A CONNECTIONIST SIMULATION OF V-TO-'V PHONOTACTICS LEARNING

Edson Françozo[†], Orlando Bisacchi Coelho[‡], Eleonora Cavalcante Albano[†], Laudino Roces[†], Renato Basso[†] and Pablo Arantes[†]

† UNICAMP – State University of Campinas, Brazil

‡ UMC – Universidade de Mogi das Cruzes & UNICAMP – State University of Campinas, Brazil

E-mail: edson@iel.unicamp.br, orlando@umc.br, albano@iel.unicamp.br

ABSTRACT

Our work hypothesis is that phonotactics is best described by non-categorical, probabilistic biases, embodied in the lexicon as constraints on lexical forms which are emergent properties of the operation of dynamical systems that shape language behaviour. In order to investigate if these biases are robust enough to help a learner infer the phonotactic constraints of a language, we developed a connectionist model of an aspect of vowel-to-vowel phonotactics: harmony and contour, namely the tendency for vowels to share or avoid repetition of phonic properties. A Simple Recurrent Network was trained to predict the next phone in the word. No phonetic information was supplied to the network. The network was able to learn a significant amount of phonetic knowledge only from the distributional biases embodied in the lexicon. For example, the model was able to correctly differentiate between vowels and consonants and to discriminate between the dynamics of contour and harmony.

1. INTRODUCTION

According to a recent trend in the phonetics/phonology literature, phonotactic patterns are best described by non-categorical constraints on lexical forms [1-3]. Our research group has been addressing these questions via statistical description of the distributional biases in phone sequences in a lexical database [4-5].

We hypothesize that he distributional biases we have found are the result of a process, which is evolving in time, by which the lexicon adapts to phonotactic preferences which are partially guided by the nature of human speech production and perception apparatus. Human languages would tend to prefer what is easier to produce and easier to distinguish (even if these constraints sometimes happen to point in opposing directions). Of course these would not be the only forces guiding lexical evolution; language contact, for instance, also plays a part. But it does not seem far-fetched to suppose that ease of production and ease of perception play an important role in lexical evolution. In fact, this is coherent with the widely accepted view that language has a biological basis.

A statistical description of a phenomenon is very helpful in characterising the forces that govern it. Nevertheless, when the phenomenon of interest is essentially dynamical, a dynamic approach can be even more enlightening. Phenomena that have a dynamical nature can be statistically described. But they can be best described by a dynamical model: a model whose behaviour is described and analysed in terms of the mathematical tools provided by the theory of dynamical systems [6]. Behaviour that can, in a first analysis, be understood as random can be correctly characterized as deterministic when analysed with the tools of dynamics. Data that can be classified as belonging to a single class can be best described as belonging to distinct classes, if the analysis takes time into consideration [7]. Accordingly, many researchers believe that in order to fully understand cognition in general one has to adopt a dynamical point of view [8-9]. Some of the main factors that play a role in perception and production of speech are of a dynamical nature [10]. That justifies applying the dynamical approach to cognition to research in phonetics and phonology.

In order to address such research issues we decided to develop a dynamical connectionist model for an aspect of the phonotactics of Brazilian Portuguese that was identified as relevant in the distributional study [4]: intervocalic harmony and contour. An interesting phonotactic feature, which seems to hold in several languages, is known as OCP (the obligatory contour principle): the preference for a certain level of contrast between consecutive vowels in a word. Contour can be described in terms of differences in opening (or tongue height), place (or tongue position), and rounding between such vowels. Harmony, on the other hand, is the absence of contrast between consecutive vowels.

The connectionist simulation reported below focuses on a representative subset of the intervocalic relationships that occur in Brazilian Portuguese: the V(C).'C(C)V phone strings present in trisyllabic nouns. The questions we pose and try to answer via the connectionist simulation are:

- Is phonetic information encoded by distributional biases in the lexicon?
- Are phonotactic constraints robust enough to help a learner infer the phonic pattern of a language?

 ISBN 1-876346-48-5

2. EXPERIMENT AND CURRENT RESULTS

Harmony and contour are constraints that hold between consecutive vowels in a word. For that reason, they are dynamic phenomena that can only be studied through a model that can focus on the transitions between vowels as they evolve in time. An SRN [11] – a discrete time recurrent connectionist network – is such a model. The SRN is a variant of the multi-layer feed-forward network where recurrent connections feed to each unity in the hidden layer a copy of its own value. Trained with the backpropagation algorithm this network is able to capture temporal information that is present in the training set.

In the simulation reported here a modified SRN was used. In order to lower the number of weights to be trained, a compression layer was added between the input and the hidden layers (Figure 1). The network was trained to predict the next phone in the word. The training set consisted of penultimate stressed trisyllabic nouns only. The 3,700 V(C).'C(C)V phone strings in the training set were fed to the network phone by phone. The context layer activation was reset after presentation of the phones corresponding to each word.

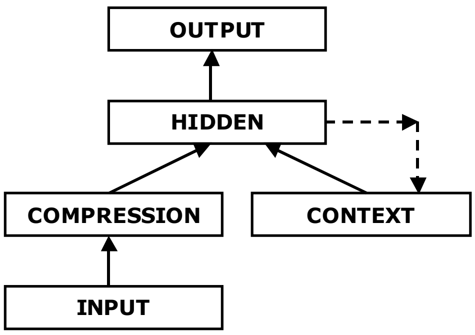


Figure 1: Architecture of the network used in the simulation.

Phones were encoded as 35 orthogonal vectors. So no phonetic information at all was supplied to the network, not even information concerning syllable boundaries. Since the focus of the experiment was on the V-to-V transitions, testing was performed by presenting the network with 12 non-words [pV.'pV.pV], spanning the whole set of pre-stressed vowels [i, e, a, o, u] and stressed vowels [i, e, E, a, O, o, u] in Brazilian Portuguese, thus controlling for the effects of consonants.

After training, in order to investigate what the network had learned, the hidden unit activations produced for each phone in the test set were analysed. A hierarchical cluster of the hidden unit activation (Figure 2) showed that, even with no featural phonetic information being input to the network, it was able to acquire a significant amount of phonetic knowledge only from the distributional biases embodied in the lexicon. For example, the network was able:

- to distinguish vowels from consonants and
- to differentiate between pre-stressed and stressed vowels[1].

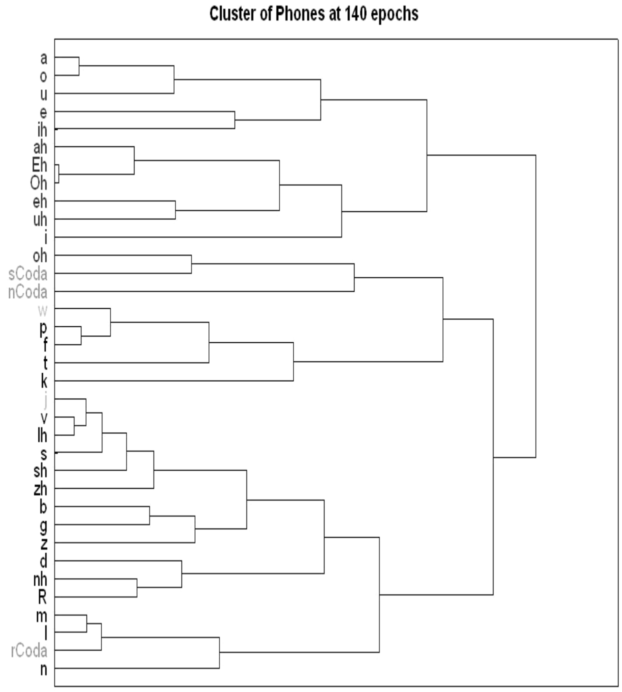


Figure 2: Hierarchical cluster of the hidden unit activations based on the entire set of letters in the training set. The codes used in the figure are[2]: double vowels represent open vowels; an *H* after a vowel codes for a stressed vowel; and an *H* following a consonant means that the consonant is at the end of the word.

The hidden unit activations were also analysed using the PCA method. Means for the 5 harmonic and 7 contour non-words [pV.'pV.pV] in the training set were taken. Such means were tested for significance with ANOVA (Table 1).

[1] Nevertheless, the network was not as successful in clustering consonants.

[2] It was not possible to use IPA fonts due to limitations of the statistical software used to produce the graphs.

PC	Harmony		Contour	
	V1	V2	V1	V2
4	0,3756	0,4096	0,4222[b]	0,2347[b]
8	0,1694[c]	0,1436[c]	0,1934[a]	0,4112[a]

Table 1: Significance of PCA means according to ANOVA. Significance: a= p>0.001; b=p>0.04; c=p>0.07 (marginal). V1: pre-stressed vowel; V2: stressed vowel.

It is possible to identify dimensions in the principal component space that discriminate for Harmony and Contour. For example, Figure 3 depicts the transition between vowels along the dimensions coded by the 4^{th}. and 8^{th}. principal components. It can be seen that harmonic vowel pairs are not strongly distinguished in any dimension, a fact that is coherent with the distributional study [5]. In contrast, contour vowel pairs differ significantly in both dimensions.

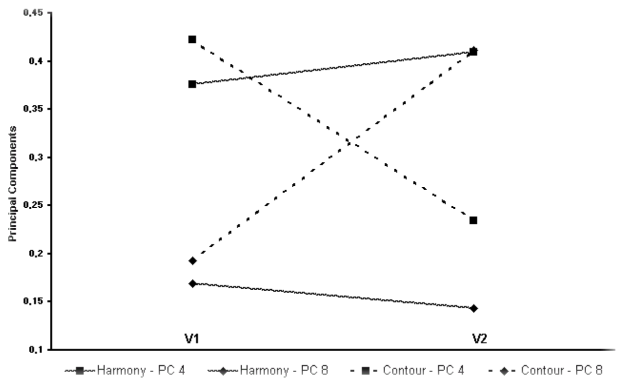


Figure 3: V1 to V2 Transitions according to PCs 4 and 8.

Also, the trajectories associated to transitions in the subspace spawn by the 4^{th} and 8^{th} components (Figure 4) show that the dynamics for harmony is opposite to the one for contour, as can be seen from the opposite directions of the respective $\overrightarrow{V_1 V_2}$ vectors (where V_i measures the mean, over the harmonic or contour instances in the training set, for the i^{th} vowel in the subspace). Therefore, PCs 4 and 8 capture together a distinct dynamic behaviour for harmony and contour. What amounts to saying that, in order to perform the task, the hidden units have to adapt in such a way that those principal components of their activations become able to encode relevant aspects of the vowel-to-vowel phonotactics of Brazilian Portuguese, namely the discrimination between harmony and contour.

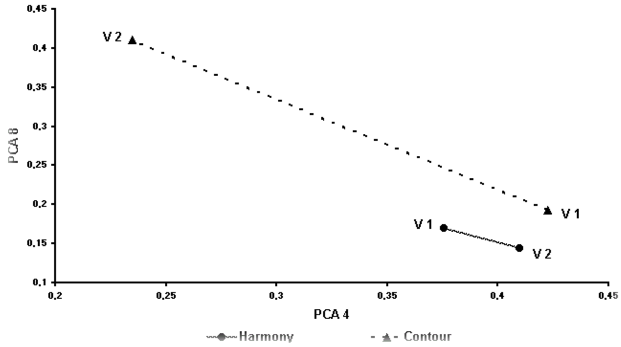


Figure 4: Harmony vs. Contour dynamics in PCs 4 and 8. The direction of the $\overrightarrow{V_1 V_2}$ vectors is opposed for vowel pairs that differ in harmony and contour.

The network´s ability to extract from the lexicon – and encode in its hidden units – information that distinguishes harmony from contour is quite robust. For example, the trajectories associated to the intervocalic transitions, as represented in the subspace spawn by the 5^{th}. and 10^{th}. principal components (Figure 5), obtained when the networks processes non-words where /p/ is the consonant, display opposing dynamics for harmony and contour.

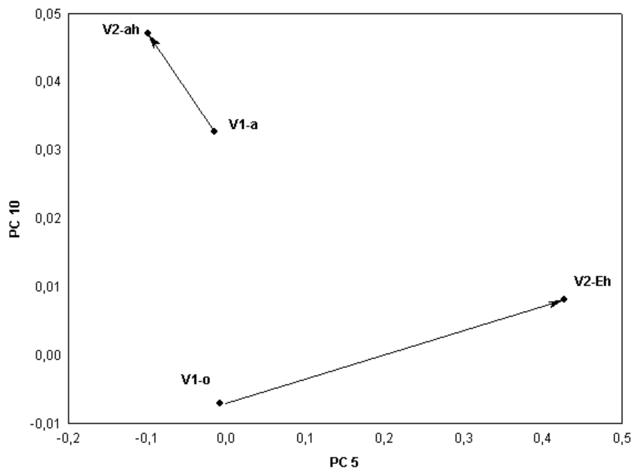
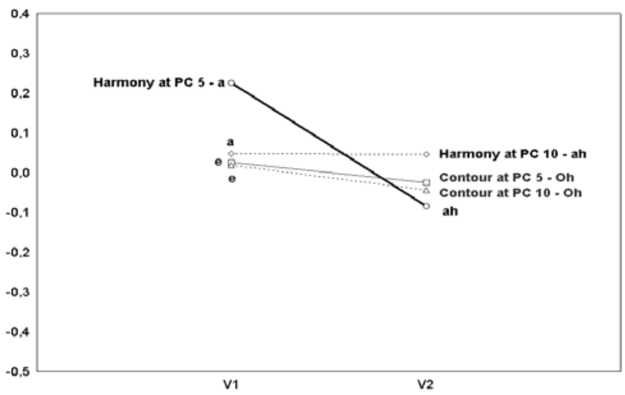


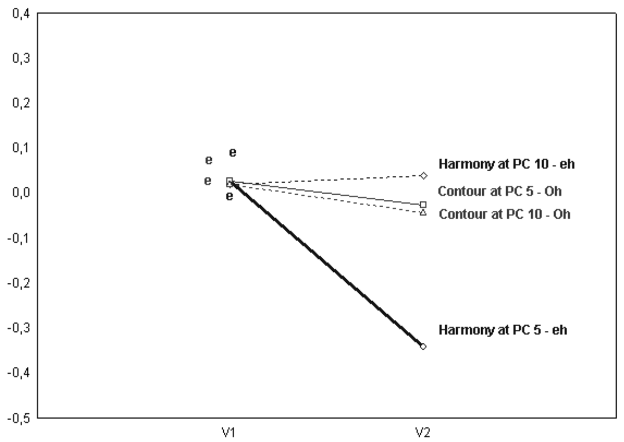
Figure 5: Dynamics for harmony and contour in PCs 5 and 10 for /p/-non-words.

PC 5 coding for harmony is also robust (Figure 6). Both the inclination and direction of the line segments that code for harmony in PC5 are the same in Figures 6a and 6b, although each figure depicts the dynamics for a different pair of vowels: /a/-/e/ in Figure 6a versus /e/-/o/ in Figure 6b.

ISBN 1-876346-48-5 © 2003 UAB



(a)



(b)

Figure 6: Intervocalic transitions according to PCs 5 and 10 for the pair of words *fachada-devoto* (a) and *defesa-devoto* (b).

3. DISCUSSION

Phonotactics is best described by non-categorical, probabilistic biases, embodied in the lexicon as constraints on lexical forms. Those biases are emergent properties of the operation of a set of dynamical systems that shape language behaviour. So, understanding phonotactics, as much as understanding cognition in general, depends on adopting a dynamical point of view. The current work also lends support to the view that the surface structure of language has information that is rich enough to allow for a dynamic connectionist network, coupled to a proper learning algorithm, to extract the information needed to perform many linguistic tasks. This work hypothesis is shared by a number of researchers that believe that such a view sheds new light on the nature-nurture debate [12]. In that framework it is not unexpected at all that there is phonotactic information embodied in the lexicon via distributional biases, since lexicon construction and phonotactics crystallization can be best understood as coevolving processes.

ACKNOWLEDGMENTS

Research funded by FAPESP (01/00136-2).

REFERENCES

[1] J. Pierrehumbert, “Syllable structure and word structure: a study of three consonantal clusters in English”, in P. Keating (Ed.) *Papers in laboratory phonology III: Phonological structure and phonetic form*, pp. 168-190. Cambridge: Cambridge University Press, 1994.

[2] S. Frisch, “Temporally organized lexical representations as phonological units”, in M. Broe & J. Pierrehumbert (Eds.) *Papers in laboratory phonology V: Acquisition and the lexicon*, pp. 283-298. Cambridge: Cambridge University Press, 2000.

[3] R. Treiman, B. Kessler, S. Knewasser, R. Tincoff and M. Bowman, “English speakers’ sensitivity to phonotactic patterns”, in M. Broe & J. Pierrehumbert (Eds.) *Papers in laboratory phonology V: Acquisition and the lexicon*, pp. 269-282. Cambridge: Cambridge University Press, 2000.

[4] E. Albano, *O gesto e suas bordas*. Campinas: Mercado de Letras, 2001.

[5] E. Albano, E., “V-to-V phonotactics, syllable structure, and morphological productivity”. *Proceedings of the Eighth Conference on Laboratory Phonology*, 2002.

[6] M. Giunti, *Dynamical Models of Cognition*, in R.F. Port and T. van Gelder (Eds.) Mind as Motion, Cambridge: The MIT Press, 1995.

[7] J. C. Principe, N. R. Euliano and W. C. Lefebvre, *Neural and Adaptive Systems: Fundamentals Through Simulations*, pp. 375-376. New York: John Wiley & Sons, 2000.

[8] T. Van Gelder, A connectionist variation on a classical theme. *Cognitive Science*, vol. 14, pp. 355-384, 1990.

[9] R. P. Port, and T. van Gelder, (Eds.) *Mind as motion*. Cambridge, MA: The MIT Press, 1995.

[10] C. P. Browman and L. Goldstein, *Dynamics and Articulatory Phonology*, in R.F. Port and T. van Gelder (Eds.) Mind as Motion, Cambridge: The MIT Press, 1995.

[11] J. L. Elman, “Language as a dynamical system”, in R. F. Port & T. van Gelder (Eds.), *Mind as motion*. Cambridge, MA: The MIT Press, 1995.

[12] J. L. Elman, E. A. Bates, M. H. Johnson, A. Karmiloff-Smith, D. Parisi and K. Plunkett, *Rethinking innateness*. Cambridge, MA: The MIT Press, 1996.