Constrained Emergence of Universals and Variation in Syllable Systems *

Melissa A. Redford, Chun Chi Chen, Risto Miikkulainen
The University of Texas, Austin, TX 78712

Abstract

A computational model of emergent syllable systems is developed based on a set of functional constraints on syllable systems and the assumption that language structure emerges through cumulative change over time. The constraints were derived from general communicative factors as well as from the phonetic principles of perceptual distinctiveness and articulatory ease. Through evolutionary optimization, the model generated mock vocabularies optimized for the given constraints. Several simulations were run to understand how these constraints might define the emergence of universals and variation in complex sound systems. The predictions were that (1) CV syllables would be highly frequent in all vocabularies evolved under the constraints; (2) syllables with consonant clusters, consonant codas and vowel onsets would occur much less frequently; (3) a relationship would exist between the number of syllable types in a vocabulary and the average word length in the vocabulary; (4) different syllable types would emerge according to, what we termed, an *iterative principle of syllable structure* and their frequency would be directly related to their complexity; and (5) categorical differences would emerge between vocabularies evolved under the same constraints. Simulation results confirmed these predictions and provided novel insights into why regularities and differences may occur across languages. Specifically, the model suggested that both language universals and variation are consistent with a set of functional constraints that are fixed relative to one another. Language universals reflect underlying constraints on the system and language variation represents the many different and equally-good solutions to the unique problem defined by these constraints.

INTRODUCTION

Computational models have been used in recent years to demonstrate how language structure could emerge given competing functional constraints on the system. Although much of this work has focused on how semantic and syntactic patterns may emerge (e.g., Steels, 1998; Kirby, 1999; Nowak and Krakaurer, 1999), some work has also considered the emergence of phonological patterns, specifically the emergence of frequent vowel systems and CV syllables (Joanisse and Seidenberg, 1997; Joanisse, 1999; de Boer, 2000). Computational models of emergent phonology grew out of a tradition in phonetics that explains universal sound patterns in terms of perceptual distinctiveness and articulatory ease constraints (this tradition is most recently represented by Ohala, 1983; Lindblom, MacNeilage, Studdert-Kennedy, 1984; Lindblom, 2000). Although these general constraints interact, individual perceptual or articulatory factors are isolated in experimental studies to explain individual sound patterns (e.g., Lindblom, 1983; Ohala, 1990; Davis and MacNeilage, 1995; Redford and Diehl, 1999). The computational models also tend to explain sound patterns in terms of a single constraint. For instance, vowel systems are shown to emerge when perceptual distance between vowels is maximized (de Boer 2000; Joanisse, 1997) and CV syllables are preferred in systems over VC syl-

* *Language and Speech* (in press, January 2001)

lables because they are more perceptible (Joanisse, 1999). The question of how specific phonetic constraints might work together to explain complex cross-language sound patterns, such as those that appear in syllable systems, has yet to be explored in detail. In addition, since specific phonetic constraints explain similarities between sound systems, a question arises as to how these same constraints might explain differences between sound systems as well.

The goal of the present paper is to examine both how multiple constraints interact to produce universal patterns and how the resulting systems may differ. A computational model of emergent syllable systems is developed based on a set of functional constraints and the assumption that structure emerges through cumulative change over time. Given a set of constraints and an initial set of randomly-generated vocabularies, the model modifies the vocabularies using artificial evolution, in effect optimizing them under the constraints. The specific constraints are proposed in the following section and their expected effect on syllable systems is discussed. The hypothesis on cross-language variation is then developed in more detail.

Emergence of Universals

Our model, which we refer to as the Emergent Syllable Systems model, or ESS, is built around six constraints that operate on sound systems. The constraints are derived from the communicative function of language. Some of the constraints affect the structure of the vocabulary as a whole, while others are more specific and constrain the organization of segments within a word. The interaction between the constraints is assumed to give rise to the cross-language universals identified in syllable systems.

Vocabulary-Level Constraints. A syllable system describes the different types of consonant and vowel groupings that occur in words of a particular language. These groupings are due to constraints that favor or disfavor different combinations of sounds. While some constraints operate on the words themselves, others operate at the level of the vocabulary as a whole. Three vocabulary-level constraints are proposed in ESS.

The first constraint follows from the principle of least effort. Assuming that all utterances consume energy, producing longer words will consume more energy than producing shorter words. Assuming that natural systems try to conserve energy, emerging vocabularies should, as a whole, favor shorter words over longer words. A word length constraint of this sort is consistent with the cross-language pattern established by Zipf (1935/1968) that short words occur more often than long words in language. The principle of least effort has been extended by phoneticians to explain why high-frequency words have a simpler sound structure (Willerman, 1994) and why articulatorily simple phonemes predominate over those with secondary and tertiary articulations in phoneme inventories (Lindblom and Maddieson, 1988).

A second constraint is that no two words should be identical. This constraint follows directly from the communicative function of language since different words must be used to denote different objects, actions, and concepts. While there are exceptions to this constraint (i.e., homophones in the vocabularies of established languages), it is unlikely that vocabularies could be used to adequately communicate different ideas if these were expressed with the same acoustic patterns. If there is a constraint on identical words, then words that are added to a vocabulary will consist of differently-combined sounds. Different combinations may result either in the creation of new syllable types or in the recombination of existing syllable types. Which of these occurs depends upon the action of a final vocabulary-level constraint.

The third constraint favors the reuse of syllable types within a vocabulary. This constraint follows from the observation that although many different consonant and vowel groupings occur across languages, the words of a single language use a very small subset of these. An analogous observation has been made on distinctive feature use in phoneme inventories: Although many different distinctive feature combinations occur across languages, individual phoneme inventories systematically exploit only a few of these (Lindblom, 1998). Lindblom (2000) argues that the reuse of sound patterns within a language reduces the metabolic cost of forming a new memory trace. If these costs are reduced, new speakers will acquire the language more easily.

In summary, three vocabulary-level constraints are proposed in ESS. These constraints favor vocabularies with short, distinctive words and the reuse of syllable types in constructing these words.

Word-Level Constraints. Even though the proposed vocabulary-level constraints will ensure overall diversity and structure in the vocabulary, word-internal structure is still free to vary under them. As previously noted, syllable systems describe different types of consonant and vowel groupings, but not every type of grouping occurs in language. According to Bell and Hooper (1978) there are three prominent patterns: (1) All languages allow initial consonants and final vowels, but not all languages allow initial vowels or final consonants; (2) all languages allow alternating sequences of consonants and vowels, but not all languages allow multiple consonant or vowel sequences; and (3) the preferred segment sequencing pattern across all languages is one where sonority, or loudness (Price, 1980), increases from the edges of a syllable to the center. A universal grouping, or syllable type, is implicit in these patterns: Consonant-vowel (CV) syllables occur in all languages.

The three cross-language patterns may be explained to emerge from three phonetic constraints on sound combination. The first two patterns, i.e., the preference for consonant onsets over consonant offsets and the preference for single consonants over consonant clusters, may both result from specific perceptual distinctiveness constraints. Initial consonants are more perceptible than final consonants in phrase-medial, monosyllabic nonsense words because they are produced with less variability and more accuracy (Redford and Diehl, 1999). In addition, in phrase-medial, monosyllabic nonsense words, single onset or offset consonants are more perceptible than consonant clusters (Redford, 1999). Single consonants are also usually produced more accurately than consonants in clusters (Byrd, 1996).

A mechanical constraint on segment articulation may explain the third pattern, i.e., why segments regularly increase and then decrease in sonority across a syllable. Segment articulation takes place within the regular open-close jaw cycle that characterizes speech (MacNeilage, 1998). In addition, segment sonority appears to be strongly related to jaw openness, such that more sonorous segments are produced with more jaw opening than less sonorous segments (Lindblom, 1983). The sequencing bias observed across languages may therefore reflect a jaw cycle constraint on segment articulation: sequential segments systematically increase and then decrease in sonority as the jaw moves from a closed position to a more open position and back again. This hypothesis was recently supported in a study of jaw movement during the production of different syllable types (Redford, 2000). The study also disconfirmed an alternative hypothesis, namely, that the jaw moves in service of segment articulation and according to a phonological principle of sonority sequencing.

Constraint Interaction. The three vocabulary-level and the three word-level constraints are proposed to account for the types of consonant and vowel groupings that occur in language. The interaction of these constraints is expected to shape syllable structure so that it reflects the cross-language regularities described above. Specifically, the vocabulary-level constraints on word length and meaning are expected to act together to maintain diverse sound structures. The three specific phonetic constraints on segment position and organization are expected to act together to select for vocabularies with high proportions of CV syllables. Individually, the phonetic constraints are expected to select against specific syllable types, namely, syllables with vowel onsets and consonant offsets and syllables with adjacent consonants or vowels. By restricting the number of syllable types, the vocabulary-level constraint on syllables will increase the frequency of the best types defined by the phonetic constraints.

The interaction of the ESS constraints also makes novel predictions about general syllable patterns that could be independently verified with language data. One prediction follows from the expectation that the general constraints on word length and meaning will generate vocabularies consisting of different, short words. Distinctiveness requires that words consist of more than one segment, but if segments may be freely recombined, then distinctiveness and brevity can be achieved by exploiting all possible combinations of a few segments. Since different segment combinations represent different syllable types, it is predicted that the average word length in a vocabulary will be inversely related to the number of syllable types. This prediction is analogous to one proposed by Nettle (1999). Nettle argued that languages with larger phoneme inventories have shorter words because more phonemes allow for more different combinations of a few segments.

Another prediction is that the combined effect of all the constraints will determine the relative frequency of different syllable types. If vocabularies were only constrained by the phonetic pressures of articulatory ease and perceptual distinctiveness, all words would be constructed out of CV syllables. However, since vocabularies are also constrained to favor distinctive and short words, non-CV syllable types will also be present. The relative frequency of these non-CV syllable types can be predicted from the interaction of the specific phonetic constraints. For instance, the next most frequent non-CV syllable types in the system should be one degree removed from the basic CV type. A consonant or vowel may be added or deleted to form the following syllable types: CCV, CVC, CVV, V, and C. Additional, less and less frequent syllable types should deviate further and further from the basic CV syllable in an iterative fashion and according to the phonetic constraints. For example, the constraint against identical adjacent segments combined with the constraint on segment position should result in a system where syllable onsets increase in complexity before codas, then onsets again, and so on. This process can be expressed as the *iterative principle of syllable structure*.

The prediction of an iterative principle of syllable structure is consistent with other hypotheses that invoke competing functional constraints, or competing motivations, to explain asymmetric relationships between linguistic types (e.g., Kirby, 1999). Here we predict asymmetries in syllable type frequency and we explain that they emerge from the interaction of a few phonetic constraints with some more general functional constraints.

In addition to predicting the relative frequency of syllable types, the iterative principle makes an implicational prediction. Specifically, the occurrence of a more complex syllable type implies the more frequent occurrence of the less complex type. For example, if syllables with final clusters occur in a vocabulary, then the vocabulary should also have syllables with single final consonants that occur with higher frequency.

These predictions regarding syllable system universals are tested in ESS and discussed with reference to real language data. The six universal constraints described in the previous section are formalized to describe a "fitness function", i.e., an optimization problem, under which syllable systems emerge. If the constraints are valid, the simulated emergent systems should replicate the patterns of consonant and vowel groupings observed across language, since identically constrained vocabularies should result in similarly structured syllable systems. More perplexing, though, is how variation between systems could also emerge in this process. In the next section, a hypothesis is developed for how variation might exist despite a single optimization problem.

Variation Across Syllable Systems

Although phonetic theory often addresses the question of how a particular sound pattern could have emerged over time, the question of how different sound patterns emerge given the same functional constraints has not been addressed. In contrast, language variation is a central issue in phonological theory, but it is almost always explained in synchronic terms. The most current and influential phonological explanation for language variation has been put forth in Optimality Theory (OT, Prince and Smolensky, 1993). In OT, universal phonological constraints are ordered according to their relative strength in each language. A stronger constraint takes effect over a weaker constraint and, depending on the structure of the language, the ordering of constraints may differ. For instance, a "no coda constraint" is stronger in a language like Hawaiian, where syllables never have final consonants, than in a language like English, where syllables often have final consonants. Although it is possible to conceive of a version of OT in which the constraints are differentially weighted rather than rank-ordered (Flemming, 1997), the constraint optimization process will still necessarily lead to different sound systems.

If the strength of individual universal constraints differs across languages, then each language must have emerged as the optimal solution to a unique problem defined by a particular ordering of the constraints. If this is correct, then language variation results from multiple optimization problems in the form of differently-ordered constraints. But what might account for the different orderings? According to Haspelmath (1999), different constraint orderings may result from social selection pressures. Speaker innovations could change the relative ordering of constraints. New ordering variants would exist in tandem with old and would be gradually selected for or against by the entire community of speakers. In other words, Haspelmath proposes that social variation and social selection could explain how the same set of universal constraints might be reordered to produce language variation.

Social pressures are undeniably important in language evolution (see Nettle, 1999), but they can also be invoked to support an alternative explanation for the emergence of language variation. An alternative hypothesis is that universal constraints are fixed relative to one another and therefore pose a single optimization problem. In the space of possible languages, there are many different, equally good solutions. According to this hypothesis, social pressures dictate whether a community remains in a single local optimum or moves within the search space to a different optimum. Certain social pressures, such as group identity and historical precedence, function as stabilizing forces, while other social pressures, such as economic or cultural imperialism, may provide sufficiently destabilizing forces to move the community out of a particular local optimum, i.e., language, and into a different one.

A hypothesis that assumes fixed constraints is preferable to one in which constraints are freely rearranged because it is simpler. This is because when universal constraints are assumed to be differently arranged by different language communities many further questions ensue. For instance, how do speakers and listeners exert control (albeit unconscious) over universal constraints, how can they come to agree on new constraint orderings, what are they giving up, and what are they gaining? These questions, which are difficult to answer, are avoided entirely when universal language constraints are assumed to be fixed relative to one another. Different languages can then emerge due to an accumulation of small, random changes over time in communities that are separated by either social or physical space. The changes experienced in a community are preserved and built upon so long as the constraints continue to be satisfied. In this way, language variation in any particular slice of time represents a set of different good solutions to a single optimization problem.

This hypothesis on how language variation emerges is indirectly tested in ESS. The architecture of the model reflects assumptions critical to the evolutionary process, namely, variation within a population and continuity in structure over time. In the following sections, the architecture of the model is explained in detail. In addition, the previously-discussed functional constraints are formalized, computational simulations are described, and specific syllable patterns are predicted to emerge. The simulated, emergent vocabularies are analyzed for similarities and differences in syllable structure in order to understand how the proposed constraints affect the system. The results are discussed with reference to the predictions on universals and variation. The results are also compared with patterns observed in real languages.

THE EMERGENT SYLLABLE SYSTEMS (ESS) MODEL

The syllable system of a language is determined from the organization of phonemes in the words of that language. Thus, in order to simulate the emergence of syllable systems, it is necessary to simulate the emergence of a vocabulary of words. ESS does just that.

Symbiotic Evolution

The ESS model is based on a version of genetic algorithms called Symbiotic Evolution (Moriarty and Miikkulainen, 1996). In this method, crossover and mutation are used to evolve a population of partial solutions that combine to yield an optimal solution to a given problem. This hierarchical architecture makes Symbiotic Evolution well suited to representing the hierarchical structure of language. In ESS, the partial solutions are words. In the first generation, the words represent random concatenations of phonemes. These phoneme strings have a maximum length, but words may be of any length under the maximum length. Subsets of words are randomly selected to form vocabularies. The vocabularies are then evaluated according to a set of specified functional constraints and assigned a fitness score. The fitness of a word depends upon how well it functions with other words to form good vocabularies. The fitnesses of all vocabularies in which a word participates are averaged to get a fitness for the word.

As a next step, the entire population of words is sorted according to word fitnesses and a percentage of the most fit words are interbred. Interbreeding takes place through crossover, where a portion of one word is concatenated to a portion of another word to form an offspring word. In addition, a fixed rate of mutation is applied. For each phoneme, there is a fixed (and small) probability that the phoneme will be replaced by another, randomly selected phoneme. The crossover and mutation operations yield novel words. These words are no longer randomly organized, since they contain parts of words that have been found to be highly fit, but diversity in structure is maintained because the crossover points are randomly specified and mutation is introduced. The novel word population replaces a percentage of the least fit words in the overall population of words. New vocabularies are then randomly selected from the revised word population, and the process continues. In this way, words that participate in successful vocabularies are more likely to reproduce and their inherent structure becomes highly represented in the population.

Constraints

The fitness measure in ESS is built from the previously-discussed functional constraints. The constraints are weighted by the experimenter. The weight values determine the penalties that vocabularies (and consequently their constituent words) receive during evaluation if they violate the constraints. The penalties are added together to determine the fitness for the vocabulary. Each of the constraints is formalized below.

1. Efficiency constraint on word production. This constraint penalizes vocabularies with longer words. The total number of segments in all the words of the vocabulary are counted and divided by the maximum number of segments allowed in our system for all words of a single vocabulary:

$$P_1 = \frac{W_1}{N_x N_p} \sum_x \sum_p 1, \tag{1}$$

where W_1 is the weight associated with this constraint, x is a word in the vocabulary, p is a phoneme in this word, N_x is the number of words in the vocabulary, and N_p is the maximum length of phoneme strings allowed in our system.

2. Intelligibility constraint on word meaning. This constraint penalizes vocabularies with identical words. Each word is compared to every other word in the vocabulary phoneme by phoneme. The penalty is assigned as

$$P_2 = W_2 \sum_x \sum_{y \neq x} r(x, y), \tag{2}$$

where W_2 is the weight associated with this constraint, x and y are words in the vocabulary, and

$$r(x,y) = \begin{cases} 1 & \text{if } \forall p \;\; x_p = y_p, \\ 0 & \text{otherwise}, \end{cases} \tag{3}$$

where x_p is phoneme p in word x.

3. Memory constraint on syllable types. This constraint penalizes vocabularies with a large number of different syllable types, which presumably requires more memory. The total number of syllable types is counted and multiplied by a constant:

$$P_3 = W_3 \sum_x \sum_a v(x,a), \tag{4}$$

where W_3 is the weight associated with this constraint, a is a syllable in word x, and

$$v(x,a) = \begin{cases} 1 & \text{if } S(x,a) \notin S_{xa}, \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

where $S(x,a)$ is the syllable a in word x and $\mathcal{S}_{xa}$ is the set of all syllables in words $y < x$ and syllables $b < a$ in word x (in other words, all syllables seen before a in the summation).

4. Perceptual distinctiveness constraint on segment position. This constraint penalizes consonants in word-final position and vowels in word-initial position:

$$P_4 = W_4 \sum_x (s(x) + t(x)) \tag{6}$$

where W_4 is the weight associated with this constraint, x is a word in the vocabulary, and

$$s(x) = \begin{cases} 1 & \text{if } \; x_1 \in \mathcal{V}, \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

and

$$t(x) = \begin{cases} 1 & \text{if } \; x_f \in \mathcal{C}, \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

where x_1 is the first phoneme, x_f the last phoneme in word x, $\mathcal{V}$ is the set of vowels and $\mathcal{C}$ the set of consonants.

5. Perceptual distinctiveness constraint on adjacent segments. This constraint penalizes identical, adjacent segments. Adjacent segments in each word are compared, and a penalty assigned if they are both consonants or both vowels:

$$P_5 = W_5 \sum_x \sum_p u(x,p), \tag{9}$$

where W_5 is the weight associated with this constraint, x is a word in the vocabulary, p is a phoneme in that word (ranging from 1 to $f-1$), and

$$u(x,p) = \begin{cases} 1 & \text{if } C(x_p) = C(x_{p+1}), \\ 0 & \text{otherwise,} \end{cases} \tag{10}$$

where $C(x_p)$ is the segment class (i.e., either vowel or consonant) of phoneme p in word x.

6. Articulatory constraint on segment sequencing. This constraint penalizes a fixed jaw and encourages a maximal difference in jaw openness between adjacent segments, which facilitates articulation.[1] The degree of jaw opening for the articulation of all segments was determined from previously published articulatory measurements (see the subsection on parameters and weights in "Simulations" below). Jaw position for adjacent segments within a word can therefore be compared and the differences calculated. Larger penalties are assigned for smaller differences in jaw position:

$$P_6 = \frac{W_6}{\sum_x \sqrt{\sum_p (O(x_p) - O(x_{p+1}))^2}}, \tag{11}$$

where W_6 is the weight associated with this constraint, x is a word in the vocabulary, p is a phoneme in the word (ranging from 1 to $f-1$), and $O(x_p)$ is the degree of jaw openness with which the phoneme p in word x is articulated.

SIMULATIONS

The specific goals of the simulations were (1) to verify that the interaction of the above functional constraints can explain the emergence of universals in syllable systems and (2) to determine whether sound system variation can result from a single optimization problem defined by these constraints. Similarities and differences between the emergent syllable

[1] All consonant segments are articulated during the closed portion of the jaw cycle and all vowels are articulated during the open portion (Redford, 1999). Thus, in a cycle that contains multiple consonants and vowels, multiple articulations must be achieved quickly as the jaw moves continuously to or from a closed position. In contrast, in a cycle with a single consonant and vowel, a single consonantal gesture, begun at the point of maximal closure, is continued through the opening phase and a single vocalic gesture, begun just before the point of maximal aperture, is continued through the closing gesture. Maximizing differences in jaw openness between adjacent segments therefore reduces the number of segments within a cycle and increases articulatory ease.

Table 1: The different instantiations of the model under which simulations were run. The six constraints are designated $C_1 - C_6$.

MODEL	C_1	C_2	C_3	C_4	C_5	C_6
Full model	on	on	on	on	on	on
1st partial	on	on	OFF	on	on	on
2nd partial	on	on	on	OFF	on	on
3rd partial	on	on	on	on	OFF	on
4th partial	on	on	on	on	on	OFF
Null model	on	on	OFF	OFF	OFF	OFF

systems were evaluated. If the proposed constraints are adequate, the simulated syllable systems should provide a good qualitative match with the data available on the world's languages. In addition, if language variation represents a range of different good solutions to a single optimization problem, then systematic differences should be identifiable in the resulting syllable systems.

Conditions

In order to test how the specific constraints, individually and in combination, affect an emergent syllable system, vocabularies were evolved under six separate conditions, i.e., six different instantiations of the model. One of these represented the full model where all six constraints were active and the other five were partial models. In other words, the full model was constrained to evolve vocabularies of short, different words (constraint 1 and 2), constructed out of a limited number of syllable types (constraint 3), with initial consonants and final vowels (constraint 4), and with segment sequences that alternated between consonants and vowels and conformed to the jaw cycle (constraints 5 and 6).

The partial models tested the effects of the constraints specific to syllables, 3–6. In four of the five partial models, each of these constraints was excluded in turn (Table 1). In the 1st partial model, vocabularies were no longer constrained to restrict the number of syllable types. In the 2nd partial model, vocabularies were no longer constrained to favor words with initial consonants and final vowels. In the 3rd partial model, alternate sequences of single consonants and vowels were no longer favored. In the 4th partial model, segment sequences did not have to conform to the jaw cycle. In the fifth partial model, all four of these syllable-specific constraints were excluded at the same time and the model was only constrained to evolve vocabularies of short, different words. This last model was called the null model since it provided a baseline against which the results from the other instantiations could be evaluated.

Parameters and Weights

The parameter settings and constraint weightings were the same for all models. Fifty vocabularies of 50 words each were created from a population of 800 words during each generation. Words could be a maximum of 12 phonemes long and were generated from the phoneme set /i, a, u, p, t, k, s, l, n/. This particular set was chosen because it represents the most frequent vowel and consonant sounds present in the world's languages (Maddieson, 1984). Phonemes were encoded according to the articulatorily-based distinctive features proposed by Chomsky and Halle (1968), i.e., according to place-of-articulation, manner-of-articulation, tongue height and tongue front-back dimension. Phonemes were also encoded as consonant or vowel and were associated with a jaw openness value. The jaw openness values were loosely based on measurement data from Lindblom (1983) and Keating, Lindblom, Lubker, and Kreiman (1994). The vowels /i, a, u/ were assigned openness values of 6, 9, and 6 (millimeters) respectively. The consonants /p, t, k, s, l, n/ were assigned openness values of 4, 3, 4, 2, 5, and 5 (millimeters) respectively.

The word population was created by randomly concatenating phonemes into strings. In subsequent generations, 100 of the best words were interbred using one-point crossover to form 100 novel words. The crossover point was randomly selected in each parent word. The 100 offspring replaced the 100 least fit words in the population. A fixed mutation rate of 1% was then applied to the revised word population.

Constraint weight (penalty) values were determined based on domain knowledge and experimentation with the model. Constraints were tested individually with different values and in conjunction with other constraints. The goal was to balance the weights so that each constraint clearly affected, but did not dominate, the optimization process. The values selected for the present set of simulations were as follows: (1) efficiency constraint on word production at -100; (2) intelligibility constraint on word meaning at -350; (3) memory constraint on syllable types at -500; (4) perceptual distinctiveness constraint on segment position at -25; (5) perceptual distinctiveness constraint on adjacent segments at -20; and (6) articulatory constraint on segment sequencing at -1000. The system turned out to be robust against minor variations of these values in repeated runs of the full model.

Twenty simulations were run in each condition. Each simulation began at a different starting point (generated by a different random seed), but the starting points were the same across all conditions. Each simulation was run for the same number of generations.

Analysis

In order to implement at least one of the constraints (constraint 3) and to evaluate the effect that the constraints had on the systems it was necessary to segment the words into syllables. The ESS syllabification routine followed the maximal onset principle and the sonority principle (Venneman, 1972; Hooper, 1976; Selkirk, 1982). The maximal onset principle was implemented by syllabifying from the end of the string to the beginning. Jaw openness values were used to implement the sonority principle. Syllabification therefore proceeded in the following manner. If the last phoneme in a word was articulated with more jaw opening than the one before it, the last phoneme was considered the nucleus for the syllable (even if it was a consonant). The next syllable boundary was then found by moving towards the beginning of the word until a change in the direction of openness occurred. For example, the word "pasli", with the openness values 4 9 2 5 6, would be syllabified as /pa.sli/, that is, as a CV and a CCV syllable. If the last phoneme before a boundary (word or syllable) was articulated with less jaw opening than the one before it, the last phoneme was considered a syllable-final consonant or vowel. The nucleus of the syllable was found when the direction of openness changed. For example, the string "pauspil", with the openness values 4 9 6 2 4 6 5, would be syllabified as /pau.spil/, that is, as a CVV and CCVC syllable. Syllable boundaries were also inserted between adjacent segments that had identical openness values (e.g., "appa" syllabified as /ap.pa/). If more than two adjacent segments had equal openness values, the onset of one of the syllables was maximized (e.g., "apppa" syllabified as /ap.ppa/).

Once words were evolved and syllabified, the most fit vocabularies from the final generation of evolution were compared and contrasted along the following dimensions: word length, number of syllable tokens and types, number of CV syllables, number of syllables with initial or final consonant clusters, and number of word-initial vowels and word-final consonants. Since each instantiation of the model was run 20 times with the same set of start-up seeds, the 20 different vocabularies produced under one instantiation of the model could be directly compared with the 20 different vocabularies produced under the other instantiations of the model. This type of analysis made it possible to evaluate the role of each constraint on vocabulary evolution and to understand the interaction between the various constraints.

Finally, the fact that all vocabularies were taken from the same point in evolution allowed each of them to be treated as a different "language" at a single point in time. This synchronic perspective made it possible to compare and contrast each of the twenty final best vocabularies of a single model—in this case the full model—to determine whether cross-language variation is consistent with a single optimization problem, i.e., to verify the second hypothesis.

Predictions

The following syllable system universals were expected to emerge. First, vocabulary structure was expected to differ in straightforward ways across different instantiations of the model. Specifically, syllable systems generated under the full model were expected to reflect all major cross-linguistic preferences. CV syllables were expected to be present in each of the vocabularies and to be the most frequent syllable type. Initial vowels and final consonants as well as consonant clusters were expected to occur infrequently. When consonant clusters occurred, syllable-initial clusters were expected to predominate over syllable-final clusters.

In contrast to the full model, the 1st partial model and the null model were expected to generate vocabularies with greater numbers of syllable types. The 2nd partial model and the null model were expected to have greater numbers of word-initial vowels and word-final consonants than the full model. The 3rd and 4th partial models as well as the null model were expected to generate vocabularies with many consonant clusters. Finally, vocabularies evolved under the null model were expected to contain shorter words and words with more varied syllable structures than those generated under the other instantiations.

The expected relationship between short words and a varied syllable structure in the null model points to another, more general pattern expected to emerge in the simulated vocabularies. Just as it was expected that the many different segment combinations, i.e., syllable types, of the null model would allow for distinctive words that were short, it was expected that distinctive words would be longer in models that placed restrictions on segment combinations. Thus, word length was expected to vary inversely with syllable type number such that vocabularies with more syllable types would have shorter words than vocabularies with fewer syllable types.

Another prediction concerns the relative frequency of different syllable types. The constraints were expected to yield syllable systems in which complex syllables would be less frequent than simpler ones. The CV syllable is the most basic syllable form and also the most frequent. We suggested that new syllable types emerge from the addition or deletion of a single consonant or vowel. Additional increases in complexity emerge in an iterative fashion with new additions of a single consonant or vowel. We named this process the iterative principle of syllable structure. It is expected that each new, more complex syllable type will be less common than the preceding one.

Finally, it was predicted that individual vocabularies generated under a single instantiation of the model should exhibit an internal syllable structure that differs from the other vocabularies generated under the same model. Like natural languages, the syllable structure of each vocabulary should be adequately described by a well-defined set of syllable types. Different sets should characterize different vocabularies. These categorical differences would lend support to the hypothesis that language variation emerges in spite of a single optimization problem.

RESULTS

Optimization Process

Evolution significantly increased the fitness of all vocabularies with the exception of those generated under the null model. Figure 1 displays the fitness curves for the vocabularies under the each of the six conditions. Each curve represents the average decrease in penalties for 20 runs.

As shown in Figure 1, the fitness curves of the full model and those of the 2nd, 3rd, and 4th partial models overlap. Also shown in Figure 1 is that overall fewer penalties were associated with the vocabularies generated under the 1st partial model and fewer still with those under the null model. The reason that fewest penalties were associated with the null model was that only the general constraints on word production and meaning (constraints 1 and 2) were active, whereas in the other models a minimum of five constraints were active. The reason that fewer penalties were associated with the 1st partial model than with the other highly constrained models was that the former excluded a memory constraint on syllable types (constraint 3) and the latter did not. We conclude from these differences in overall penalties that the general constraints on word production and meaning conflict with the more specific phonetic constraints on word production and perception (constraints 4, 5, 6) and that the memory constraint conflicts with some of the other constraints, though there is not enough information to determine exactly which ones.

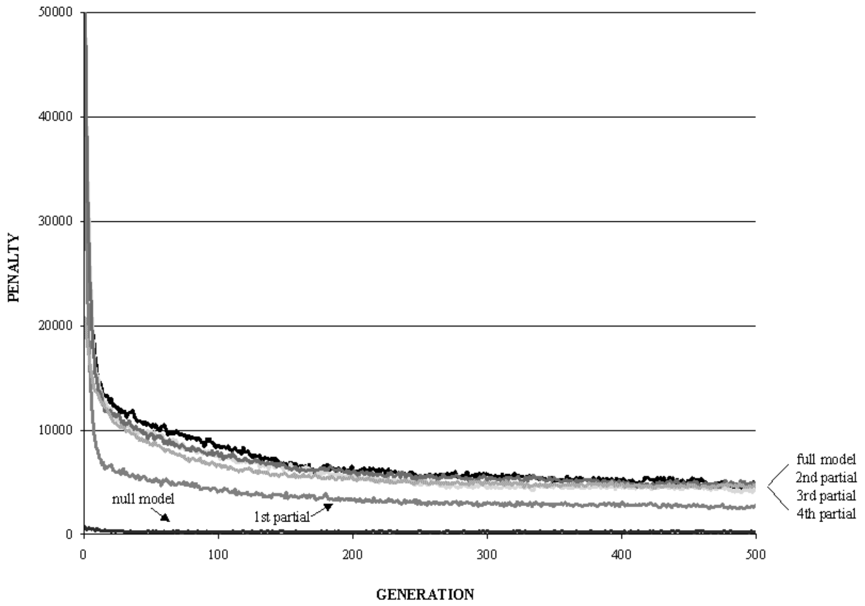

Figure 1: Average fitness curves are displayed for the 20 best vocabularies generated under the different models. Generations 0 to 500 are shown along the x-axis and penalties are shown along the y-axis. The figure shows a decrease in penalties across the first 350 generations of evolution for all instantiations of the model except the null model. Penalties remain constant for the 500 generations of evolution under the null model and for the last 150 generations of evolution under the other models.

Figure 1 also shows that there was no improvement in fitness after 350 generations of evolution. The simulations were run until generation 500 to ensure that fitness increases had leveled off.

The constrained evolution in ESS allows phonological structure to emerge. A standard way to quantify an increase in structure is to measure entropy. However, it is unclear how entropy should be defined in this case. An accurate definition would require enumerating every possible regularity in the system, which is not tractable. Even if we limited the enumeration to syllable types, the definition would be unnatural. The majority of possible syllable types never occur at all in the simulations and many others are weeded out very early. A more reasonable way to measure an increase in the structured set of phonological patterns is to count the number of different segment combinations (i.e., syllable types) that occur in the best vocabularies of each generation of evolution. In initial, randomly organized vocabularies, many different syllable types will be expected. As vocabularies become more structured under the pressure of the constraints, syllable type number should decrease.

Figure 2 displays a syllable type count for the best vocabularies of the full model, the 1st partial, and the null model across 500 generations of evolution. Only these three models are shown since they differed substantially from one another. Vocabularies generated under the full model increased in phonological structure more than those of the 1st partial model. No increase was observed in vocabularies of the null model. The other partial models resulted in vocabularies similar to the full model. These results indicate that syllable structure emerged as a result of the three specific phonetic constraints on segment organization (constraints 4-6) and the more general memory constraint on number of syllable types (constraint 3).

The effect of evolution on vocabulary structure is also seen in Table 2, which shows an example initial (generation 0) and final (generation 500) vocabulary from the full model. Whereas no regular phonological pattern exists in the initial vocabulary, the final vocabulary is regularly structured and exhibits patterns common to natural syllable systems. The syllabification routine identified 29 syllable types in the initial vocabulary compared with 3 for the final vocabulary, i.e., CV, CVV, and CVC. In addition, the frequency of the different syllable types in the final vocabulary is consistent with what might be observed in natural languages: CV syllables predominate (83 tokens), CVV and CVC syllables are almost equally frequent (18 vs. 21 tokens), word-initial vowels do not occur, word-final consonants occur, but these are rarer than word-final vowels.

In summary, a clear increase in organization, as measured by a decrease in the number of syllable types, is seen in the vocabularies generated under the highly-constrained models. In contrast, the vocabularies of the null model show no such increase in organization. In the next section, the emergent structure of the final best vocabularies is compared and contrasted for all instantiations of the model.

Universals

The final best vocabularies (from generation 500) were compared and contrasted in terms of (1) the relative frequency of CV syllables, (2) syllable-initial and syllable-final consonant clusters, and (3) word-initial vowels and word-final consonants to determine how individual constraints affect the emergence of language-like universals. The relationship between syllable type number and word length was also examined to establish whether vocabularies with the greatest number of syllable types had the shortest average word length and vice versa. Finally, the iterative principle of syllable structure was evaluated by examining the occurrence of different syllable types and their relative frequencies in the vocabularies generated by the full model.

Syllable Types. Figure 3 displays the average percentages of CV syllables, syllable-initial and syllable-final clusters, as well as word-initial vowels and word-final consonants across the six different instantiations of the model.

Figure 3 shows that the basic syllable structure of the vocabularies varied from one condition to another. Repeated measures analyses of variances (ANOVA) confirmed that these differences were

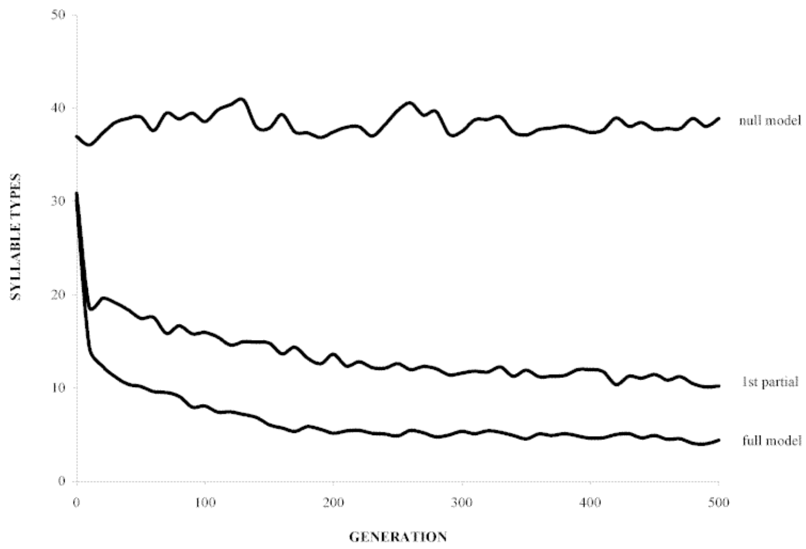

Figure 2: The average number of syllables types are displayed for the 20 best vocabularies generated under the full model, the 1st partial model, and the null model across 500 generations of evolution. The figure shows a gradual decrease across the generations in the number of syllable types associated with the vocabularies of the full model and the 1st partial model. The number of syllable types remains constant across 500 generations of evolution in vocabularies generated under the null model.

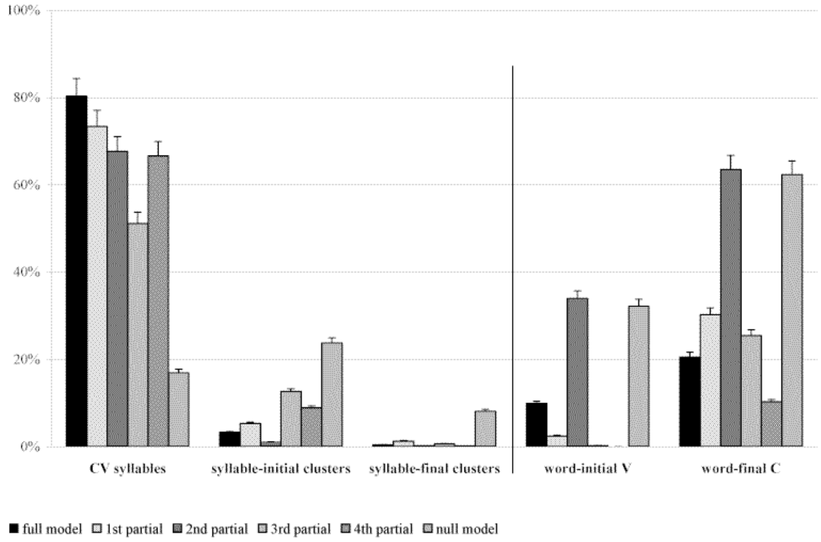

Figure 3: The average percentage of CV syllables, syllable-initial and syllable-final consonant clusters, word-initial vowels and word-final consonants are displayed for the final best vocabularies from each of the six instantiations of the model. The figure shows that the vocabularies differed on each of these measures depending on which constraints were active during the optimization process. The full model included all constraints, the 1st partial model excluded the constraint on number of syllable types, the 2nd partial model excluded the constraint on segment position, the 3rd partial model excluded the constraint on adjacent segments, the 4th partial model excluded the constraint on segment organization, and the null model excluded all syllable-specific constraints.

Table 2: Example initial and final vocabularies from the full model. Blanks in the initial vocabulary indicate words of length 0. It is evident that no regular syllable structure exists in the initial vocabulary, whereas words of the final vocabularies are clearly composed of just a few syllable types.

Initial Vocabulary		Final Vocabulary	
nial	ska	kiliaku	puliasu
lpnssu	klplptai	pitaa	sitilsu
p	lppnntustnu	kukin	saliaku
tpltintit	aiik	situn	pata
aiaautk	iapansk	kupipsu	pat
nlasanusiua		tilianu	lipu
ittlliiluau		tinianu	kapulsu
ut	kplk	pisiaku	kapun
p	nsininikaps	tilu	piliatu
alsnsu	ktptailiklt	pitiltu	putu
s	s	kituaku	kupipsu
u	uia	pitu	pitu
aipnaspstta	p	kuniaku	kitiaki
kisikaspatp	iinaptaplsu	latip	kupunku
	sntikn	tatilsu	silu
iit	upnppiipknu	kuliatu	ninuanu
stiaitsk	ku	kupun	kap
pisun	iiapistitpp	nipultu	pitin
katppaktllt	k	tiliatu	tatua
tnunptnikls	stiaitsk	titilsu	kilia
tl	uupsnklpkuu	ninianu	kuti
tis	nik	pitin	kuti
kstn	kiskpp	kunus	titilsu
saip	iapp	pip	tatuanu
utsikalliap	n	sati	kunu

statistically significant. Vocabularies generated under the different instantiations differed in number of syllable types ($F(5,114) = 640.69, p < 0.01$), in the proportion of CV syllables to the total number of syllables ($F(5,114) = 46.88, p < 0.01$), in the proportion of syllable-initial ($F(5,114) = 34.7, p < 0.01$) and syllable-final consonant clusters ($F(5,114) = 69.49, p < 0.01$) to the total number of syllables, and in the proportion of word-initial vowels ($F(5,114) = 11.69, p < 0.01$) and word-final consonants ($F(5,114) = 47.45, p < 0.01$) to the total number of word-initial segments. These main results are analyzed in more detail below.

CV syllables occurred in every vocabulary of every full and partial model. The frequency of CV syllables differed, though, as a function of which constraints were active. Given 6 consonants and 3 vowels, the likelihood that a consonant and vowel would co-occur by chance was 21.8%. CV syllables occurred more frequently than this in all highly-constrained instantiations of the model. CV syllables were most frequent, however, in the full model and the 1st partial model, that is, in vocabularies constrained to favor short, different words (constraints 1 and 2), word-initial consonants and word-final vowels (constraint 4), non-identical adjacent segment types (constraint 5), and segment sequences obeying the jaw cycle (constraint 6). As shown in Figure 3, CV syllables were less frequent in those partial models that excluded only one of the more specific constraints. When vocabularies were only constrained to favor short, different words, as in the null model, CV syllables occurred at below chance levels. These results suggest that the ubiquity of CV syllables in the simulated vocabularies and their high frequency within a vocabulary is due to the combined effect of constraints 4, 5, and 6.

In contrast to CV syllables, syllables with initial or final consonant clusters were infrequent compared with other syllable types across all instantiations. Also unlike CV syllables, syllables with clusters occurred in only a portion of the vocabularies generated under the full model and the 2nd partial model. Syllables with either initial or final clusters occurred in 11 of the 20 vocabularies under the full model and in

10 of the 20 vocabularies under the 2nd partial model. Syllables with clusters occurred in 19 of the 20 vocabularies generated under the 4th partial model and in every vocabulary generated under the 1st partial, 3rd partial, and null model. These results suggest that syllables with clusters occur infrequently in the full model because this model incorporated constraints on syllable types (constraint 3), on adjacent segments (constraint 5), and on segment sequences (constraints 6). One of these constraints was absent in the 1st, 3rd, and 4th partial models and all were absent in the null model.

The results also indicated that the constraint on segment position (constraint 4) effectively encourages initial consonants and final vowels. With 3 vowels and 6 consonants and no constraint on where they can occur, chance dictates that 33% of the words should begin with a vowel and 66% should end in a consonant. On average, the frequency of initial vowels and final consonants were at these levels in the vocabularies of the 2nd partial model and null model (see Figure 3), neither of which included constraint 4.

Although both initial vowels and final consonants occurred well below chance levels in the vocabularies of the other models, final consonants were at least twice as frequent as initial vowels. Examination of the proportion of vocabularies with initial vowels and/or final consonants revealed a similar difference in frequency. Word-initial vowels occurred in only a small subset of the vocabularies of the highly-constrained instantiations. Word-initial vowels occurred in 6 of the 20 vocabularies of the full model, in 8 of the 1st partial model, 3 of the 3rd partial model, and in none of the vocabularies of the 4th partial model. In contrast, word-final consonants, though also penalized in these models, occurred in most of the vocabularies generated under the different instantiations. Word-final consonants occurred in 18 of the 20 vocabularies generated under the full model and the 4th partial model and in every vocabulary generated under the other partial models. These differences in initial vowel frequency versus final consonant frequency may reflect the different underlying probabilities associated with the occurrence of each (i.e., 33% for vowels versus 66% for consonants).

In addition to encouraging initial consonants and final vowels, the constraint on segment position (constraint 4) was further expected to affect the frequency of syllable-initial and -final consonant clusters. As expected, initial clusters predominated over final clusters, but because this difference occurred under every instantiation of the model, including those that excluded constraint 4, the difference cannot be attributed to the constraint. Instead, the higher frequency of onset clusters was probably an artifact of the syllabification routine, which incorporated a principle that maximized syllable onsets over syllable offsets (the maximal onset principle).

Word Length. Another regularity predicted to emerge in the simulated vocabularies was based on the expectation that the shortest words are most likely to occur when there are no restrictions on the types of segment combinations allowed. Accordingly, it was expected that the null model, with the fewest restrictions on segment combinations, would have the shortest words (and the most syllable types), whereas the full model, with the greatest restrictions on segment combinations, would have the longest words (and the fewest syllable types). To evaluate this hypothesis we compared the average word length with the average number of syllable types in the six models. Word length was expected to vary inversely with syllable type number.

Word length was measured as a function of segment number. We found that, contrary to our expectations, words were actually longer in vocabularies that had more syllable types. As shown in Figure 2 above, the final vocabularies of the full model had the fewest syllable types and those of the null model had the most. This difference paralleled a difference in average word length. The average length of words generated under the full model was 7 segments. In contrast, the average length of words generated under the null model was 11 segments, which is not much shorter than the maximum of 12 allowed by the model. The same relationship held for the 1st and 3rd partial models, where the average word length (9 and 10 segments) and the average number of syllable types (10 and 6 types) were both greater than in the full model (7 segments and 4 types). The relationship was violated in the 4th partial model, where average word length was greater than in the full model (10 vs. 7 segments), but the average number of syllable types was equally small (4 types).

The observed relationship between word length and number of syllable types was in direct contradiction to the inverse relationship we expected. Word length was shortest in vocabularies with the fewest number of syllable types. One possible explanation for this result is that the regular organization of phonemes into syllable patterns preserves intelligibility while reducing production costs associated with word length.

The Iterative Principle of Syllable Structure. A final prediction about syllable structure universals was that the functional constraints would determine the relative frequency of the different syllable types. Simpler syllable types were expected to occur more frequently than complex ones in a systematic fashion we termed the iterative principle of syllable structure.

The CV syllable is the simplest and most frequent syllable type. The first degree of complexity is achieved by either adding a consonant or vowel to the CV form, or by deleting the consonant. Further increases in complexity take place by adding further single consonants or vowels. For example, a CVC syllable is one degree more complex than a CV syllable and a CCVC syllable is one degree more complex than a CVC syllable or two degrees more complex than a CV syllable. More complex syllable types are less frequent than simpler ones in a systematic fashion. For example, syllable types that are two degrees more complex than a CV syllable should be less frequent than syllables that are one degree less complex, and so on. In the present example, CCVC syllables are predicted to be less frequent than CVC syllables, which are predicted to be less frequent than CV syllables. To evaluate this hypothesis, the frequencies of all syllable types were determined for the vocabularies of the full model and the null model, and compared.

In addition to the CV syllable, 14 other syllable types occurred in at least one of the 20 final best vocabularies of the full model. In Table 3, these syllable types are presented in order of increasing complexity together with their frequency in the full model. As predicted, the least complex syllable types were the most common and the most complex were the least common. CV syllables accounted for 76.3% of all syllable tokens; of the non-CV syllable types, 20.82% of the tokens were one degree more complex than CV and 2.3% were two degrees more complex. Fewer than 1% of the non-CV type tokens were more than two degrees more complex than the basic CV syllable.

In contrast to the full model, CV syllables accounted for only 17.11% of the syllable tokens generated under the null model. Of the non-CV syllable type tokens, 27.89% were one degree more complex than CV, 23.09% were two degrees more complex, and 31.9% were three or more degrees more complex. These results from the full and null model support the hypothesis that functional constraints affect syllable type frequency in a systematic fashion.

The iterative principle, which may be used to define syllable complexity, also suggests a relationship between the occurrence of particular syllable types in the system. Since more complex syllable types are built in an iterative fashion from less complex types, more complex types should only occur in a system that has the simpler types. For example, if a CCVC syllable type exists, then a CCV and CVC syllable type should also exist. This prediction was usually upheld in the full model, but certain exceptions did occur. For example, Table 3 shows that although a VVCC syllable type occurred, the simpler types VCC, VVC, and VC did not. It is possible that a systematic relationship between different syllable types would be more robust in a system that grew in complexity and size over time.

The above results suggest that the iterative principle of syllable structure is useful for predicting the complexity and frequency of different syllable types. With more detailed study it could lead to insights into why certain types occur and others do not.

Variation

Although a number of common sound patterns exist across languages, languages are often characterized in terms of categorical differences. For example, a language may be characterized according to whether

Table 3: Frequencies of different syllable types in the full model listed in order of increasing complexity. In general, more complex syllable types are less frequent than simpler ones.

Basic		Basic+1		Basic+2		Basic+3		Basic+4	
CV	2682	CVC	380	CVVC	1	CVCC	18	VVCC	1
		CCV	113	CCVC	2	CCC	2		
		CVV	78	CCCV	3				
		V	97	CVVV	1				
		C	64	VV	8				
				CC	66				

or not it allows syllables with codas (e.g., CVC), syllables with clusters (e.g., CCV), syllables without onsets (e.g., V), or any of a number of other characteristics.

The vocabularies generated by the ESS model also exhibit categorical differences. Table 4 displays three different vocabularies of the full model. The syllable structures of each of these vocabularies conform to the universals discussed above, but each also differs from the others in categorical ways. For example, syllable-initial clusters and syllable-final consonants occur in Vocabulary A, but syllable-final clusters do not. In contrast, Vocabulary C, which also allows syllable-final consonants, only has syllable-final clusters. Vocabulary B is distinguished from A and C in that no syllable-final consonants occur.

These three vocabularies can be further contrasted with the other vocabularies generated under the full model (not shown in Table 4). For example, word-initial vowels did not occur in any of the three vocabularies presented in Table 4, but they did occur in 6 of the 17 other vocabularies.

The different vocabularies represent different good solutions to the problem defined by the six constraints. This is evident when variances of syllable type frequency are compared for the full and null models (Figure 4).

Figure 4 shows the variance for 4 syllable types. If there were only one good solution, all 20 runs of the full model would converge on this same solution. The variances would be small and due to chance, like in the null model. However, in every case, variances are greater in the full model than in the null model. Since all vocabularies were run to the same point in evolution, the greater variance in the full model indicates that the optimization problem was solved in different ways on different runs of this model. This result suggests that language variation could be due to the fact that multiple good solutions exist for a single problem posed by universal functional constraints.

GENERAL DISCUSSION

These simulations show that the constraints of ESS define language-like syllable systems. The sound patterns that emerge in every simulation are similar to those observed across languages. The variation observed among identically-constrained vocabularies supports the hypothesis that language variation represents different good solutions to a single optimization problem. Below, the emergent regularities and variation across vocabularies in ESS are compared with those of real language. The relationship between ESS and language evolution is then examined and the status of the model's assumptions is explored.

Universals

As expected, a number of general syllable patterns emerged in the simulated vocabularies of ESS. Some of these patterns replicated previously-described cross-language regularities in syllable systems, while others represent novel hypotheses about language structure. The simulations therefore demonstrate of how functional constraints might interact to define complex sound patterns, such as syllable systems, in language.

Table 4: A sample of three final best vocabularies evolved under the full model. Each vocabulary exhibits language-like regularities, but each also differs categorically from the others.

Vocabulary A		Vocabulary B		Vocabulary C	
sutuap	suskanal	tisitalu	pipipitu	pipik	sunatu
tuntapi	satuapi	tisinalu	tu	pasila	lupila
sutkata	satkata	kipinu	lipusapu	pu	kipinu
suskapal	tn	kulinu	sulinu	pinul	piputu
tuntakin	sutuasi	tipinilu	pulinuti	kipa	pukula
taslasi	litkapi	punina	pisunusi	pu	piputu
sutlasi	saukata	tipinalu	pusinipu	pinuk	susus
su	supnika	tulipisu	pilinapu	pipik	tupul
sutuasa	lunkapa	nulinu	pisunusi	pinulak	pukula
lun	sutkapi	pipinu	lulinu	pitutuk	pupila
sutuasi	satua	tisinu	tipunalu	si	siputuk
sutkapal	lupiaki	pulini	tipulani	sututis	pupilk
putkana	saskasik	pipinu	tisinu	piniku	pupila
saspapa	titsapa	tulikisi	tuputisa	pinika	pupina
satkap	kuklaka	tilikipu	pulipilu	sinuta	lukula
saspapa	susnapa	kulinu	talinatu	kipalu	lukula
luskapal	taklaka	pulupalu	lipupuli	lapik	pipup
susnapi	satsapa	tulinala	luskni	pinitis	sinitis
litkasi	tit	tulipi	puninupi	pupil	lapil
putlasin	sutkasi	tipinu	tulinalu	nup	papila
tit	nutlasin	nulusa	lisusi	sipitu	kipinu
lktkasis	puniati	pitinapu	pusina	putilp	pasil
saklaki	sitkasi	pulina	puninupi	lipiku	nuliku
suslapa	latkasi	nipasisa	puputinu	kipipu	lapikis
su	pustapik	kulinitu	tu	punitis	pupilu

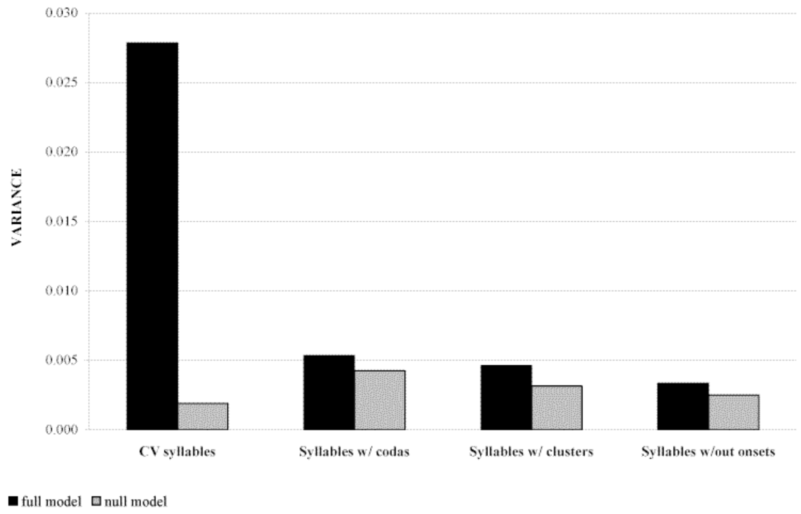

Figure 4: Variance is shown for CV syllables, syllables with final consonants, consonant clusters, and vowel onsets in the full and null models. These particular syllable types were used to compare and contrast syllable structure across the different vocabularies in the full model. The figure shows that variances were greater in the full model than in the null model indicating that the optimization problem was solved in different ways on different runs of the full model.

The frequencies of different syllable types in the emergent vocabularies correspond to those same types in natural languages. The most dominant pattern in ESS was that CV syllables, which occur in all languages, occurred in every vocabulary and under every instantiation of the model, but at above-chance levels only in the highly-constrained models. In particular, CV syllables were most frequent in the full and 1st partial models, which incorporated all specific phonetic constraints on segment position and organization. It may therefore be concluded that the phonetic constraints combine to strongly favor CV syllables over other syllable types in a vocabulary.

This result is interesting since CV syllables also appear to be the most frequent syllable type in most of the world's languages. Although such a claim can be fully substantiated only by studying a large sample of languages, even a small initial sample provides convincing evidence, as shown in Table 5. One hundred words were randomly selected from different language dictionaries. The words were then syllabified using the same principles employed by ESS. With the exception of Efik, CV syllables were the most frequent type in all languages. Furthermore, in most languages, CV syllables accounted for at least half of the total number of syllables in the sample. ESS suggests that different phonetic constraints combine to make CV syllables the most frequent syllable type in languages.

Unlike CV syllables, syllables with initial vowels, final consonants, and consonant clusters do not occur in all languages. ESS points to a potential explanation: These structures are each disfavored by a particular phonetic constraint. For instance, initial vowels and final consonants occurred at chance levels in vocabularies generated under the 2nd partial model and the null model, both of which excluded a perceptually motivated constraint on segment position. By contrast, initial vowels and final consonants occurred much more rarely in vocabularies that included such a constraint.

In addition to the expected regularities, an interesting and unanticipated syllable pattern emerged in the vocabularies: Final consonants occurred more frequently than initial vowels under every instantiation of the model. The reason is simply that there were more consonants than vowels in the system. Nevertheless, the pattern is consistent with the typological data. Bell and Hooper (1978) noted that in 20–40% of the world's languages words must begin with a consonant, but in only 10–25% must they end in a vowel. Results from ESS suggest that this asymmetry occurs because most of the world's languages have more consonants than vowels (Maddieson, 1984).

Another unexpected pattern was that vocabularies with fewer syllable types had shorter words. For example, the full model generated vocabularies with fewer syllable types and shorter words than the partial models. An analysis of the small set of diverse languages displayed in Table 4 suggests that this relationship may not hold for language. English has the most syllable types (N=18) and the shortest words (measured by the total number of syllable tokens, 206, divided by the number of words sampled, 100). In fact, these data suggest that perhaps no relationship exists between syllable type and word length at all. Hawaiian, with the fewest number of syllable types (N=4), has shorter words than five other languages—Czech, Spanish, Alabama, Dakota, and Luganda—that have 7 to 11 syllable types. Even so, it is interesting to see a correspondence between syllable type number and word length emerging in ESS. Given the constraints of ESS, it suggests that increasing phonological structure (fewer syllable types) decreases speech production costs (word length). This relationship between syllable type number and word length may not hold in natural languages because all may be characterized by a similar degree of phonological structure.

Finally, the iterative principle of syllable structure was predicted to emerge from constraint interaction. In addition, the simplest syllable types were predicted to occur more frequently than more complex types. Both patterns were confirmed in ESS. The typological data are also consistent with these predictions. Blevins (1995) notes, for example, that "if clusters of n Cs are possible syllable-initially, then clusters of (n - 1) Cs are also possible" (p. 217) and the same relation holds in syllable-final position. Data from Table 4 provide further evidence on how simple and complex syllable types occur. As previously noted,

Table 5: The frequency of different syllables types are displayed for a diverse group of languages. Syllable types were derived from a random sample of 100 words per language from the pronunciation guidelines provided by dictionaries. (Sources: Czech, Poldauf 1986; English, Halsey 1979; Spanish, Inelva 1985; Alabama, Sylestine et al. 1993; Dakota, Riggs 1968; Efik, Goldie 1964; Luganda, Murphy 1972; Mansaka, Svelmoe and Svelmoe 1990; Hawaiian, Judd 1939; Japanese, Takahashi 1953.)

Language	CV	CVC	CCV	CVV	V	VC	CCVC	CCVV	other
Czech	136	56	42	4	9	5	13		10 (N=4)
English	77	51	8	7	7	8	9	2	37 (N=10)
Spanish	195	65	20	25	12	15	5	2	11 (N=2)
Alabama	181	86		50	20	23			5 (N=2)
Dakota	194	56	31		54	3	17		2 (N=1)
Efik	42	81	4	2	71	14	6		16 (N=2)
Luganda	214	29	25	29	2		2	1	3 (N=1)
Mansaka	112	92	3		14	11			
Hawaiian	178			57	28				9 (N=1)
Japanese	198	30	5	22	10	1			1 (N=1)

CV syllables are the most common syllable type in each of these languages (with the one exception of Efik), and the next most frequent types are, as predicted, one degree more complex than the basic CV syllable, i.e., CVC, CCV, CVV, and V. More complex syllables occur even less frequently.

The data in Table 5 also support another prediction of the iterative principle: if a more complex syllable type occurs, then the less complex form also occurs. For example, a CCVV syllable only occurs in those languages that also have CVV syllables.

In summary, the emergent syllable systems of ESS show the same syllable patterns as natural languages. The patterns in ESS emerge from the interaction of a number of functional constraints, which implies that such patterns in natural languages may also be due to interacting constraints. Whereas previous demonstrations of emergent sound structure explain each pattern in terms of a single constraint, the present simulations demonstrate how a number of well-defined and well-justified constraints may interact to explain more complex patterns.

Variation

Even though the vocabularies of the full model were similar in certain respects, they also displayed categorical differences. For example, some vocabularies did not include syllables with final consonants, some did not include syllables with vowel onsets, and others did not include syllables with consonant clusters. In contrast, the variation was minimal among the vocabularies of the null model. This result suggests that variation in the full model was not randomly generated; instead it emerged from an optimization process that found a number of equally-good sets of syllable types to satisfy the constraints.

The ESS model suggests that categorical differences in natural languages may be due to chance. In the space of possible syllable types, many different subsets are available. The subsets selected for by the constraints are never mutually exclusive—all share the basic CV syllable type—but different sets of the less frequent types are equally-good at satisfying the conflicting constraints.

Even though the ESS results are strong, they are based on rather small vocabularies. Larger vocabularies, such as those of natural languages, demand many more distinctive combinations of segments, and it may be less likely that different syllable structures could emerge by chance: Larger vocabularies may require larger subsets of syllable types, which would have a greater chance of overlapping. In future work, we will include larger word and phoneme inventories in the ESS model to study the effects of size on the emergence of syllable structure.

In summary, the variation observed between the identically-constrained, evolved vocabularies of ESS demonstrates that language variation may be due to chance in the optimization process. This hypothesis can be contrasted with the one proposed by Optimality Theory (OT, Prince and Smolensky, 1993), which explains that language variation emerges from different optimization problems. The former hypothesis is favored over the latter because it makes simpler assumptions about language history. The demonstrations offered in this paper do not, however, allow conclusive rejection of the OT hypothesis. Accordingly, in future work with ESS we will more rigorously test these two competing hypotheses.

In order to perform such a test, we plan to match the vocabularies of ESS with data from different languages, and search for those weight settings that result in better matches. Finding the appropriate weights manually would be laborious, because the weight space is potentially infinite. Instead, another level of evolution will be incorporated into the ESS model as an outer loop above the vocabulary evolution. At this level, constraint weights will be evolved, and their fitness determined based on how closely the resulting syllable systems match those of different target languages. If cross-language variation results from different optimization problems, it is expected that constraint weightings will differ significantly between different languages. If, however, cross-language variation is due to chance in the optimization process, constraint weightings will be similar for all languages. In sum, the proposed modifications to ESS would test both hypotheses on cross-language variation and would therefore provide a stronger basis for favoring one hypothesis over the other.

ESS and Language Evolution

The current version of ESS is aimed at providing a meaningful account of how syllable structure universals and variation emerge. The explanatory value of ESS is based on four critical assumptions: (1) Language is a generative system, that is, a system in which a finite set of units is combined into a potentially infinite set of larger units. (2) Language phonology is an abstraction of regularities in the sound system. (3) Functional constraints affect the structure of language. And, (4) language structure emerges over time. The first of these assumptions is uncontroversial. The particular definition of phonology given in the second assumption should also be uncontroversial, since no claim is being made about the nature of phonological representation. The third and fourth assumption, however, are more controversial and so will be discussed in more detail.

The assumption that functional constraints affect the structure of language is not new. Many phoneticians have explained phonological regularities in terms of phonetic constraints (e.g., Ohala, 1983; Lindblom, MacNeilage, Studdert-Kennedy, 1984; Lindblom, 2000) just as other linguists have explained syntactic and semantic regularities in terms of cognitive or communicative constraints (e.g., Givon, 1979; Du Bois, 1987; Croft, 1990). These explanations of language universals are controversial, though, because they assume that language structure results from external pressures on the system. This assumption is in direct contrast to one held by many linguists, namely, that language structure results from a mental and domain-specific representation of language, i.e., a grammar.

Newmeyer (1998) has convincingly argued that, when language structure is analyzed without reference to language history, it cannot be explained to emerge from external pressures alone. He uses the specific example of cross-language variation to show that speakers must have some form of language-specific knowledge or grammar. He notes, for instance, that even if different interactions of functional constraints account for different language structures, speakers will need to know how the constraints interact in a particular language in order to produce appropriately structured speech. Since this language-specific knowledge would constitute a type of grammar, defining the grammar should adequately explain language structure.

Although we believe that Newmeyer (1998) is correct to argue that speakers have a mental and domain-specific representation of language, this does not mean that language structure results from this representation.

Newmeyer's particular critique of the functionalist program is dependent on a synchronic analysis of language. We agree that speakers do not produce structured language at a given point in time because functional constraints affect their output. However, this type of critique is no longer relevant when a diachronic perspective is adopted, as in the ESS model. Although functional constraints may not adequately explain language structure at a single point in time, these pressures may operate over time to preserve speaker and listener innovations, thereby affecting language structure. The ESS results demonstrate this by showing how regularities and differences in syllable structure emerge solely from the action of functional constraints over time. The results do not address, nor depend on, how language structure is represented in the speaker's mind. They do suggest, however, that grammar is not the only explanation for language structure.

For those who explain language structure in terms of grammar (i.e., formalists), functional explanations often remain suspect even when a diachronic perspective is adopted. This is because, like functional explanations, formal explanations can be recast from a synchronic perspective into a diachronic one. From a diachronic perspective, a formalist might explain that natural selection pressures operate on an innate, universal grammar that defines language structure (e.g., Pinker and Bloom, 1991; Pinker, 1994). In this view, changes in language structure over time reflect modifications to the genetic code underlying language.

At this point in time it is not possible to truly distinguish between a view of language evolution in which language structure results from functional constraints on the system and one in which it results from an underlying mental representation such as an innate, universal grammar. Although the concept of a universal grammar may be attractive, the biological evidence for it remains controversial. In particular, it has not been possible to identify genes that would be responsible specifically for language. On the other hand, hypotheses regarding functional constraints on the system can be tested in phonetic or psychological experiments. The constraints can then be formalized in a computational model such as ESS and their affect on language structure studied through simulated evolution. In light of this, we believe a functional approach is more promising for understanding the emergence of language structure.

In summary, the assumptions that underlie ESS are either uncontroversial or relate to a theoretical framework that may be controversial, but provides promise for understanding the emergence of language structure. The model therefore provides a theory-based account of how syllable structure universals and variation might emerge over time.

Finally, it should be noted that the current version of ESS is not intended to model the actual evolution of language as it might occur in the real world. Sound patterns of language do not represent judicious selections of word types from random concatenations of phonemes. Some phoneme combinations are unpronounceable in this model and would never occur in the real world. In addition, language vocabularies do not have a fixed size, but instead grow from a small set of words to a much larger one. ESS will be modified in the future to provide a more faithful model of the actual evolution of sound patterns. We will evolve phonemes as well as words according to the general phonetic principles of articulatory ease and perceptual distinctiveness. Initial vocabularies will be small and articulatory ease constraints will dominate. As vocabularies expand in size, perceptual distinctiveness constraints will play a larger role. With these modifications, we should be able to compare changes that take place during the optimization process in the model with those that take place over time in real languages.

CONCLUSION

We developed the ESS model to explore how syllable system universals and variation may have emerged from functional constraints on natural language. In its present instantiation, the model demonstrates how specific phonetic constraints might interact with more general communicative and cognitive constraints to produce the known cross-language regularities in syllable systems. The model also provides novel insights into several other patterns that occur in languages. For instance, the model predicts that the syllable types emerge following an iterative principle of syllable structure. Different syllable types are built in step by step fashion from a basic CV type. Types that are further and further removed from the basic CV form are less and less frequent. Results from the simulations and analysis of a small sample of diverse languages support this principle. Finally, ESS shows how variation in syllable structure can emerge despite universal constraints that are fixed relative to one another. The optimization problem posed by the constraints was solved in different ways on different runs of the model, resulting in categorical differences between vocabularies.

Overall, the ESS results demonstrate that computational modeling of emergent sound structure can provide valuable insights into why certain regularities occur in languages. The simulations also support a view of language evolution in which complex structure emerges over time due to pressures exerted on the system by functional constraints.

Acknowledgments

The authors would like to thank Sergei Bogdanov, Simon Kirby, Jason Eisner, and an anonymous reviewer for their insightful comments on earlier versions of this paper. This research was supported in part by National Science Foundation grant #IRI-9504317 and Texas Higher Education Coordinating Board grant #ARP-444 to the third author as well as by an NSF graduate research fellowship and an NIH (NIDCD) National Research Service Award to the first author.

REFERENCES

Bell, A. and Hooper, J. Bybee. (1978). Issues and evidence in syllabic phonology. In Bell and Hooper (Eds.), *Syllables and Segments*, (pp. 3–22). Amsterdam: North Holland.

Blevins, J. (1995). The syllable in phonological theory. In Goldsmith (Ed.), *The Handbook of Phonology*, (pp. 206–244). Cambridge: Blackwell.

de Boer, B. (2000). Emergence of sound systems through self-organization. In Hurford, Knight, Studdert-Kennedy (Eds.), *The Emergence of Language: Social Function and the Origins of Linguistic Form.* Cambridge: Cambridge Univ. Press.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics,* 24, 209–244.

Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English.* New York: Harper and Row.

Croft, W. (1990). *Typology and Universals.* Cambridge: Cambridge University Press.

Davis, B.L. and MacNeilage, P.F. (1995). The articulatory basis of babbling, *Journal of Speech and Hearing Research,* 38, 1199–1211.

Du Bois, J. (1987). The discourse basis of ergativity. *Language,* 63, 805–855.

Flemming, E. (1997). Phonetic optimization: compromise in speech production. *University of Maryland, Working Papers in Linguistics,* 5, 72–91.

Givon, T. (1979). *On Understanding Grammar.* New York: Academic Press.

Goldie, H. (1964). *Dictionary of the Efik Language in Two Parts.* Ridgewood: Gregg Press.

Halsey, W. D. (1979). *MacMillan Contemporary Dictionary.* New York: MacMillan Publishing Co.

Haspelmath, M. (1999). Optimality and diachronic adaptation, *Zeitschrift fuer Sprachwissenschaft,* 18/II.

Hooper, J. Bybee (1976). *Introduction to Natural Generative Phonology.* New York: Academic Press.

Inelva, S.A. (1985). *English-Spanish / Spanish-English Compact Dictionary.* Barcelona: Biblograf.

Joanisse, M. (1999). Exploring syllable structure in connectionist networks. In *Proceedings from the 14th International Congress of Phonetic Sciences, (ICPhS-99 San Francisco, CA)*, pp. 731–734.

Joanisse, M.F. and Seidenberg, M.S. (1997). [i e a u] and sometimes [o]: perceptual and computational constraints on vowel inventories. In Shafto, Langley (Eds.), *Proceedings of the 19th Annual Meeting of the Cognitive Science Society, (COGSCI-97, Palo Alto, CA)*, (pp. 331–336). Hilsdale, NJ: Erlbaum.

Judd, H.P. (1939). *The Hawaiian Language.* Honolulu: Hawaiian Service.

Keating, P.A., Lindblom, B., Lubker, J., Kreiman, J. (1994). Variability in jaw height for segments in English and Swedish VCVs, *Journal of Phonetics,* 22, 407–422.

Kirby, S. (1999). *Function, Selection, and Innateness: The Emergence of Language Universals.* Oxford: Oxford University Press.

Lindblom, B. (2000). Developmental origins of adult phonology: the interplay between phonetic emergents and the evolutionary adaptations of sound patterns, *Phonetica*, in press.

Lindblom B. (1998). Systemic constraints and adaptive change in the formation of sound structure. In Hurford, Studdert-Kennedy, Knight (Eds.) *Approaches to the Evolution of Language*, (pp. 242–264). Cambridge: Cambridge University Press.

Lindblom, B. (1983). Economy of speech gestures. In MacNeilage (Ed.), *The Production of Speech*, (pp. 217-246). New York: Springer-Verlag.

Lindblom, B., MacNeilage, P.F., and Studdert-Kennedy, M. (1984). Self-organizing processes and the explanation of phonological universals. In Butterworth, Comrie, Dahl (Eds.), *Explanations for Language Universals*, (pp. 181-203). Berlin: Mouton.

Lindblom, B. and Maddieson, I. (1988). Phonetic universals in consonant systems. In Hyman and Li (Eds.), *Language Speech and Mind* (pp. 62-78). London: Routledge.

MacNeilage, P.F. (1998). The frame/content theory of the evolution of speech production. *Behavioral and Brain Sciences,* 21, 499–511.

Maddieson, I. (1984). *Patterns of Sound.* Cambridge: Cambridge University Press.

Moriarty, D.E. and Miikkulainen, R. (1996). Efficient reinforcement learning through symbiotic evolution. *Machine Learning,* 22, 11–32.

Murphy, J.D. (1972). *Luganda-English Dictionary.* Washington, D.C.: Consortium Press.

Nettle, D. (1999). *Linguistic Diversity.* Oxford: Oxford University Press.

Newmeyer, F.J. (1998). *Language Form and Language Function.* Cambridge: Massachusetts Institute of Technology Press.

Nowak and Krakauer (1999). The evolution of language. *Proceedings from the National Academy of Sciences,* 96(14), 8028–8033.

Ohala, J. (1990). Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In Ziolkowski, Noske, Deaton (Eds.), *The Parasession on the Syllable in Phonetics and Phonology, Vol. 2: Papers from the 26th Regional Meeting of the Chicago Linguistic Society*, (pp. 319–338). Chicago:CLS.

Ohala, J. (1983). The origin of sound patterns in vocal tract constraints. In MacNeilage (Ed.), *The Production of Speech*, (pp. 189-216). New York: Springer-Verlag.

Pinker, S. (1994). *The Language Instinct.* New York: William Morrow and Co., Inc.

Pinker, S. and Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences,* 13(4), 707-784.

Poldauf, Ivan. (1986). *Czech-English Dictionary.* Prague: State Publishing House for Educational Literature.

Price, P.J. (1980). Sonority and syllabicity: acoustic correlates of perception. *Phonetica,* 37, 327–343.

Prince, A. and Smolensky, P. (1993). *Optimality Theory.* Ms., Rutgers and University of Colorado.

Redford, M.A. (2000). *The Constraint of the Jaw Cycle on the Sequential Articulation of Segments*, submitted.

Redford, M.A. (1999). *An Articulatory Basis for the Syllable.* Ph.D. dissertation, Psychology, University of Texas, Austin.

Redford, M.A. and Diehl, R.L. (1999). The relative perceptibility of syllable-initial and syllable-final consonants. *Journal of the Acoustical Society of America,* 106, 1555–1565.

Riggs, S.R. (1968). *Dakota-English Dictionary.* Minneapolis: RossHaines.

Selkirk, E. (1982). The syllable. In van der Hulst and Smith (Eds.), *The Structure of Phonological Representations, Part 2.* Dordrecht: Floris.

Steels, L. (1997). Synthesising the origins of language and meaning using co-evolution, self-organisation and level formation. In Hurford, Knight, Studdert-Kennedy (Eds.), *Approaches to the Evolution of Language*, (pp. 384-404). Cambridge: Cambridge Univ. Press.

Svelmoe, G. and Svelmoe, T. (1990). *Mansaka Dictionary.* Dallas: SIL, Inc.

Sylestine, C., Hardy, H.K., Montler, T. (1993). *Dictionary of the Alabama Language.* Austin: UT Press.

Takahashi, M. (1953). *Romanized Japanese-English Dictionary.* Kobe: Taiseido.

Vennemann, T. (1972). On the theory of syllabic phonology. *Linguistic Berichte,* 18, 1–18.

Willerman, R. (1994). *The phonetics of pronouns: articulatory bases of markedness.* Ph.D. dissertation, Linguistics, University of Texas, Austin.

Zipf, G.K. (1935/1968). *The Psycho-Biology of Language.* Boston: MIT Press.