Why Adult Language Learning is Harder: A Computational Model of the Consequences of Cultural Selection for Learnability.

Robert N. Nelson Jr. (rnnelson@purdue.edu)
Department of English, Purdue University
West Lafayette, IN. 47906

Abstract

This paper reports on a limited model of language evolution that incorporates transmission noise and errorful learning as sources of variation. The model illustrates how the adaptation of language to the statistical learning mechanisms of infants may be a factor in the apparent ceiling on adult second language achievement. The model is limited in its focus to only phonotactics because the probabilistic imbalances that have been found in phonotactics have been found to be effective cues in the very first language learning task, speech segmentation (Saffran & Theissen, 2003; Mattys & Jusczyk, 2001), and in the organization of lexical memory (Vitevitch, Luce, Pisoni & Auer, 1999). The argument that this model supports is that these probabilistic imbalances are the result of the cultural selection of more learnable variants across generations of learners, and that this process has produced sequences that help the child learner while confounding the adult learner. The child learner is aided by specific phonotactic cues that correlate with word and syllable boundaries (e.g. the English prohibition on word initial '-ng' and Czech word-final voicelessness). These cues are often invisible or misleading to the adult learner (e.g. Broselow, Chen & Wang, 1998; Flege & MacKay, 2004), contributing to errors in both perception and production.

Keywords: Language Evolution; Critical Period Effect.

Background

Most accounts of our maturational loss of linguistic adaptability have focused on age-correlated changes in the learner. This report focuses on the object of learning, language, and asks why language is harder for adults to learn than for infants. The hypothesis that this project advances is that the adaptation of language to children is a factor in the age-associated decline in language learning achievement. To this end, a model is presented which shows that, by integrating errors that occur during use and learning, an artificial language that is initially only constrained by articulatory considerations develops a more fine grained and informative structure whose distributional characteristics are highly similar to both natural language text and speech samples. Importantly, the model also shows that networks, when trained on one of the artificial languages evolved through this process and then tested on another, display the kind of cue blocking effect that held by Ellis (2006) to be the source of 'fragility' in second language learning.

Unlike other models of the cultural selection of language, such as Kirby (1996), which are concerned with the emergence of grammatical universals, this model is limited to an account of the structural imbalances observable in language phonotactics. Like Kirby (1996), however, this model holds that adaptation occurs through iterations of cultural transmission. To explore this hypothesis, an evolutionary model of errorful learning with noisy intergenerational transmission was developed. While ease of articulation and discrimination are the primary forces that shape phonotactics, this model assumes that for any two or more sequences that require near equal articulatory and discriminatory effort, the one that is most 'learnable' will be selected.

All languages exhibit regular phonotactic patterning. While some of this structural imbalance is physiological in origin, much of it is not, and constraints like vowel harmony in Finnish and Turkish or Czech word-final voicelessness, which cannot be ascribed to articulatory pressures, must be implicit knowledge that is culturally transmitted. Also not due to articulatory constraint are non-absolute, probabilistic imbalances such as the pairing `/uf/`[1] which occurs in English much more frequently than would be predicted by the independent probabilities of either `/u/` or `/f/` (Kessler & Treiman, 1997). Certain imbalances, like the prohibition on obstruents in Mandarin syllable codas, are believed to influence both the production and perception of second languages (e.g. Broselow, Chen & Wang, 1998).

The model presented here assumes that any phonotactic sequence that is easier to learn than a competitor sequence will be selected for representation in lexical memory. And, while such selection is certainly influenced by the independent probabilities of the phonemes, true selective advantage is the predictive efficiency of phones or sequences of phones. This model shows that phonotactic imbalances may be the result of generations of language learners selecting the most learnable forms of a language, and that these patterns (1) make language easier for children to learn, (2) are partially responsible for the commonly observed differences in outcomes between child and adult learning, and (3) account for some of the distributional characteristics of phonemes in natural languages.

The Model

This is not a model of language evolution, but rather the change through adaptation of a part of language. As Dell, Reed, Adams, and Meyer (2000) note, all languages have patterns at many different levels, and all of these patterns

[1] As in 'stuff'.

are available and informative to the learner. It should be reasonable to assume, then, that to the degree that languages evolve at all, all of these different levels of systematicity may be following individual (albeit mutually constraining) trajectories through an evolutionary 'design space' (Dennett, 1995; Eigen & Winkler-Oswatitsch, 1992).

The Population

The population that was exposed to learnability-based selection was comprised of words from the lexicon of an artificial language. The corpora showed structure at three levels. They were composed of 100-120 simple sentences ('NVN'). The lexicon was accordingly divided into two classes: N (67 words), and V (33 words). A pseudo-random number generator assigned words to slots in the sentences. Words were strings of 3 to 11 'phones', and were generated from a list of 15 phones that were each composed of 13 features. There were 10 C (consonant) class phones and 5 V (vowel) class phones. In accordance with the observation that words in natural languages are built from syllables that tend to adhere to an onset-rhyme-coda structure, a CV or CVC alternating structure was imposed on them. A finite state transducer (FST) generated 100 strings of a form such that no more than two phones from the C (consonant) class could occur before a phone from the V (vowel) class in any string. Importantly, in any C or V position, any C- or V-class phone occurred with equal probability in the initial lexicon. The phones were composed of thirteen features and thus represented on a thirteen dimensional vector. Each place on the vector stood for a real feature, so, for example, the 'phone' /k/ was composed of the features +velar, +stop, and –voice:

{0, 0, 0, 0, **1, 1,** 0, **0,** 0, 0, 0, 0, 0} → /k/

While the phone /a/ was composed of the features +voice, +back, and +low:

{0, 0, 0, 0, 0, 0, 0, **1**, 0, **1,** 0, 0, **1**} → /a/

Fifteen phones were generated for the initial 'phone inventory', but during the operation of the model the number of phones was extended by mutation to 21.

The Landscape

The landscape that shaped the population through learnability-based selection was sequence learning. The ability of Simple Recurrent Networks (SRN) to approximate a Bayesian analysis allowed them to function as the landscape here. Predictive dependencies between phonemes in the stream of speech has been shown to be extremely valuable for infants in the process of speech segmentation (Saffran & Theissen, 2003; Gómez & Gerken, 2000), and SRNs have been shown to model this data nicely (Elman, 1990; Christiansen, Allen & Seidenberg, 1998;).

Sources of Variation

Elements of culture can accrue modifications during use, and these modifications can be transmitted to later generations that improve the fitness of the enculturated individual with respect to the environment or the fitness of the culture with respect to the learner (Sperber, 1996, identifies this as an epidemiological feature of cultural variants). Ease of learning has been proposed as a selective feature in language evolution by William Labov (1994), Morten Christiansen (1994), Terrence Deacon (1997), and Simon Kirby (1996). This model is different from those above, however, in that it identifies sources of variation ('mutations') that have strong correlates in the 'real world' use of language. The sources of variation operationalized in this model are:

1. *Random bit-switching.* Random bit-switching here represents types of variation that may enter a language through inter-learner phenomena: contact between languages, dialects and/or idiolects. These mutations tend to be divergent, in that they increase the variety of forms available to the learner.
2. *Integration of error.* Integration of error is meant to operationalize types of variation that may enter a language through intra-learner phenomena like the realization of the past tense of 'bring' as 'brang' on analogy with other present/past pairs. These errors tend to reduce the amount of variety available to the learner.

One reason to consider random mutation in a model of language change is that all transmission of information implies noise. And it thus seems reasonable to adapt Shannon & Weaver's (1949) original description—the stochastic 'flipping' of a bit as it passes through a noisy channel. This may seem to be an impossible simplification of the human situation, which is concerned with the transmission of cultural meaning through language, but it fits because when comprehending an utterance we select or construct the meanings of messages based on a sparse signal sent by an interlocutor through a noisy environment.

Random bit-switching involved the switching of two positions on the feature vector if a random number between 0 and 1 was greater than a set threshold (0.99)[2]. For example, if the module took in the stop /k/ and then generated the number .991, /k/ could become the fricative /x/ with the switching of the bits in the sixth and seventh positions:

a) {0, 0, 0, 0, 1, <u>1, 0,</u> 0, 0, 0, 0, 0, 0} → /k/
b) {0, 0, 0, 0, 1, <u>0, 1,</u> 0, 0, 0, 0, 0, 0} → /x/

The 'mutated' version of the phone would then replace one instance of the original in the corpus. After training, network performance would be evaluated on both the original and the mutated form. If the network performed better on the mutated form, it was passed into the lexicon, replacing the previous form. Thus, for example, the word /kekt/ became /kext/ in the sixth generation of the 'L2' run of the model. The model was limited to four types of switch so that only pronounceable phonemes were produced. Each bit-switch was meant to represent exposure to a variant as the result of contact with a specific 'dialect' containing that variant:

1. +PL → +FRIC. This mutation replaced stops with homorganic fricatives.
2. +FRIC → +PL. This mutation replaced fricatives with homorganic stops.

[2] All random numbers were generated with the Mathematica Random function, which uses the time of day at start-up as seed.

3. +FRNT → +BCK. This mutation moved the vocalic place of articulation.

4. +BCK → +FRNT. This mutation moved the place of articulation forward.

The second method of introducing variation involved the inclusion of a network's mistakes at one generation into the corpus for the following generation. Since the task of the network was predictive auto-association, the input and teaching patterns were identical. However, the actual output was always close to the target (of the same C or V class, and matching the +/-voice feature), but seldom 'correct'[3]. Integration of error was meant to instantiate types of variation that may enter a language through intra-learner phenomena. These errors in learning are realized as overgeneralizations, not unlike the production of the past tense of 'bring' as 'brang' on analogy with present/past pairs like 'sing : sang'. The specific method was straightforward: after learning was stopped, a test corpus was presented to the network. Every vector produced by the network in response was compared by its cosine with every possible target phone. The phone which showed the highest cosine with the actual output was considered to be the actual output of the model. With a very low average probability ($p<=0.01$), these highest cosine phones were substituted into a temporary lexicon that could be tested against the original lexicon. Importantly, the level of network error (MSE) was interpreted as its confidence level and, accordingly, where error was highest, the threshold for mutation was lowered. So, while the average likelihood of any variant being passed over to the lexicon was 0.01, the likelihood for the few phones showing the highest MSE was as high as 0.1. This operationalizes the observation that the less well practiced a skill is, the more likely mistakes are to occur. The temporary lexicon was then tested against training lexicon and the winning patterns were passed onto the next generation. Integration of error produced the most winning variants, with 62% being passed on, versus 42% of the random bit-switch variants.

While the idea of variation has been important in the evolutionary modeling of language (cf. Kirby, 1996), there has not been much discussion of the role of mutation in language evolution. This may partly be due to the sense of deliberateness in linguistic innovation that would seem to make the models Lamarckian. Indeed, the difference between human and animal culture may be that human cultural is deliberately cumulative with respect to modifications (e.g. 'the ratchet effect', Tomasello, 1999). Edelman (1992) in fact holds that all cultural evolution is Lamarckian. Dennett (1995), however, has disagreed, noting that this charge assumes the viewpoint of the holder of the variant. A more rational viewpoint would be that of the variant, since it is the one 'struggling' for survival and unable to directly affect its own fitness.

Nearly as important as the source of variation is the rate of mutation. There are two ways in which evolutionary change in this model is gated. First, by setting a parameter that constrains language change by limiting the number of variants that can pass into the corpora. This parameter is intended to reflect the way that communicative requirements constrain language change in vivo. The second way that change is modulated is through the interaction of the two types of mutation (random bit-switching and integration of error). Integration of error is compressive in that it tends to collapse the language onto the most frequent phones and n-grams. Random bit-switching brakes this compressive effect by providing low-frequency, high-information patterns that are resistant to change.

Results

The results are presented in three sections. The first two describe tests of two artificial languages that were generated by the same FST ('L1' and 'L2') and then subjected to 20 generations of learnability-based selection. The third section examines the distributional characteristics of the 'evolved' corpora.

Section One The first measure compared the performance of 10 pairs of SRNs (13-40-13) trained on both the 0^{th} and 20^{th} generation of the 'L1' lexicon. The results are summarized in figure 1. The cosine averages on the y-axis are for each respective entire corpus, and reflect the average similarity of the network output to the target of learning.

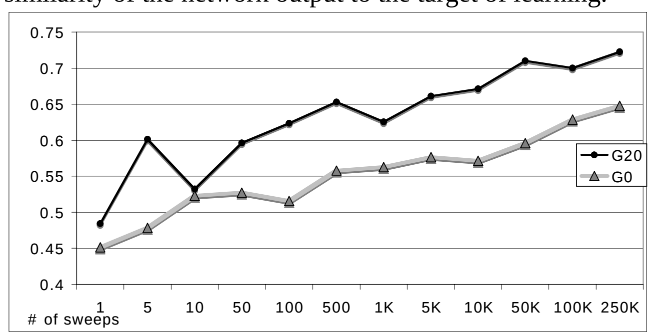

Figure 1.

This figure shows that 20 generations of selection for learnability had created, in the lexicon, enough distributional, structural imbalance to increase the average cosine of the network output and the target by 14.2%. The exact nature of the structural imbalance will be discussed below, but it is generally due to the efficient 'chunking' of initially random characters into high frequency bi- and tri-grams (i.e. emergence of phonotactic patterns) and the predictive value of low frequency phones (especially those recruited through mutations).

Section Two For the second measure, networks were trained on a corpus generated from L1G0 and L1G20. Then they were evaluated on their ability to process two new corpora: one generated from the L1G20 lexicon, and one generated from the L2G20 lexicon (i.e. the 'L2' lexicon that had undergone 20 generations of learnability based selection). The results are summarized in Figures 2 and 3. Note that, in Figure 1, the two data points at each number of sweeps represent performance by the same network on two

[3] An output was correct if its cosine with the actual target was higher than its cosine with any other possible target.

different data sets, whereas data points in Figures 2 and 3 at the corresponding pairs represent performance of two different networks. Figure 2 shows that training networks on L1G0 does not increase or decrease their ability to predict sequences in L2G0. Figure 3 shows that that networks trained on the L1G20 found sequences from the L2G20 lexicon to be equally predictable in the early stages of learning (the first 50 sweeps), but not so in later stages.

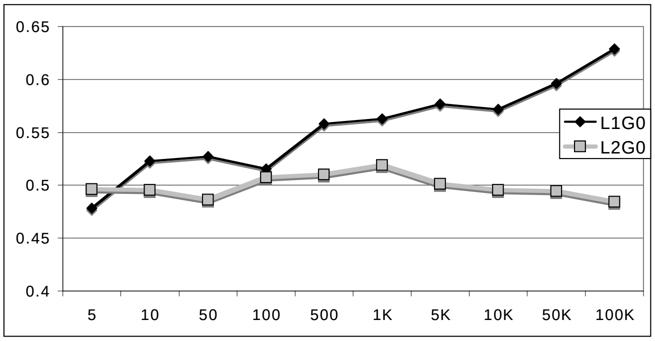

Figure 2.

The y axis in figures 2 and 3 shows the mean cosine of the network output with the target, while the x axis shows the number of sweeps through the training data.

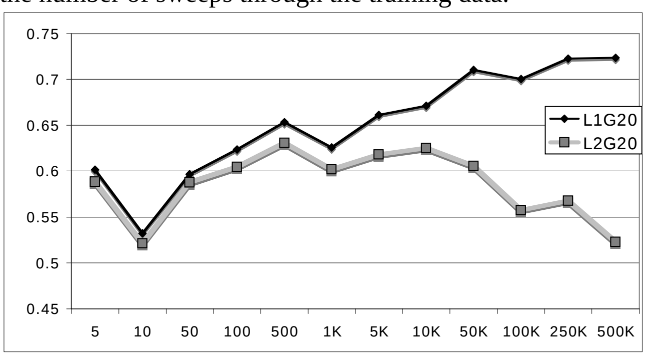

Figure 3.

Figure 3 shows that networks trained on L1 sequences lose sensitivity to the underlying similarities of the L1 and L2 lexica while gaining sensitivity to finer-grained dependencies in the L1. However, whether these patterns of network learning are meaningful depends partially on how they hold over natural languages. Figure 4 shows the results of training similar networks (18-40-18 SRNs) on transcriptions of data from the Childes database (MacWhinney, 2000). Networks were trained on the Berenstein (English) data and then tested on the Miyata-Aki (Japanese) data.

Early in training, networks track each other because of (a) the similar underlying structure of both languages (both Ls are composed of CV and/or CCV sequences), and (b) since half of all C-class phones are +voice while all of the V-class phones are +voice, a good early 'hypothesis' for these networks is to always guess +voice (thus being right 66% of the time). However, as learning proceeds from the 50th sweep, networks become sensitive to multiply correlated weak cues in L1G20. These are statistical cues that, on their own, possess very weak predictive value but become strong predictors of patterns in the data when correlated with each other. At 10,000 sweeps, networks begin to rely predominantly on these systems of weak correlations, thus performance on L2G20 and the Japanese data falls of significantly in their respective groups of networks.

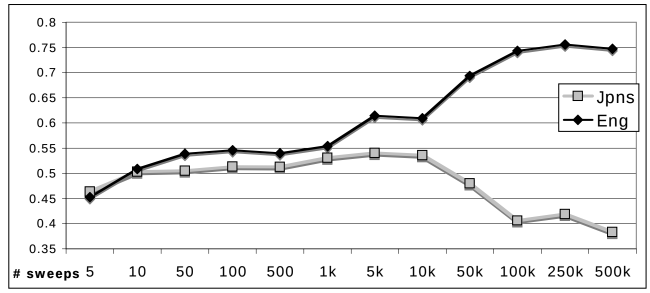

Figure 4.

Section Three Cultural evolution of the type modeled here is driven by a 'rich-get-richer' effect that exploits statistical variations in (1) the initial random distribution of features and (2) the early rounds of mutation/selection. This type of effect has received a lot of attention recently as a statistical mechanism behind the topography of Scale-Free Networks (e.g. Albert & Barabasi, 2002). However, in the context of this model, it can be best understood as an instance of the Polya's Urn contagion model[4]. If the mechanism posited in this model, integration of transmission and learning errors, operates in real world, the distributions that are the product of the model should look like the distributions we see in real life.

Figure 5 shows a log-log plot of the frequency of characters in three 100 sentence corpora alongside a frequency count of phonemes from a 750,000 phoneme corpus of spoken British (RP) English (Fry, 1947). The second data set ('TS') is composed of the first 100 sentences of the second chapter of Mark Twain's (1876) 'Tom Sawyer'. The third data set ('G20') was generated from the L1G20 lexicon, the fourth data set is the G0 corpus. The process embodied by this model thus produces corpora whose distributions are highly similar to both natural language text and speech samples.

In any word produced by the L1G0 FSA, all characters are equally likely and the most compressed representation of any word in the L1G0 lexicon would be the word itself. However, by the 20th generation, the lexicon includes several different recurring sequences. These sequences reduce the entropy of the lexicon so any representation

[4] Notionally, Polya's urn asks us to imagine a game played with an urn that is filled with n red pebbles and n black pebbles. The game has one rule: if you first draw a red pebble, you have to put it back and then also replace one of the black pebbles with a red one (or vice-versa if a black pebble is selected first). The pebbles are then remixed. It is now a little more likely that you will draw a red pebble the next time. Eventually, the urn will be filled with either only red or only black pebbles, the other color having been driven out.

(magnetic, neural, etc.) of the L1G20 lexicon requires less information than a representation of the G0 lexicon

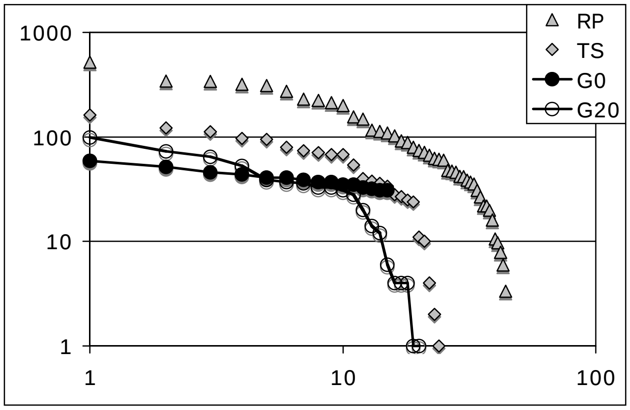

Figure 5.

. Consider the by-order decline in entropy illustrated in Table 1. This figure shows that, while the imbalance in frequencies of individual characters does reduce entropy, the greatest gain of informativeness in the lexicon is through character dependencies (2nd order entropy).

Table 1: Entropy of L1G20

	0th Order	1st Order	2nd Order
G0	4.3	3.88	3.23
G20	4.3	3.76	2.4

Adaptation in this model decreases entropy, and thus the amount of information a learner needs to construct a language, by creating high frequency 'chunks' that can become routinized, and lower frequency chunks that are highly informative. Low frequency phones are created in two ways in the model. The first way is by driving their numbers down though gradual replacement (the 'rich-get-richer' effect). The second is when mutations recruit novel combinations of features from the space of possible phones. The resulting distribution of high and low frequency phones and sequences of phones increase the learnability of the lexica. High frequency 'chunks' are not very predictive, yet occur so frequently that they are easy to predict (and thus produce low error in the networks). On the other hand, very low frequency characters that have high 'suprisal' (like 'q-' in English orthography) are difficult to predict (i.e. produce high error) but, once they occur, are reliably predictive of the subsequent phone.

Recall that the mutation rate increased when the network was less confident of its output (i.e. the rate was sensitive to network error signal). As a result of this, positionally constrained sequences of phones (like /N/ in English) emerged under the error-integration condition as a response to the high prediction error associated with word boundaries. Figure 7 shows the average relative probabilities for the V- and C-class phones for each possible phone position in every word. The graph shows that most of the category changes, from V-class to C-class and vice versa, occurred in the first and last three positions. These class changes led to the emergence of a VV sequence, [ou], which occurred in the first and second phone positions of five words by 20th generation.

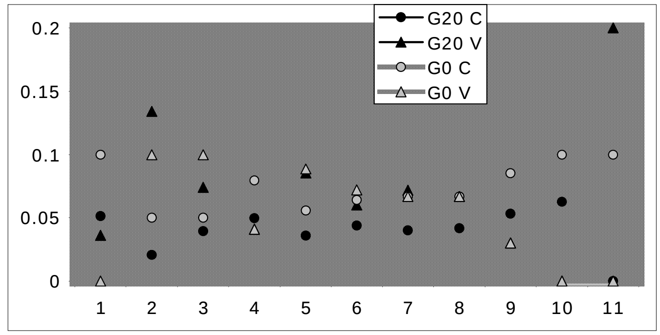

Figure7.

Discussion

This simple model evinces an evolutionary process that is very similar to the Polya's Urn contagion model. The results described in section one show that this simple winner-take-all process increases the number of recurring sequences in the lexicon. This in turn increased the ability of the networks to predict items in these sequences. Section two showed that change through iterations of learning produced lexica that were broadly similar but specifically highly dissimilar. The broad similarities meant that networks' initial hypotheses about the data generalized well from one set to the other, while the specific dissimilarities meant that continued learning of the subtle cue correlations of the L1 data produced a weight set that was less able to derive useful information from the L2 set. In a sense, continued learning increased the overshadowing of cues and produced a blocking (of associative learning) effect (Kamin, 1969; Rescorla & Wagner, 1972), which is hypothesized by Ellis (2006) to be the source of 'fragility' in second language learning.

This occurs because the evolutionary process embodied by the model produces distributions of features that are similar across languages. For example, the number of voiced consonants in all artificial languages evolved in this model is increased by 12 to 14.5% by the 20th generation. This means that early learning about general features of the distributional nature of the language will generalize well from language to the next. However, while some features may be common across languages, their particular combinations are not. As learning continues and networks become sensitive to multiply correlated cues, their ability to generalize thus drops and the systems of cue correlations learned for the L1 becomes maladaptive in the face of an L2.

This interpretation is in concert with a number of developmental studies (e.g. Dell et al., 2000; Mattys & Jusczyk, 2001), but especially Coady & Aslin, (2004) and Iverson, et al., (2003). Coady & Aslin, (2004) found that older children, but not younger ones, were sensitive to "fine-grained acoustic-phonetic information in the developing lexicon" and that this sensitivity continued to develop over time while Iverson et al. (2003) found that early language learning experience altered low level perceptual processes. This acquired insensitivity partially grounds the critical period effect as a predictable outcome of learning in parallel processing, representationally distributed, sub-symbolic networks. It is also consistent with findings from the Second Language Acquisition literature, in which L1 phonotactics is shown to be sound predictor of difficulties with the production and perception of L2 phonology (Broselow, Chen & Wang, 1998; Flege & MacKay, 2004).

In sum, cultural evolution by selection for learnability increases the learnability of evolved lexica while producing a 'critical period effect' in agents that learn them. Results also show that the accumulation of adaptations results in a lexicon that is rich in probabilistic information (i.e. entropy is minimized), which, in turn, predicts the type of distribution of phonemes and graphemes in natural languages, as well as the same type of by-order decline in entropy found in English by Shannon & Weaver (1949). Finally, the model also predicts the emergence of phontactics sensu stricto (typical sequences that have typical positions) as a response to increased uncertainty at word boundaries.

References

Albert, R., & Barabasi, A. L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics, 74*(1), 47-97.

Carpenter, G. A., Cohen, M. A., & Grossberg, S. (1987). Computing with Neural Networks. *Science, 235*(4793), 1226-1227.

Coady, J. A., & Aslin, R. N. (2003). Phonological neighbourhoods in the developing lexicon. *Journal of Child Language, 30*, 441-469.

Broselow, E., Chen, S.-I., & Wang, C. (1998). The Emergence of the Unmarked in Second Language Phonology *Studies in Second Language Acquisition 20*, 261-280.

Christiansen, M. H. (1994). *Infinite Languages, Finite Minds: Connectionism, Learning and Linguistic Structure.* Unpublished PhD Dissertation, University of Edinburgh, Scotland.

Deacon, T. W. (1997). *The Symbolic Species: The Co-evolution of Language and the Brain*: W.W. Norton.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: a study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26*(6), 1355-1367.

Dennett, D. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon & Schuster.

Edelman, G. M. (1992). *Brilliant Air, Brilliant Fire: On the Matter of the Mind*. New York: Basic Books.

Fry, D. B. (1947). The frequency of occurrence of speech sounds in Southern English. *Archives Néerlandaises de Phonétique Expérimentale, 20*.

Gaygen, D. (1997). *The effects of probabilistic phonotactics on the segmentation of conituous speech.* State University of New York at Buffalo, Buffalo, NY.

Eigen, M., & Winkler-Oswatitsch, R. (1992). *Steps towards life : a perspective on evolution* (P. Woolley, Trans.). New York: Oxford University Press.

Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science, 14*(1), 179-211.

Flege, J. E., & MacKay, I. R. A. (2004). Percieving vowels in a second language. *Studies in Second Language Acquisition, 26*(1), 1-34.

Gomez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Science, 5*(4).

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Ketterman, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*.

Kamin, L. J. (1969). Selective association and conditioning. In S. Scarr, S. W. Scarr & L. J. Kamin (Eds.), *Fundamental issues in associative learning*. Halifax, Nova Scotia: Dalhousie University Press.

Kauffman, S. A. (1993). *The origins of order : self organization and selection in evolution*. New York: Oxford University Press.

Kessler, B., & Treiman, R. (1997). Syllable Structure and the Distribution of Phonemes in English Syllables. *Journal of Memory and Language, 37*(3), 295-311.

Kirby, S. (1996). *Function, Selection and Innateness: the Emergence of Language Universals.* University of Edinburgh, Scotland.

Labov, W. (1994). *Principles of linguistic change. Vol. 1: Internal factors*: Blackwell Ltd. Oxford.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk. Third Edition.*: Mahwah, NJ: Lawrence Erlbaum Associates.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning:The effectiveness of reinforcement and non-reinforcement. In R. A. Rescorla, A. R. Wagner, A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory*. New Haven, CT.: Yale University Press.

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology, 39*(3), 484-494.

Shannon, C. E. & Weaver, W. (1949). *The Mathematical Theory of Communication*. University of Illinois Press. Urbana.

Sperber, D. (1996). *Explaining Culture: A Naturalistic Approach*: Blackwell, Ltd. Oxford.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Boston, MA.: Harvard University Press.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language, 68*(1-2), 306-311.