a somewhat different problem. While the voiceless
stops in English appear to be crucially ordered with respect to each other, /k/ and /t/
appear in either order in word-initial onsets in Russian. On one hand, it might be possible
to argue that /k/ and /t/ are empirically equal on the sonority scale and that sonority
plateaus are acceptable. The computational network offers an additional possibility,
however. The dynamic contextual influences represented by alpha and beta make it
possible to transform an ill-formed sequence 1-1-5-1-1 into a well-formed syllable.
3.5.2 Sonority reversals

An even more troublesome phenomenon for traditional sonority analyses involve cases where the sonority values of adjacent segments are actually reversed (e.g. where the first segment in an onset has higher sonority than the second). Whether the weak or strong version of the Sonority Sequencing Principle is adopted, sonority reversals are a prima facie violation. Two basic kinds of violations appear in (40b). In the case of English and German, the problem concerns the idiosyncratic behavior of sibilants such as /s/. In Russian and French, the sonority reversals involve a wide variety of segments that are more easily categorized than the /s/.

One of the most difficult problems for analyses that rely on the traditional sonority hierarchy regards the syllabification behavior of /s/ in languages such as English, Spanish and German. Whether one places it in the sonority hierarchy with other segments that share the same general class features (i.e. voiceless continuants) or in some unique category, /s/ appears to defy the assignment of any single sonority value.

\begin{itemize}
\item[(44)] a. stop \hspace{1cm} *t\text{top} \hspace{1cm} *\text{top}
\item[b. pots] \hspace{1cm} *potf \hspace{1cm} *pot\text{\textdagger}
\item[c. *t\text{top}] \hspace{1cm} *t\text{fop} \hspace{1cm} *t\text{\textdagger\textdagger}
\item[d. post] \hspace{1cm} raft \hspace{1cm} ?ba\text{\textdagger\textdagger} (ba\text{\textdagger\textdagger})
\item[e. posts] \hspace{1cm} *po\text{f}\textf \hspace{1cm} *po\text{\textdagger\textdagger}\textf
\end{itemize}

The initial problem with the patterning of /s/ follows the assumption that /s/ patterns with the voiceless continuants in the universal sonority hierarchy. If this were the
case, the forms in (44a), (44b), and (44c) would be incorrectly predicted to be ill-formed (compare forms with /t/ or /θ/). Even the forms which are correctly predicted fail to directly support the sonority hierarchy without additional provisos. The rejection of (44c) depends on the Minimum Sonority Distance Principle while the acceptance of its mirror image in (44d) requires the seldom stated stipulation that the MSD only apply to the onset of the syllable.

Potential fixes don't fare much better. If we were to assign /s/ a value lower than /t/, we could account for the peripheral /s/’s (assuming that the difference between /t/ and /s/ was sufficient to exceed the MSD) but we would then fail to account for the acceptability of the medial /s/ in (d); and regardless of where we place /s/ in the hierarchy, forms such as (e) /posts/ will always prove problematic. Three other solutions prove equally unsatisfying. It would be possible, but rather implausible, to argue that English has two different /s/’s, one which appears peripheral to voiceless stops and the other which appears medial to voiceless stops. Alternatively, one could argue that /sp/, /st/, and /sk/ form a special class where two segments can fill a single position within the syllable template (Selkirk 1982).19 While an auxiliary template might account for the peripheral /s/ in syllable-initial clusters (although /sm/, /sn/ and /sf/ are also well-formed), it does not account for peripheral /s/ in syllable-final clusters.20 A third solution would be to propose word-initial and/or word-final extrasyllability or appendices, particularly when the complex clusters result from affixation (Goldsmith 1990, Bosch 1988; cf. Sievers' notion of "auxiliary syllable" Nebensilbe). In addition to proposing an auxiliary template for /s/ + voiceless stop clusters, Selkirk uses word-level extrasyllability to account for tri-consonantal clusters in words like acts /kts/ or quadri-consonantal clusters such as texts /kstʃ/ or sixths /kstʃs/.

In general, all of the above solutions account for the problem of apparent sonority reversals by circumventing it, using additional machinery that is explicitly not sensitive to the Sonority Sequencing Principle. The computational network offers a much simpler solution that requires no additional machinery or exceptions to general principals. In our early work with automatic learning devices, we noted that the network performed differently from one run to another with respect to the sonority value assigned to /s/. In half of the cases, /s/ patterned with the voiceless continuants. In these simulations, the network performed poorly on peripheral /s/’s, particularly since we tended to limit alpha and beta to negative values. In the other half of the simulations, the network selected a value of /s/ lower than any other segment in the inventory, and in those cases network performance was superior. When alpha and/or beta were permitted to take positive as well as negative values, the network correctly syllabified all of the forms containing /s/. In the general description of the network architecture in chapter two, it was noted that

19 Defining a special class of /sp/, /st/, and /sk/ is, in a formal sense, not much different than proposing two different /s/’s.

20 Selkirk (1982) also uses the auxiliary template to account for the /st/ cluster in next [nekst]. Considering the /st/ in next to be an example of the auxiliary template creates several problems, however. The /s/-clusters cannot appear in all positions filled by ordinary stops (e.g. first position in a coda cluster. *nespt). It would also be rather odd to interpret next [nek+st] differently than axe [æks] and axed [æks+t], though the latter might excluded as bi-morphemic.
negative values of alpha and/or beta reduce the number of segments that can fit within an onset or coda cluster while positive values would increase the number of potential clusters.

The learning device correctly accounts for the distribution of /s/ in English most easily by assigning it the lowest sonority value and assigning a negative value for alpha and a positive value for beta, thereby simultaneously accounting for the onset/coda asymmetry (sharp rise / slow fall).\textsuperscript{21} As we examine each of the four potential positions for /s/ within the syllable, the network solution appears quite plausible. While traditional analyses suggest that medial /s/'s should be expected while the peripheral /s/'s are idiosyncratic, the paradigm in (44) suggests the opposite. The only unacceptable cluster is the stop+/s/ onset (44c, Figure 3.45), precisely the expected result with a negative alpha and positive beta (the selected coefficients are for illustration purposes).

\textsuperscript{21}A slightly different proposal will be considered in 5.1.2.
Most crucially, the network also successfully syllabifies the case where /s/ is both medial and peripheral (e.g. posts, Figure 3.49).

![Figure 3.49](image)

Whether or not we analyze the distribution of /s/ in English and other European languages as a sonority reversal, unambiguous sonority reversals appear in Slavic languages such as Polish and Russian:

(50)  
**Russian:** rta, lba, mgla  
**Polish:** wtedy, zbudowac', kto'ry

In addition to the examples of sonority reversals (though neither language exhibits as many sonority reversals in the surface phonetics as they arguably represent phonologically; cf. Rubach & Booij 1990; Gorecka ms.), it has frequently been observed that Polish onsets exhibit much greater potential complexity than English or German. On the other hand, Gorecka argues that Polish codas are relatively constrained both in their length and in the obligation to observe the Sonority Sequencing Principle. While the Polish and Russian onsets appear to be prima facie violations of the sonority hierarchy, they can be straightforwardly accounted for by considering positive values for alpha and/or beta (similar to the treatment of /s/ above). In order evaluate whether the network can specify descriptively adequate values for Polish, we used the learning algorithm described in chapter six to identify optimal coefficients. Using a corpus of 575 high-frequency Polish forms, over the course of 10 complete replications of the learning procedure, the network obtained the following performance (see Appendix II.9 for data):
(51) Polish (575 words)  

<table>
<thead>
<tr>
<th>Metric</th>
<th>Avg.</th>
<th>Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pct. correct (including intervocalic clusters)</td>
<td>.941</td>
<td>.951</td>
</tr>
<tr>
<td>Pct. correct (excluding intervocalic clusters)</td>
<td>.971</td>
<td>1.000</td>
</tr>
<tr>
<td>Value of $\alpha$</td>
<td>.150</td>
<td>.048</td>
</tr>
<tr>
<td>Value of $\beta$</td>
<td>.091</td>
<td>.072</td>
</tr>
</tbody>
</table>

3.5.3 Syllable peaks that are not sonority peaks

In Clements' discussion of empirical difficulties created by the sonority hierarchy three types of phenomena are included in his final category. In English, the contrast between you /ju:/ and we /ui/ (as well as the general non-occurrence of iw or uy), the prediction of the syllable peak and the resulting glide formation can not be based solely on the sonority hierarchy, requiring some form of directionality as well. To the extent that English selects either the higher sonority segment or the segment on the right as the peak (with the non-peak becoming non-syllabic), the distribution supports the above selection of values for alpha and beta in English. In the competition between two empirically equivalent vocoids, a negative alpha and/or a positive beta would cause the segment on the right to have higher derived sonority and thereby be realized as the peak.

Unfortunately, the English data is somewhat more complex (e.g. buoy /bu-i/ where both /u/ and /i/ are syllable peaks). As described in chapter two, the network can not represent two adjacent segments that are peaks of different syllables since the wave definition of a syllable requires both a peak and a trough. Even if the English examples such as buoy were dismissed as peripheral, adjacent syllable peaks appear frequently in Spanish (when one vowel is accented), French and Turkish (where the two vowels can be identical). As a result, an empirically adequate network model must accommodate adjacent peaks.

Several possible explanations are available. First, it might be possible to argue that phonetically the two adjacent peaks are actually separated by some form of transitional onset (glide, glottal stop, or /h/). While a significant number of cases permit such an interpretation, native speaker judgments generally argue against the presence of a transitional onset in French (Laks, personal communication). On the other hand, it might be possible to propose the existence of a null phonological onset in the spirit of Kaye, Lowenstamm & Vergnaud's obligatory onset condition (cf. Charette 1988). It would also be possible to change the definition of a syllable peak from one which requires a segment to be a local maximum to one which only requires that it exceed a threshold (parallel to the definition of stress peaks in languages which do not demand stress clash avoidance, e.g. Eskimo 4.4.1). Finally, it might be necessary to ultimately change the representation of time and segment adjacency within the network to one which more directly represents real time relationships within the phonological string (cf. 2.7).
3.6 Conclusion

Before concluding the discussion of syllabification within a dynamic computational network, it is instructive to observe the distribution of syllabification systems within the convergence zone of alpha/beta values plotted in chapter two. Noting in general that values of alpha tend to affect the distribution of segments to the left of the syllable peak while values of beta influence the distribution of segments to the right of the peak with positive values increasing potential complexity and negative values decreasing potential complexity, we can plot the various languages that have been described above.

![Convergence Values Diagram](image)

Figure 3.52

In addition to noting the parametric variation between languages, the model provides interesting theoretical possibilities for the description of dialectal variability and language change. In addition to serving as a device that can parse and generate syllables, the network offers interesting predictions concerning the underlying mechanism that determines the various syllabification phenomena that appear in the world's languages.
Chapter 4

Stress Assignment

In the previous chapter, we demonstrated that syllabification phenomena can be effectively modeled using the resources of a dynamic computational network. We will now turn to the next logical level of phonological structure by examining stress assignment procedures that can be modeled by the network. Several factors make the computational network an ideal device for modeling metrical phenomena. Within the majority of current metrical theories the syllable typically serves as the interface between the segmental representation and the metrical representation. In a similar fashion, within the dynamic computational network, the syllable serves both as the principal output unit of the segmental network and as the basic input unit of the metrical network. The segmental network provides a natural device not only for the construction of syllables but also for the identification of weight distinctions that are important in quantity-sensitive stress systems.

The input layer of the metrical network also allows us to represent the variety of additional parameters that affect stress placement -- proximity to an edge, lexically and/or morphologically assigned stress, etc. More critically, lateral inhibition/excitation permits the effective mapping between complex patterns of input activation and the output stress profile. Rhythmic stress patterns (i.e. iterative binary feet, perfect grid) exemplify the cyclic waves that are propagated by a spreading activation network with negative alpha and/or beta (cf. 2.4.2.2, Figure 2.15). Stress phenomena such as unbounded feet, stress clash avoidance, stress retraction, cyclic stress, and extrametricality are similarly modeled by appropriate settings of alpha and beta combined with appropriate input values.

In addition to the fact that the network is ideally suited for the modeling of stress assignment systems, metrical theory serves as an ideal domain for the evaluation of the network. The complexity and range of variation among stress systems is sufficiently small so that it can serve as a tractable data set. Dresher & Kaye (1990) hypothesize a solution space of 216 possible stress systems given eleven parameters with significant cross dependencies. While Dresher & Kaye admit the need for additional parameters that would increase the size of the solution space, the range of possible stress systems still appears to be quite constrained. Halle & Vergnaud's Essay on Stress (1987) discuss in some detail 34 different stress systems which arguably represent paradigm cases. Touretzky & Gupta (1992) further narrow the discussion to 19 crucial systems (ten major patterns with mirror images of nine of the ten). While the latter discussion overly limits the domain of analysis, the presence of a relatively limited and well-defined set of stress systems permits the careful comparison of competing theories. Metrical theory has also achieved a degree of formality that permits rigorous evaluation. While there are several variants of metrical theory that warrant discussion, Halle & Vergnaud (1987) will serve as the principal point of departure for this analysis. In addition to having a well-defined set of parameters, Halle & Vergnaud formalize an ordered set of rules and a formally constrained production device. Dresher & Kaye (1990; cf. Dresher 1991, 1992) further

---

1Without cross dependencies, the eleven parameters would describe $2^{11} = 2048$ formally possible systems.
extend the theory by proposing a learning device to acquire the parameters.\textsuperscript{2} Since the proposed dynamic computational network combines a description of grammatical competence with a production and acquisition system, we will be able to compare its account of stress assignment with standard metrical theories.

The plan of the current chapter will be as follows: First we will describe the architecture of the metrical network, emphasizing the available parameters of variation (degrees of freedom). We will then briefly describe traditional metrical theory (à la Halle, Vergnaud, Idaardi, Dresher, Kaye, et al.), outlining its principles, parameters, production system, and acquisition devices. The third and fourth sections will test the metrical network on a representative set of quantity insensitive and quantity sensitive stress systems, comparing and contrasting the theoretical machinery used in the network vis-à-vis that of metrical phonology.\textsuperscript{3} The final section will address general evaluative issues such as generative power, learnability and the choice of appropriate evaluation metrics.

4.1 Architecture of Metrical Network

The metrical network possesses the same gross structure as the syllable network discussed in chapter three. It consists of three layers of units (input, processing, identification) which are further segregated into vertical columns associated with each syllable peak identified by the segmental network.\textsuperscript{4} While the structure is similar, the variables encoded within the network are somewhat different. Whereas the segmental network uses a sonority vector as its input, the metrical network possesses the following adjustable parameters which help define its power (degrees of freedom): Initial (I), Final (F), Weight (L/H), Inherent Lexical and/or morphological stress (M).

4.1.1 Input: Positional activation (initial/final)

One of the most important determinants of stress placement is positional activation associated with particular positions in the string of syllables. While it would be possible to

---

\textsuperscript{2}Touretzky and Gupta (1991) describe a connectionist framework that uses a perceptron learning model to map between strings of syllables marked for syllable weight and stress patterns in 19 representative languages.

\textsuperscript{3}The current model has been tested on the following languages: Latvian, French, Maranungku, Weri, Garawa, Lakota, Polish, Paiute, Warao, Koya, Eskimo, Malayalam, Yapese, Ossetic, Rotuman, Komi, Eastern Cheremis, Khalkha Mongolian, Aguacatec Mayan, Lenakel, Aklan, Cayuava, Western Aranda, Pirahã, Latin, Cairene Arabic, Tiberian Hebrew, Southern Paiute, Klamath, Macedonian, Russian, Sanskrit, Indonesian, Yidin\textsuperscript{5}, Winnebago, Creek, Lithuanian, Spanish, Odawa, and English. Only a small subset of these languages will be explicitly discussed in this chapter, however, since the analyses are only meant to be descriptive of the network's capabilities.

\textsuperscript{4}Halle & Vergnaud (1987) cite evidence that in some languages, the line 0 primitives need to be moras instead of syllable heads (e.g. Southern Paiute, Cayuava, Winnebago). The treatment of Winnebago within the current framework will be considered in 4.4.3. While it may be possible to account for these "mora-counting" languages as quantity-sensitive phenomena that use only syllable heads as primitives, it is also possible (if necessary) to modify the segmental network so that it recognizes moras instead of syllable peaks (using thresholds instead of local maxima).

---
assign positional activation to an arbitrary number of positions in the string (e.g. first two, last three, etc.), the current model constrains the power of the network by only assigning positional activation to two positions: initial (I) and final (F). It should be noted that unlike the parameter settings in Dresher and Kaye (1990), it is possible that both I and F could take non-zero values in the description of a given language. It is also possible that I or F could be negative, generating much the same result as extrametricality in traditional analyses.

4.1.2 Input: Syllable weight (heavy/light/bias)

Traditional metrical analyses distinguish between two types of stress assignment systems: quantity-insensitive and quantity-sensitive. In the former, stress is assigned to a syllable solely as a result of its position with respect to one (or both) of the edges of the string. The internal structure of the syllable is irrelevant to stress placement in such systems. In quantity-sensitive systems, however, the internal structure of the syllable is relevant to stress placement. Typically, syllables are segregated into two types: light and heavy, with languages varying as to what kinds of syllables are assigned to each category. Dresher & Kaye, for instance, identify languages that consider syllables with long vowels (CVV) to be heavy (nucleus sensitive) and others that consider all closed syllables (CVC) to be heavy (rime sensitive). Prince (1983 check reference) further distinguishes between languages with respect to what coda segments can serve to make a syllable heavy. Clements (1987) notes that such languages form a cline corresponding to the sonority hierarchy (i.e. if a nasal closes a syllable, a liquid will also close it but not necessarily the reverse). Everett & Everett (1984) extend the range of weight sensitive phenomena by noting that syllable weight can also be sensitive to the content of onsets in Pirahã (cf. Harris' (1983) discussion of Spanish on-glides). Once syllables are categorized as heavy and light, they can influence the pattern of stress, with heavy syllables typically attracting stress and light syllables often being destressed. Halle & Vergnaud represent the traditional handling of heavy syllables by assigning them line 1 asterisks in basically the same way other line 1 asterisks appear.

In the metrical network, information concerning syllable weight is input from the segmental network along with the identification of syllable peaks. In other words, each syllable peak outputs an activation value that is determined by its weight (chapter three discusses the ability of the syllable network to efficiently compute weight). While it is possible that such weight information could be represented as a continuous variable, the current model uses a step function so that the weight output of the segmental network corresponds to the number of moras identified in traditional metrical analyses. The

---

5 A somewhat similar configuration is permitted by Halle & Vergnaud by allowing different parameter settings on each line in the metrical grid. For instance, line 0 could be -BND L while line 1 is -BND R.

6 The attraction of stress to heavy syllables is only a strong tendency. In pre-accenting or post-accenting languages such as Squamish (Kuipers 1967; Demers & Horn 1978), a syllable adjacent to heavy syllable ends up being stressed on the surface, whether or not it is itself heavy.

7 It is quite possible that a continuously valued weight function would more accurately model complex stress systems such as English, but such a discussion is beyond the scope of this study. The fact that
network is unique, however, by permitting the connection between the mora counter and
the metrical input unit to take a continuously valued weight. In other words, the metrical
network can learn to assign a variety of values to either light or heavy syllables. In
quantity-insensitive languages, where all syllables are treated equally, the network can
accommodate that fact in one of two ways. For such languages, the step function that
counts moras can be designed so that all syllables have the value 1. As an alternative, the
connection between the syllable network and the metrical network can have a zero weight
so that all quantity distinctions are eliminated. At present, we adopt the former option.
Since languages differ with respect to what counts as a mora, the mora counting function
will have to be variable even for quantity-sensitive languages. Quantity-insensitive
languages would simply be a special case where no syllable crosses the threshold that
would assign more than one mora to the syllable. Assigning all quantity-insensitive
syllables the same input value as we use for light syllables in quantity-sensitive languages
permits the introduction of a third weight option, a variable we will call bias. One of the
features of grid-based analyses is that we distinguish between those syllables which are
active (line 1 asterisk) and those that are not (line 1 dot). Translating this into a network
representation, we might expect that those syllables which are active have positive input
activation and all other syllables have zero input activation. It is equally possible that
these default syllables could have some non-zero activation value, however. For instance,
Goldsmith (in press) discusses syllabification differences which can be captured by
assuming that the bias is positive in one case (Yapese) and negative in the other
(Rotuman). Although Goldsmith suggests that non-zero bias should be uniquely available
in quantity-sensitive systems, there is no formal reason for excluding it as a possible
variable in quantity-insensitive systems as well.

4.1.3 Input: Lexical stress

In addition to positional activation and activation due to syllable weight or
quantity, languages frequently assign inherent stresses within the lexicon or morphology.
Perhaps the biggest distinction between the current approach and grid analyses is the fact
that each input activation can take a continuous range of values, both positive and
negative. In the grid, all line 0 asterisks are the same, regardless of their source. The fact
that positional stresses tend to be stronger than stresses produced by binary constituent
construction and those associated with heavy syllables is rather the result of line 1
constituent construction that tends to add additional asterisks at the periphery of the
string. Additional distinctions are produced by rules that add or subtract asterisks rather
any potential differences between the asterisks that are initially assigned. 8

8 Halle and Vergnaud do distinguish between two different kinds of constituent boundaries: intrinsic vs.
dependent, that are differentially affected by subsequent rules (e.g. the Domino Condition). This
distinction potentially has the effect of distinguishing between different kinds of asterisks, though less
efficiently than permitting input parameters to take non-equivalent values.
4.1.4 Processing: Lateral inhibition ($\alpha/\beta$)

In the discussion of syllabification in chapter three, it was noted that roughly equivalent results can be achieved using either a feedforward network with a limited number of input connections to adjacent units or a recurrent network with lateral inhibition between a unit and its left and right neighbors. Since the former is computationally less expensive, it was adopted as the architecture of choice for the segmental network. Metrical phenomena provide a stronger warrant for a recurrent network. One of the principal tasks in metrical phonology is to account for putatively iterative phenomena such as perfect grid or rhythmic stress. Traditional accounts generate rhythmic patterns by building binary constituents (ternary in the case of Cayuvara; cf. Halle & Vergnaud 1987). A recurrent net with negative values for alpha and/or beta naturally produces the same rhythmicity, whereas a feedforward network with local connections would not. One of the biggest differences between syllabification phenomena and metrical phenomena appears at precisely this point. While languages may indeed prefer the rhythmicity inherent in CV syllables, it is not unusual for a syllable to, in fact, contain more than two segments. In such situations, feedforward connections are often better behaved than recurrent connections. When syllabification phenomena are modeled using the latter, alpha and beta typically take rather small values, thereby minimizing the long distance effect of the connection. In rhythmic stress systems, alpha and/or beta typically take relatively large negative values in order to insure that rhythmicity is reproduced over an arbitrarily long string. Recurrent networks also prove to be advantageous in the description of quantity-sensitive stress systems that only realize a single surface stress. In languages such as Komi, Eastern Cheremis, Mongolian and Aguacatec Mayan, a positive value of alpha and/or beta creates a cumulative stress pattern that selects left or rightmost heavy syllables. Goldsmith (in press) describes the coordinate space of convergent values of alpha and beta (Figure 4.1). When alpha and/or beta are negative rhythmic stress patterns are produced, when alpha and/or beta are positive cumulative stress patterns are generated, and when alpha and beta have significant values but different signs one end of the string has rhythmic stress while the other end fails to.
While metrical phenomena frequently require iterative alpha and beta connections ($\alpha_1/\beta_1$), several stress systems are appropriately modeled by simple feedforward connections ($\alpha_0/\beta_0$). Languages such as Malayalam, Ossetic, Yapese, and Rotuman, which are quantity-sensitive but limit stress to the first two or last two positions prove to be excellent candidates for strictly feedforward local connections (see 4.3). In a learning model, it is possible that all four parameters are available, with two (or more) converging on a zero value.

4.1.5 Identification: Local maxima, Global maxima, Thresholds

Once the network takes a set of input activations and computes a derived activation based on contextual influences (i.e. lateral inhibition/excitation), the network must identify which syllables are to be stressed and what degree of stress is to be realized. While the identification issue is formally similar to that found in syllabification systems, fewer phenomena are straightforwardly amenable to an identification procedure that simply identifies local maxima. Stress systems that are not stress clash avoidant permit adjacent syllables to be stressed. In such situations, it is not possible for both syllables to be a local maximum. To accommodate such phenomena, stresses are defined as those

---

9The juxtaposition of two consecutive stress peaks is formally similar to the situation where two consecutive vowels both serve as syllable peaks (cf. 3.4.3).
units that exceed a particular threshold (\(\Theta\)).\(^{10}\) Thresholds can also be used to identify variable degrees of stress. While the identification stress peaks may prove to be more complex than the discovery of syllable peaks, the computation of relative values for non-heads is less constrained. Whereas syllable tactics (e.g. the Sonority Sequencing Principle) generally require a precise sequencing of non-head, stress systems that identify only a single primary stress generally don't impose any similar constraint on non-stressed syllables.\(^{11}\)

To summarize, the following sources of variation are available within the network:

(2) Parameters of variation (modifiable weights)

**Positional activation:** Initial (I), Final (F)

**Weight activation:** Heavy (H), Light/Bias (L)

**Lateral/feedforward inhibition:** \(\alpha / \beta\)

**Stress assignment:** Local maximum, Global maximum, Threshold (\(\Theta\))

Of the above, the vast majority of the work is done by \(I, F, H, \alpha\) and \(\beta\). Moreover, it should be noted that no language will require more than a small subset of the above parameters for its description. In fact, it may be possible that all languages can be described without recourse to one or more of the parameters. For instance, the use of *bias* for quantity-insensitive languages (or for light syllables in quantity-sensitive languages) can be eliminated, though such a move may result in a more complex description of some systems. Throughout the analyses of representative stress systems that will follow, the use of the above parameters permits a careful evaluation of computational complexity, generative power, etc.

### 4.2.0 Alternative accounts

Within metrical phonology, a significant division has arisen between analyses that rely on trees (Hayes 1980; Halle & Clements 1983; Hammond 1988) and those that rely on grids (Liberman 1979; Liberman & Prince 1977, Prince 1983). More than simply representing a theoretical choice between competing formalisms, the two approaches potentially reflect the *particle* and *wave* metaphors described in chapter one. In the former, stress is a property of syntactically defined constituents, whether it be heads in an

---

\(^{10}\)While the introduction of threshold units may appear to increase the power of the system, they are independently necessary in the syllable network for well-formedness judgments (i.e. local maxima which fail to exceed the threshold are ill-formed). The computations necessary to identify thresholds are, in fact, less costly than those that identify local maxima.

\(^{11}\)Languages that permit more than two degrees of stress (e.g. English) do pose difficult stress assignment challenges for both traditional metrical theories as well as the computational network. While it might be hoped that the relative derived activation level of each stress peak should correspond to its realized degree of stress in either a continuous or step-function fashion, such an analysis falls outside of the scope of this discussion.
X-bar graph or strong elements in binary trees. In the latter, stress is a computational property of a column of asterisks which can be added, subtracted and moved. Halle & Vergnaud (1987) take a moderating view that represents metrical structure through the use of grids but constructs and manipulates those grids by means of metrical constituents. The metrical network described above also accommodates both descriptive viewpoints but in the mirror image of Halle & Vergnaud's approach. In the network model, the stress coefficients represented by the column of asterisks in a metrical grid are directly computed with any necessary constituent structure subsequently read off of the resulting periodic waves.

In order to be able to be to compare and contrast the alternative approaches in terms of explanatory adequacy we will assume that the analyses offered by Halle & Vergnaud (1987) and by Dresher & Kaye (1991; cf. Dresher 1992) successfully model the phenomena they discuss. After discussing Halle & Vergnaud’s model in some detail, we will briefly describe Dresher & Kaye’s principles and parameters model since its parameters can be more directly compared with the variable coefficients that are used within the proposed metrical network. In addition to demonstrating that the network can successfully process the same range of data, we will examine the machinery necessary to accommodate the various stress systems.

4.2.1 Halle & Vergnaud -- An Essay on Stress (1987)

While Halle & Vergnaud profess to search for a middle road, their theoretical approach proves, in fact, to be more closely aligned with constituency theories, attested by the fact that the book begins with Section 1.1 "Metrical Constituents." For Halle & Vergnaud the principal task is to manipulate constituent structures so that grids can be constructed from constituent heads. Rather than directly manipulate the metrical grid once it is constructed, Halle & Vergnaud typically choose to modify the constituent structures that control the building of the grid. Even when rules appear to do the former, they are usually supported by elaborate procedures that modify constituents. For instance, in their discussion of Klamath, Halle & Vergnaud posit a rule (49h) which modifies the metrical grid: "Delete line 0 asterisk on the penultimate syllable if its rime is non-branching . . ." Recognizing that such a rule potentially serves as an embarrassment within their formalism, Halle & Vergnaud add a footnote which argues,

The context for the deletion rule (49h) will be characterized by means of a mechanism similar to that introduced for Tiberian Hebrew in section 2.3.3. Specifically, we would posit an independent reduction plane in which all syllables with a branching rime are accented and in which bounded left-headed constituents are constructed from right to left. A line 0 asterisk on the stress plane is deleted whenever it corresponds to a governed position on the reduction plane (p. 74, fn. 13).

While the placement or movement of asterisks is generally accounted for by constituency-sensitive operations, this is not exclusively the case. Rules such as (3) provide an
extremely powerful (and insufficiently constrained) procedure for directly modifying the grid.

(3) Boundary metathesis: *) → *) / * ______ line 1 (word finally)
    *  *  * line 0

While the hybrid use of constituent structures and grids significantly increases the complexity of their account, Halle & Vergnaud choose to build their grids from constituents for several reasons. The constituent structure formalism makes metrical representations congruent with syntactic structures on one hand and, more generally, with logical structures that can be manipulated by declarative programming languages such as Prolog (cf. Halle & Vergnaud, pp. 139-141). More importantly, it permits a explicit formal description of the machinery of the theory. Halle & Vergnaud's formalism involves three major types of theoretical machinery: parameters, principles (conditions, conventions, axioms, etc.), and rules. The challenge is to propose sufficiently powerful parameters and rules to account for the range of metrical phenomena while maintaining sufficiently strong constraints to limit the power of the formalism to describe only possible metrical systems. While proponents of the "principles and parameters" approach to grammar generally prefer to account for all phenomena without recourse to derivational rules (cf. Dresher & Kaye 1990; Dresher 1991), Halle & Vergnaud use a variety of constituency-sensitive and grid-sensitive rules to account for metrical phenomena.12

In Halle & Vergnaud (1987), four major parameters control the construction of metrical constituents. Initially, constituents can be either bounded or unbounded and either head-terminal or not head-terminal.

(4) a. [+BND, +HT] (+HT constituents illustrated as left-headed)
   *
   ( *  *  * )

b. [+BND, -HT] (amphibrachs)
   *
   ( *  *  *  * )

c. [-BND, +HT]
   *
   ( *  *  *  *  *  *  * )

12I t should be noted that Dresher & Kaye admit that the set of parameters that they identify do not account for all of the types of stress systems reported in Hall & Vergnaud (1987). They hold out hope, however, that additional parameters could be identified that would allow all systems to be described.
d. [-BND, -HT]  
(not exemplified - violates Recoverability Condition)

\[
\begin{array}{cccccc}
* & * & * & * & * & *
\end{array}
\]

While setting the head-terminal and boundedness parameters logically entails four possibilities, the negative setting for the head-terminal parameter is marginal, being excluded in the case of unbounded constituents and exemplified in bounded constituents only if one constructs ternary feet. Once values are set for \( \pm \text{BND} \) and \( \pm \text{HT} \), certain configurations permit the setting of additional parameters. In the case of binary constituents \((+\text{BND}, +\text{HT})\), it is possible to set the direction of headedness \((L \text{ or } R)\) and the direction of constituent construction \((\text{Left-to-Right or Right-to-Left})\). In the case of unbounded constituents, the directionality of constituent construction is moot but a choice can be made between left-headed and right-headed constituents. This leaves seven of the sixteen logically possible settings:

(5) a. \([+\text{BND}] [+\text{HT}] \text{ Left-headed constructed Left-to-Right}\)

\[
\begin{array}{cccccc}
* & * & * & * & * & *
\end{array}
\]

\[
\begin{array}{cccccc}
* & * & * & * & * & *
\end{array}
\]

b. \([+\text{BND}] [+\text{HT}] \text{ Right-headed constructed Left-to-Right}\)

\[
\begin{array}{cccccc}
* & * & * & * & * & *
\end{array}
\]

\[
\begin{array}{cccccc}
* & * & * & * & * & *
\end{array}
\]

c. \([+\text{BND}] [+\text{HT}] \text{ Left-headed constructed Right-to-Left}\)

\[
\begin{array}{cccccc}
* & * & * & * & * & *
\end{array}
\]

\[
\begin{array}{cccccc}
* & * & * & * & * & *
\end{array}
\]

d. \([+\text{BND}] [+\text{HT}] \text{ Right-headed constructed Right-to-Left}\)

\[
\begin{array}{cccccc}
* & * & * & * & * & *
\end{array}
\]

\[
\begin{array}{cccccc}
* & * & * & * & * & *
\end{array}
\]

e. \([+\text{BND}] [-\text{HT}]\)

\[
\begin{array}{cccccc}
* & * & * & *
\end{array}
\]

\[
\begin{array}{cccccc}
* & * & * & *
\end{array}
\]

\[
\begin{array}{cccccc}
* & * & * & *
\end{array}
\]

The fact that -HT, +BND is only consistent with amphibrachs is used by Halle & Vergnaud to justify excluding a parameter that chooses between binary and ternary constituents. As a consequence, dactyls and anapests are excluded.
f. $[-\text{BND}] [+\text{HT}]$ Left-headed

\[
(* \ast \ast \ast \ast \ast \ast \ast \ast )
\]

g. $[-\text{BND}] [+\text{HT}]$ Right-headed

\[
(* \ast \ast \ast \ast \ast \ast \ast \ast )
\]

Since many more stress systems need to be exemplified, additional parameters and/or rules are necessary to account for additional systems. For instance, given the options found in (5a-g), a language with a single stress could only place that stress on the first or the last syllable. To place such a stress in non-peripheral positions, Halle & Vergnaud offer two additional mechanisms. First, a single syllable at the beginning or end of the string can be marked as extrametrical. As a result, a language with a single penultimate stress could be analyzed as having $[-\text{BND}] [+\text{HT}]$ right-headed constituents with final extrametricality. The second alternative is slightly more complex. Noting that $[+\text{BND}] [+\text{HT}]$ left-headed constituents constructed from right-to-left results on stress on the penult and not the ultima, Halle & Vergnaud offer a procedure that will erase the overgenerated stresses. It is first necessary to construct constituents over the line 1 asterisks. If the line 1 parameters are now set at $[-\text{BND}] [+\text{HT}]$ $[\text{Right-headed}]$ we end up with the single asterisk over the penult syllable on line 2 just as the extrametricality solution has a single asterisk over the penult on line 1. Halle & Vergnaud then introduce a Line Conflation rule that eliminates the overgenerated stresses by erasing the line that was used to generate the line 2 asterisk in the first place.

On the other hand, the fact that only seven of the combinations of the above parameters yield possible metrical systems suggests that conditions need to be identified to constrain parametric combinations. Halle & Vergnaud offer the following conditions to attempt to formally constrain the parameter settings and types of rules that can apply:

**Recoverability Condition:** Given the direction of government of the constituent heads in the grammar, the location of the metrical constituent boundaries must be unambiguously recoverable from the location of the heads, and conversely the location of the heads must be recoverable from that of the boundaries.

**Exhaustivity Condition:** The rules of constituent boundary construction apply exhaustively.

**Maximality Condition:** Each constituent constructed by a rule of boundary construction must incorporate the maximal substring, provided that other requirements on constituent structure are satisfied.
**Faithfulness Condition:** The output metrical structure respects the distribution of heads (accented elements), in the sense that each head is associated with constituent boundaries in the output structure and that these are located at the appropriate positions in the sequence. Constituent boundaries are erased in the output when none of the elements enclosed by the boundaries is marked as head.

While the parameters and conditions are designed to formally constrain the theory, they reveal some potentially surprising gaps. In the machinery of the theory, there is no formal difference between the assignment of line 0 parameters and the assignment of line 1 parameters. In point of fact, however, all of the 25 languages for which parameters are specifically identified are posited as being \(+HT -BND\) on line 1.\(^{14}\) In fact, with very few exceptions, they are \(+HT -BND R\). Of the exceptions, Koya and Lithuanian use redundant parameters on line 0 and line 1, potentially making the latter unnecessary to capture the correct stress patterns. Only Western Aranda and Chamorro in the non-cyclic stratum uniquely exemplify \(+HT -BND L\) on line 1. The fact that all of the languages under discussion exemplify \(+BND\) on line 1 proves to be very significant. If languages can be

\(^{14}\)Halle & Vernaud's parameter settings

<table>
<thead>
<tr>
<th>Language</th>
<th>HT</th>
<th>BND</th>
<th>L/R</th>
<th>HT</th>
<th>BND</th>
<th>L/R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Koya</td>
<td>+HT</td>
<td>-BND</td>
<td>L</td>
<td>+HT</td>
<td>-BND</td>
<td>L</td>
</tr>
<tr>
<td>Weri</td>
<td>+HT</td>
<td>+BND</td>
<td></td>
<td>+HT</td>
<td>+BND</td>
<td></td>
</tr>
<tr>
<td>Maranungku</td>
<td>+HT</td>
<td>+BND</td>
<td></td>
<td>+HT</td>
<td>-BND</td>
<td>R</td>
</tr>
<tr>
<td>Yidin(^{3})</td>
<td>+HT</td>
<td>+BND</td>
<td></td>
<td>+HT</td>
<td>+BND</td>
<td></td>
</tr>
<tr>
<td>Cayuva</td>
<td>-HT</td>
<td>+BND</td>
<td></td>
<td>+HT</td>
<td>+BND</td>
<td></td>
</tr>
<tr>
<td>Akan</td>
<td>+HT</td>
<td>+BND</td>
<td></td>
<td>+HT</td>
<td>+BND</td>
<td></td>
</tr>
<tr>
<td>W. Aranda</td>
<td>+HT</td>
<td>+BND</td>
<td>L/R</td>
<td>+HT</td>
<td>-BND</td>
<td>L</td>
</tr>
<tr>
<td>Cl. Arabic</td>
<td>+HT</td>
<td>-BND</td>
<td>L</td>
<td>+HT</td>
<td>-BND</td>
<td>R</td>
</tr>
<tr>
<td>E. Cheremis</td>
<td>+HT</td>
<td>-BND</td>
<td>L</td>
<td>+HT</td>
<td>-BND</td>
<td>R</td>
</tr>
<tr>
<td>Macedonian</td>
<td>+HT</td>
<td>+BND</td>
<td>L/R</td>
<td>+HT</td>
<td>-BND</td>
<td>R</td>
</tr>
<tr>
<td>Chuvash</td>
<td>+HT</td>
<td>-BND</td>
<td>L</td>
<td>+HT</td>
<td>-BND</td>
<td>R</td>
</tr>
<tr>
<td>Hindi</td>
<td>+HT</td>
<td>-BND</td>
<td>L</td>
<td>+HT</td>
<td>-BND</td>
<td>R</td>
</tr>
<tr>
<td>Huasteco</td>
<td>+HT</td>
<td>-BND</td>
<td>L</td>
<td>+HT</td>
<td>-BND</td>
<td>R</td>
</tr>
<tr>
<td>Creek</td>
<td>+HT</td>
<td>+BND</td>
<td>R/L</td>
<td>+HT</td>
<td>-BND</td>
<td>R</td>
</tr>
<tr>
<td>Cairene Arabic</td>
<td>+HT</td>
<td>+BND</td>
<td>L/R</td>
<td>+HT</td>
<td>-BND</td>
<td>R</td>
</tr>
<tr>
<td>(reduction)</td>
<td>+HT</td>
<td>+BND</td>
<td>R/L</td>
<td>+HT</td>
<td>-BND</td>
<td>R</td>
</tr>
<tr>
<td>Klamath</td>
<td>+HT</td>
<td>-BND</td>
<td>L/R</td>
<td>+HT</td>
<td>-BND</td>
<td>R</td>
</tr>
<tr>
<td>Spanish (c)</td>
<td>+HT</td>
<td>+BND</td>
<td>L/R</td>
<td>+HT</td>
<td>-BND</td>
<td>R</td>
</tr>
<tr>
<td>(nc)</td>
<td>+HT</td>
<td>+BND</td>
<td>R/L</td>
<td>+HT</td>
<td>-BND</td>
<td>R</td>
</tr>
<tr>
<td>Seneca</td>
<td>+HT</td>
<td>-BND</td>
<td>L</td>
<td>+HT</td>
<td>-BND</td>
<td>R</td>
</tr>
<tr>
<td>English</td>
<td>+HT</td>
<td>+BND</td>
<td>L/R</td>
<td>+HT</td>
<td>-BND</td>
<td>R</td>
</tr>
<tr>
<td>Odawa</td>
<td>+HT</td>
<td>+BND</td>
<td>L/R</td>
<td>+HT</td>
<td>-BND</td>
<td>R</td>
</tr>
<tr>
<td>Lithuanian</td>
<td>+HT</td>
<td>-BND</td>
<td>L</td>
<td>+HT</td>
<td>-BND</td>
<td>L</td>
</tr>
<tr>
<td>Chamorro</td>
<td>+HT</td>
<td>+BND</td>
<td>L/R</td>
<td>+HT</td>
<td>-BND</td>
<td>R</td>
</tr>
<tr>
<td>(nc)</td>
<td>+HT</td>
<td>+BND</td>
<td>L/R</td>
<td>+HT</td>
<td>-BND</td>
<td>L (p. 215)</td>
</tr>
<tr>
<td>Lenakel (n)</td>
<td>+HT</td>
<td>+BND</td>
<td>L/R</td>
<td>+HT</td>
<td>-BND</td>
<td>R</td>
</tr>
<tr>
<td>(v)</td>
<td>+HT</td>
<td>+BND</td>
<td>L/R</td>
<td>+HT</td>
<td>-BND</td>
<td>R</td>
</tr>
<tr>
<td>Pirahā</td>
<td>+HT</td>
<td>+BND</td>
<td>LR* L/R</td>
<td>+HT</td>
<td>-BND</td>
<td>R (p. 225 sonority)</td>
</tr>
</tbody>
</table>
+BND on line 1 and above, an unlimited number of impossible stress patterns can be generated.

(6) Bounded Line 1 constituents

\[
\begin{align*}
&\text{Line 3} \\
&\text{Line 2} +HT -BND R \\
&\text{Line 1} +HT +BND L R/L \\
&\text{Line 0} +HT +BND L R/L \\
\end{align*}
\]

In the above example, the single stress in the word would appear on the fourth from the last syllable. Inappropriate settings of the parameters and/or the introduction of new lines with unlimited reapplication of line conflation could result in a single stress being placed on any arbitrary syllable within a word. So if excessive generative power is a reason to reject a grammatical formalism, Halle & Vergnaud's model would suffer the same fate as "less-constrained" alternatives (cf. 6.5).

4.2.2 Dresher & Kaye (1990) -- Principles and Parameters

 Whereas Halle & Vergnaud utilize principles, parameters and rules to account for the range of stress systems that appear in the world's languages, Dresher & Kaye (1990) attempt to account for the data without recourse to rules. For each of the eleven parameters that are listed, we indicated the coefficient within the dynamic computational network that achieves the same effect.

(7) Parameters for Metrical Structures (Dresher & Kaye 1990)

P1: The word-tree is strong on the [Left/Right] \((\alpha/\beta)\)
P2: Feet are [Binary/Unbounded] \((\alpha/\beta)\)
P3: Feet are built from the [Left/Right] \((I/F \ \alpha/\beta)\)
P4: Feet are strong on the [Left/Right] \((I/F \ \alpha/\beta)\)
P5: Feet are quantity sensitive (QS) [Yes/No] \((H)\)
P6: Feet are QS to the [Rime/Nucleus] \((\text{syllable network})\)
P7: A strong branch of a foot must itself branch [No/Yes] \((\alpha/\beta)\)
P8A: There is an extremetrical syllable [No/Yes] \((I/F)\)
P8: It is extremetrical on the [Left/Right] \((I/F)\)
P9: A weak foot is defooted in clash [No/Yes] \((\alpha/\beta)\)
P10: Feet are noniterative [No/Yes] \((\alpha/\beta)\)

In comparing the power of the network architecture to that provided by either Halle & Vergnaud's constituency/grid formalism or Dresher & Kaye's parameters, several different conclusions are possible, depending on how one evaluates the power of certain descriptive devices. Dresher & Kaye's formalism certainly involves more parameters. As
described, it utilizes 11 different parameters, though they admit that additional parameters would be required to accommodate all of the stress systems discussed in Halle & Vergnaud. In order to reduce their solution space, Dresher & Kaye also outlaw certain impossible configurations. Moreover, in designing a learning algorithm, they identify specific cues for the setting of each parameter, they have an extrinsically ordered identification algorithm, and they use batch-mode processing that makes the entire data corpus simultaneously available for analysis. Finally, Dresher & Kaye do not make it at all clear how their device could function effectively as or with a separately defined production system. While it may well be appropriate on methodological grounds to separate competence and performance issues, it still seems that a model that simultaneously accounts for competence and performance would be theoretical superior.

4.3 Quantity-insensitive stress systems

In the case of quantity-insensitive stress systems, stress proves to be a function of the position of a syllable within the string, independent of its phonological content. Given the parameters suggested by Dresher & Kaye, we would expect the following basic stress systems:

- Word-initial stress (alternating secondary stress): *Maramungku* (4.3.1)
- Word-final stress (alternating secondary stress): *Weri* (4.3.2)
- Second syllable stressed (alternating secondary stress): *Paite* (4.3.3)
- Penultimate stress (alternating secondary stress): *Warao* (4.3.3)
- Word-initial stress (no secondary stresses): *Latvian* (4.3.4)
- Word-final stress (no secondary stresses): *French* (4.3.4)
- Second syllable stressed (no secondary stresses): *Lakota* (4.3.4)
- Penultimate stress (no secondary stresses): *Polish* (4.3.4)

In addition to these basic stress patterns we will consider three significantly more complex stress patterns that demonstrate the advantages of the dynamic computational network. In Garawa (4.3.5), stress is both initial and penultimate, with alternating stresses propagating forward towards the beginning of the word, except that the second syllable is unstressed. In Lakota (4.3.6), different stress patterns are exemplified in nouns and verbs. Finally, we will consider the problem of antepenultimate stress. While no language appears to assign stress solely to antepenultimate syllables, languages such as Macedonian (4.3.7) can assign a stress to one of the last three syllables but no earlier syllables.

In the discussion of network analyses of typical stress patterns, we will frequently identify alternative solutions. In very few cases are the entire set of available parameters needed to generate the appropriate stress pattern. The examination of alternatives will

---

15Tourzetzky & Gupta (1991) also note stress systems that can not be captured by Dresher & Kaye's parameters (e.g. Garawa, Aklan).

16An earlier version of the current discussion is reported in Goldsmith (in press). In that paper, originally presented at the University of Illinois, Weri, Maramungku, Warao, Garawa and Lenakel are discussed in some detail.
illustrate the descriptive power of the network and permit the selection of superior analyses based on appropriate evaluative criteria. While the network has been trained on both arbitrarily long and arbitrarily short syllables, the illustrations will demonstrate the stress pattern in seven-syllable words (in cases with a crucial distinction between words of even and odd parity, six-syllable words will also be illustrated).

4.3.1 Maranungku

In Maranungku, stress falls on the first syllable and every following odd-numbered syllable. Halle & Vergnaud (1987) achieve the appropriate stress pattern by setting the parameters as follows:

(8) Maranungku: [+HT] [+BND] L L/R

The metrical network achieves the correct result by setting Initial (I) greater than zero and $\beta < 0$. As a result, the pulse associated with the initial segment is propagated down the string by lateral inhibition.

![Maranungku](image)

**Figure 4.9**

It should be noted that the above solution is very fault tolerant. The solution only refers to the parameters $I$ and $\beta$. More importantly, a wide range of values achieves the appropriate result ($I > 0$, $\beta < 0$). If bias is introduced, a positive value of $L$ will also generate the correct stress pattern whether or not $I$ is also used.
The choice between $I > 0$ and $L > 0$ is not entirely arbitrary, however. The former solution is superior due to a greater degree of stability. Unlike Halle & Vergnaud's model where the directionality parameter must be set at either L/R or R/L for each line, the current framework permits both alpha and beta to take non-zero values. In most cases, one or the other will approach zero as a result of the training regimen. But in situations where both alpha and beta are non-zero, the smaller the number of active input units, the better behaved the network will be. For example, if beta is negative and alpha is positive, the network with a single initial activation will continue to provide the correct result (Figure 4.11), with the wave simply being damped more quickly. In a network with a positive bias, however, the presence of a positive alpha will make the right end of the tree non-rhythmic, resulting in an incorrect parse (Figure 4.12, note the star on the last syllable which should be stressed but isn't). In addition to the stability of the network, the use of a single initial activation simplifies the recognition task. In general, two recognition alternatives present themselves: the identification of local maxima and/or the use of linear threshold units. In most cases, the two alternatives prove equivalent. In Figure 4.9, for example, one could set a threshold for stress at 0.00 with all local peaks also exceeding the threshold.
In the case where alpha and beta are both negative, an interesting pattern emerges. When only an initial activation is posited, any negative value of alpha and beta will produce the correct stress pattern, regardless of which has the greater magnitude (Figure 4.13). The only limitation is that $\alpha + \beta \geq -1$. If not, the network will not be convergent.

While the network in Figure 4.14 appears to have the same property, it is only because of the fact that it has an odd number of syllables. If a string consists of an even number of syllables, the network becomes much more sensitive to the presence of a negative alpha. As alpha becomes increasingly negative, the activation from the final syllable sends a leftward pulse that collides with the rightward pulse. Depending on the relative magnitude of alpha and beta, a variety of different results will appear. When $\alpha = -.40$ and $\beta = -.60$ the right to left pulse results in the final syllable incorrectly receiving stress rather than the penult (Figure 4.15). When both alpha and beta are -.50, the resulting wave is symmetrical, with an oscillation between syllable three and syllable four receiving stress (Figure 4.16).
As a result, we propose the following parameters for the stress assignment system of Maranungku:

(17) Positional activation: \( I > 0; \ F = 0 \)
Weight activation: not applicable // \( L = 0 \)
Lateral inhibition: \( \beta < 0; \ \alpha = 0 \)

4.3.2 Weri

The stress pattern in Weri (Boxwell & Boxwell 1966) proves to be the mirror image of Maranungku, with stress falling on the final syllable and every odd-numbered syllable counting back from the final syllable. As a result, the metrical network can account for such a stress pattern with the following parameters:

(18) Positional activation: \( I = 0; \ F > 0 \)
Weight activation: not applicable // \( L = 0 \)
Lateral inhibition: \( \alpha < 0; \ \beta = 0 \)
As in the case of Maranungku, a wide range of values generates the correct stress pattern. Of most interest is the fact that with a positive bias ($L$) and a negative $\alpha$, the final parameter ($F$) does not need to be set (Figure 4.20). In the proposed analysis, the crucial difference between Maranungku and Weri is the setting of $I$ vs. $F$ with a less critical distinction between settings of $\alpha$ and $\beta$. In the bias solution, only $\alpha$ and $\beta$ differ.
While the latter solution might appear simpler, the previously discussed instability created by increasing the number of active units mitigates its value. As a result, we avoid the introduction of bias as a variable unless a strong warrant demands it.

The fact that disjoint alternative solutions are often available illustrates an interesting property of the network. In addition to evaluating the relative merits of the alternative solutions, we can entertain the possibility that two languages which have similar surface stress patterns could have different underlying parameters. The difference could potentially be reflected in different historical developments as the parameters change. While the two systems would start out as phenomenologically similar, they could diverge in rather surprising ways over time.

### 4.3.3 Paiute // Warao

Whereas Maranungku and Weri place primary stress on the peripheral syllables in the string, Paiute (Hayes 1980) and Warao (Osborne 1966) assign primary stress to the second and penultimate syllables, respectively. In each case, rhythmic stress is assigned to alternate syllables beginning with the primary stress. In traditional analyses, the stress pattern is achieved by marking the peripheral syllable as extrametrical and then creating binary constituents in the same way as Maranungku and Weri.

![Figure 4.21](Image)

![Figure 4.22](Image)

As in the case of Maranungku and Weri, the introduction of a negative bias ($L$) as a variable results in the same stress pattern as generated by the use of negative values for the initial ($I$) or final ($F$) parameters. In addition to the potential difficulties created by introducing additional active input units, the use of a negative bias complicates the description of what it means to be a stress peak.
4.3.4 Latvian/French/Lakota/Polish

Up to this point, we have discussed stress systems that exhibit perfect grid (iterative binary constituents in traditional metrical analyses). We now turn to a group of languages that only have a single stress per word:

- Initial: Latvian
- Final: French
- Second: Lakota
- Penultimate: Polish\(^{17}\)

Within metrical phonology, each of the above stress patterns can be described in terms of unbounded constituents with extrametricality employed in Lakota and Polish to exclude a peripheral syllable. Halle & Vergnaud (1987), however, propose an account of penultimate stress in which left-headed binary constituents are constructed from right to left. Since such an account would overgenerate secondary stress, the rather opaque line conflation rule would be required to erase the Line 1 asterisks.

In the metrical network, five different alternatives are available for the description of a single initial or final stress (Latvian/French). By far the simplest (and preferred) is an account where we only utilize the \( I \) parameter.

---

\(^{17}\)As noted by Halle & Vergnaud (1987) and Touretzky & Gupta (1990), describing Polish as a language with single penultimate stress is somewhat of an oversimplification. Its prototypical stress pattern illustrates the means by which a metrical network can model penultimate stress. The exceptional cases where either antepenultimate stress or multiple stresses appear will be discussed below (cf. Rubach & Booij 1985).
(25) **Latvian**
Positional activation: \( I > 0; \ F = 0 \)
Weight activation: not applicable \( /L = 0 \)
Lateral inhibition: \( \alpha = 0; \ \beta = 0 \)

**French:**
Positional activation: \( I = 0; \ F > 0 \)
Weight activation: not applicable \( /L = 0 \)
Lateral inhibition: \( \alpha = 0; \ \beta = 0 \)

![Figure 4.26](image1)

![Figure 4.27](image2)

In the above analyses, the stressed syllable has a positive value equal to \( I/F \) while all other syllables have the value 0. Since the recognition device can identify a local maximum, it is also possible to describe the stress patterns of Latvian and French with non-zero values of alpha and/or beta. In the case of initial stress (Latvian), if beta is greater than zero but less than one, each successive segment approaches zero leaving a single maximum at the left end of the string (Figure 4.28).

![Figure 4.28](image3)

![Figure 4.29](image4)
Another class of solutions is suggested by Halle & Vergnaud's *line conflation* rule. In several cases of single stress, Halle & Vergnaud propose the construction of binary constituents which will help place the stress in the appropriate position. The extra stresses which are *overgenerated* by the procedure are then erased by superimposing line 2 over line 1 (Halle & Vergnaud 1987, pp. 49-52). While the formal apparatus is different, the metrical network allows a similar operation to be modeled. A line conflation analysis of Latvian would build constituents identical to Maranungku and then eliminate the rhythmic stresses. If we adopt for Latvian the same parameters as Maranungku ($I > 0; -1 < \beta < 0$) and use a linear threshold unit ($\theta$) equal to $I$, the additional local maxima will be excluded as stress peaks (Figure 4.30).

![Figure 4.30](image1)

![Figure 4.31](image2)

Two final network configurations that generate a single word-initial stress deserve brief mention, particularly since they potentially accommodate some of the more complex phenomena that will be encountered later in the discussion. In Figure 4.32, we observe the effect of a positive bias ($L = 1$) with a positive $\alpha$ ($\alpha = .50$). The wave grows constantly from right to left with the leftmost syllable becoming the peak of activation. An even more radical alternative can generate a rising right-to-left wave with a single final activation ($F = 1$) if alpha is permitted to take a value greater than one (Figure 4.33).
In both of the illustrated cases, the choice of values for the parameters ($L, \alpha, \beta$) is significantly less error-tolerant than the proposed analysis illustrated in 4.26. More importantly, a greater potential exists for non-convergence in the network. As a result, the choice of parameters in (25) is to be preferred.

The modeling of a single stress in second position (Lakota) or penultimate position (Polish) is slightly more complex. Four different options are available, two of which appear quite plausible. In Figure 4.32, a positive bias with a positive alpha created a rising contour with an initial maximum. If this configuration is modified by simply positing a negative value for $L$, stress on the second syllable would result (Figure 4.34). In the mirror image system, a positive beta, positive bias and a negative $F$ would generate a single penultimate stress (Figure 4.35). In the context of constituency-sensitive theories, the assignment of negative values for $I$ or $F$ juxtaposed with a positive value for $L$ proves to be conceptually equivalent to the use of extametricality.
The assignment of a single stress on the second or penultimate syllable could also be achieved by introducing a threshold unit ($\theta$) to the description of Paiute (Figure 4.36 vs. Figure 4.21) or Warao (Figure 4.37 vs. Figure 4.22).

![Graphs of Lakota and Polish](image)

In such an analysis, the difference between Lakota and Paiute is that the latter assigns stress to all local peaks (or alternatively, all syllables with an activation greater than 0.00), while Lakota only assigns stress to syllables that equal a threshold determined by $B^*F$ (or alternatively, the global maximum, cf. Creek, 4.4.4). In the discussion of Maranungku above, it was suggested that a recognition device that assigned stress to local maxima would generally prove equivalent to one that used threshold units to assign stress. The case of Lakota and Polish suggest that the two choices would not always be equivalent, providing evidence that at least some syllabification systems are better served by the latter.

Before leaving the discussion of Lakota and Polish, it would be instructive to consider a potential solution that is decidedly less desirable. Prince (1992) criticizes the current network architecture on the grounds that it permits the description of impossible stress configurations. The particular configuration that he objects to is one with positive bias and positive values for both alpha and beta. He observes that such a network could arbitrarily assign stress to any syllable in the string depending on the alpha/beta ratio. If such a concern were not heeded we might propose the following parameters:

(38) **Lakota**

- Positional activation: *not applicable*
- Weight activation: $L > 0$
- Lateral inhibition: $\alpha = .50$, $\beta = .10$

**Polish**

- Positional activation: *not applicable*
- Weight activation: $L > 0$
- Lateral inhibition: $\alpha = .10$, $\beta = .50$
Unfortunately, such a solution would lead to precisely the difficulties that Prince notes. Slight changes in either alpha and beta would move the stress successively to each position in the string as Figures 4.41-44 illustrate.