A spreading activation model of word production and lexicon evolution that might produce more minimal pairs for less confusable contrasts

Russell Richie

Introduction

Questions of why a particular animal -- say a giraffe -- possesses some feature -- its long neck -- are often answered by appeal to *function* -- because the long neck allows the giraffe to eat leaves in the tops of trees. This is a *functionalist* explanation. As in biology, many phenomena of human language, from syntax to phonology, have seen functionalist explanations (Haspelmath, 1999; Bybee, 1999, 2006). For example, it has been argued that one reason SVO word order predominates (~45% of world languages, a plurality) is that it is robust to noise in communication: if the subject or the object is dropped or misheard, the other arguments syntactic role can still be recovered (cf. SOV order; Gibson et al., 2012). Similarly, it has been shown that the meanings of words[[1]](#footnote-1) tend to maximize within-category similarity, and minimize between-category similarity (Regier et al., 2007). Despite the intuitive appeal of such accounts, merely identifying a way in which a particular structure or phenomenon *is* functional does not suffice as an explanation of that structure’s origin. Instead, one must show *how* structure evolved under pressure for satisfaction of some functional constraint. In this paper, I take up this call, and attempt to show how a recently identified functional characteristic of human lexicons might have arisen.

This characteristic is the following: that human lexicons tend to have more minimal pairs for more perceptible contrasts (Graff, 2012). That is, minimal pairs like ‘bought’ and ‘taught’, which differ by the reliable contrast /b/ vs /t/, are more common than minimal pairs like ‘fought’ and ‘thought’, which differ by the highly confusable /f/ and /θ/. Such a pattern is functional because it increases the overall discriminability of words in the lexicon. However, as pointed out previously, this does not entail that human lexicons have this property *because* of its communicative benefits. There may be mechanisms of lexicon evolution that produce the minimal pair ~ perceptible contrast phenomenon that have no communicative element, and hence no way for perceptibility to exert an influence on lexicon evolution. In fact, Graff (2012) himself may have identified such a mechanism.

The model of Graff (2012)

Graff (2012) described a spreading activation model of word production that could account for related phenomena of lexical contrast maintenance (of which his finding that perceptible contrasts have more minimal pairs is an example). In this model, there is a layer for phonemes (e.g., a node for /c/, a node for /a/, a node for /t/), a layer for word-forms (e.g., a node for ‘cat’), and a layer for concepts (e.g., a node for ‘furry four-legged feline’). To speak, the model activates a concept, and activation spreads to the corresponding words at the word layer, and then to phonemes corresponding to these words. Additionally, activation can flow back from the phoneme layer to the word layer, and the words in the word layer laterally inhibit each other, such that highly active words strongly inhibit all other words. To simulate lexicon evolution, every generation, randomly generated synonyms for existing concepts are added and linked up to the network at the beginning of the generation, agents randomly choose concepts to express and activation consequently flows throughout the network, and, finally, the synonyms with the weakest activation after activation of their respective concept are discarded at the end of each generation. Simulations of 1000 generations like this generated lexicons with low average similarity between words. Simulations without lateral inhibition between words, however, generated high average similarity between words. Why does lateral inhibition have this effect? When a target word that is similar to many other words is activated, the target word activates its phonemes, which activate these many other, similar words in the lexicon, which then laterally inhibit the target word, reducing its activation, and making it more likely to be discarded at the end of a generation.

Graff (2012)’s model is useful for understanding the general architecture of a possible lexical contrast mechanism, although it does not explain his specific finding that perceptible contrasts have more minimal pairs, as all the phonemes in his model are essentially the same – they are just numbered nodes. One way to simulate Graff’s finding, then, might be to add a distinctive feature layer beyond the phoneme layer, and allow feedback from the feature layer to the phoneme layer. Confusable sounds tend to have similar articulations: the highly confusable /f/ and /θ/ are both [-voice] and [+continuant]; /m/ and /n/ are both [+voice] and [+nasal] (see Figure 1). If a language had a confusable minimal pair like ‘ram’ and ‘ran’, then activation of a word like ‘ram’ would feed all the way down to the feature layer, and back up to minimal pairs like ‘ran’ that share features, inhibiting ‘ram’ and possibly leading to its death in the language. But if a language had a non-confusable minimal pair like ‘ram’ and ‘rack’, then activation of ‘ram’ should not activate ‘rack’ as much, leading to less inhibition of ‘ram’ from ‘rack’, and the continued coexistence of that minimal pair. Exploring a model like this will form the rest of this paper.

Notice that the mechanism of lexicon evolution in Graff’s model is *purely speaker-driven*. That is, lexical contrast maintenance in his account is just driven by the word production process. This should come as somewhat surprising given that Graff claims that his minimal pair finding is evidence of *communicative* *efficiency*, as the model has no communication in it, no interaction between a speaker and a listener. If the minimal pair finding were really evidence of *optimization* of communicative efficiency per se, then we should expect the mechanism of emergence to involve something like listeners misunderstanding and consequently rejecting minimal pairs more often for more confusable sound contrasts.

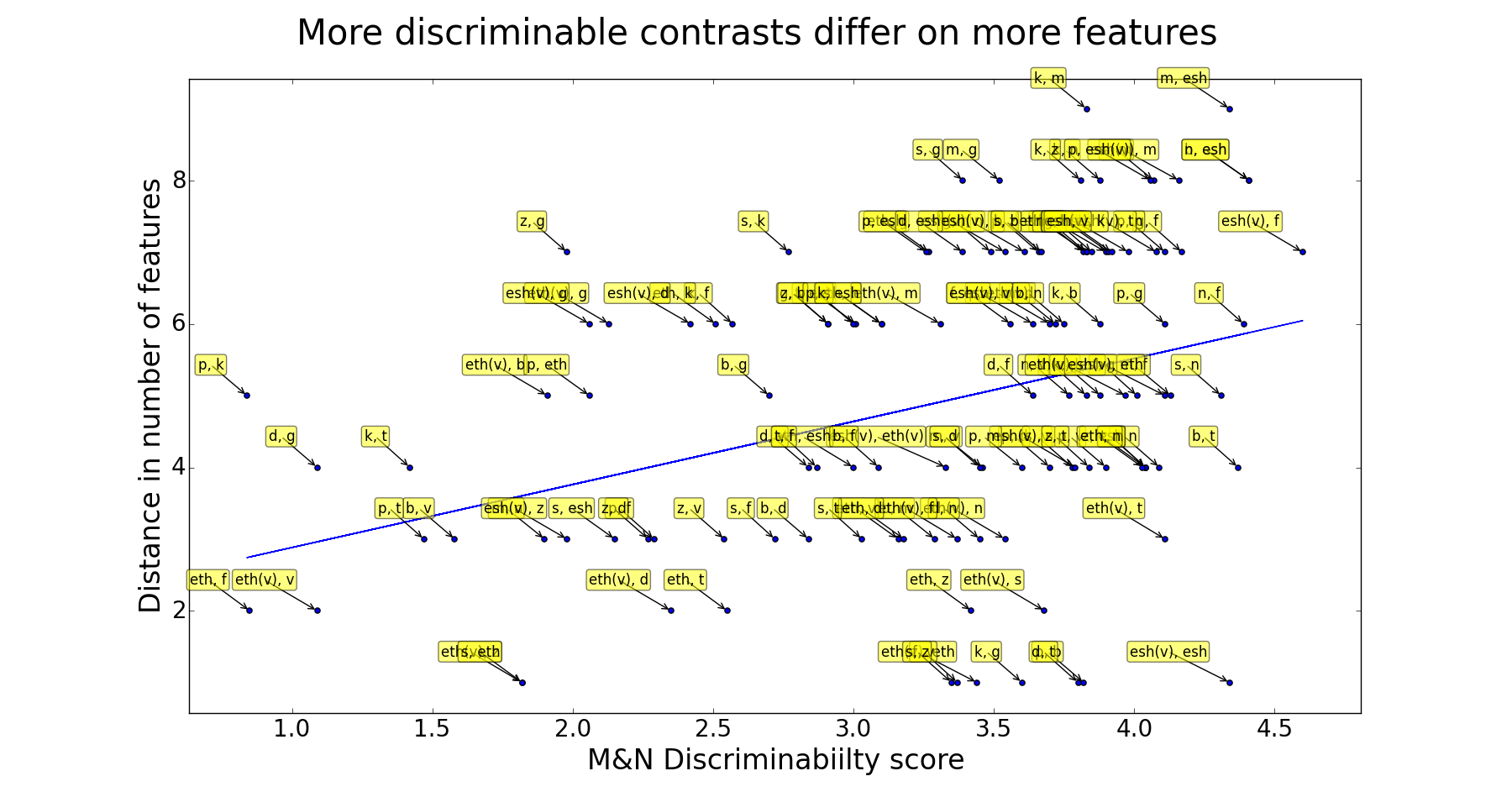
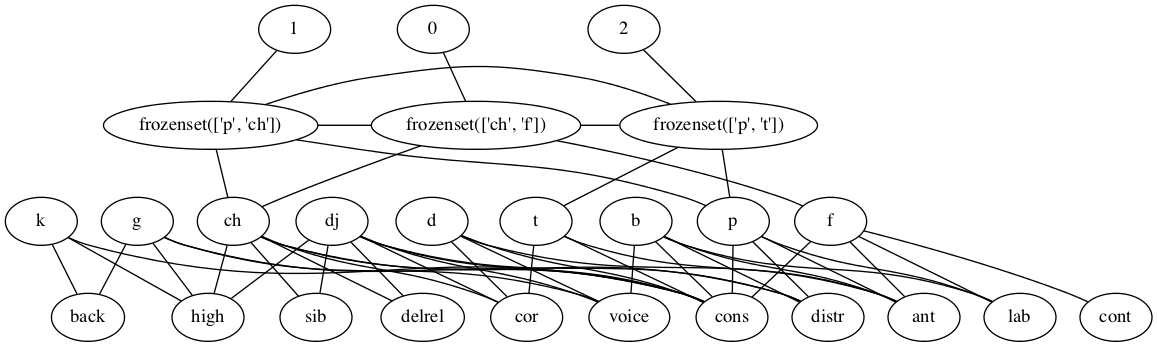


Figure 1. On the x-axis is a contrast’s discriminability score, taken from Miller and Nicely (1955). On the y-axis is the number of features on which two phonemes of a contrast differ. Taken from here. More perceptible (discriminable) contrasts tend to share fewer features (*r* (118) = .36, *p* < 10-4). Identically, more confusable contrasts tend to share *more* features.

We now describe in some detail an extension to Graff (2012)’s model that may explain his minimal pair findings.

The current model

The architecture of the model is pictured below in Figure 2. As in Graff (2012)’s model, there is a layer for concepts, a layer for word-forms, and a layer for phonemes. There are bidirectional, excitatory connections between all adjacent levels, and bidirectional, inhibitory connections between words. Added to the model in this work is a layer for distinctive features (Mannell, no date). This layer has 15 features, including features for voicing, nasality, various places of articulation (e.g., coronal, labial, back, high), and various manners of articulation (e.g., delayed release, + for affricates and – for fricatives). The Mannell (no date) typology also determined the set of 24 English phonemes to be included.

 Figure 2. The general architecture of the current model. Concepts, at the top, are simply numbered 0 to *n*, the number of concepts. Below concepts are word-forms, below them, phonemes, and below them, features. This particular network uses a limited number of concepts, words, phonemes, and features for ease of display.

Some further notes on these additions are required. First, minimal pairs require some notion of linearization/sequentiality in word-form representations, so that one can say that a particular minimal pair differs only on the *first* phoneme. Right now, neither Graff (2012)’s model nor the present model possess such linearization. Instead, word-forms are randomly generated by sampling phonemes without replacement, and treated as sets. Thus, ‘bap’ would be a minimal pair with ‘pad’. This does not solve the issue of linearization, however, because there is still a single layer of distinctive features for all phonemes. Thus, the ‘b’ and ‘p’ of ‘bap’ both feed down to a single ‘labial’ feature. Thus, the word-form information of ‘bap’ is somewhat lost down at the featural layer. Any future modifications of this model might instead have non-overlapping sets of phonemes for onsets, nuclei, and codas (as in the original spreading activation model of word production by Dell, 1986), and corresponding non-overlapping sets of features for these phonemes, so that word-form information is faithfully represented at the featural layer.

Second, the current simulations are run with, as stated, all 15 features and 24 phonemes from the Mannell (no date) data, 5 phonemes per word, and 10 concepts/words in the lexicon. With so few words in the lexicon, and so many possibilities for word-forms, minimal pairs are extremely unlikely[[2]](#footnote-2). As a result, it is hard to assess whether the model produces Graff (2012)’s minimal pair finding. An analysis analogous to average word-form similarity by phonemes, whereby average word similarity at the *featural* level, could be conducted. Note that, due to the correlational analysis presented above, more minimal pairs for more discriminable contrasts translates to more minimal pairs for contrasts differing by many distinctive features. A pressure for more minimal pairs like this is a special case of a pressure for word-forms that are more distinctive on their featural specifications *overall*, i.e., on all their phonemes, not just on those single phonemes on which they differ. Hence, if the model produces words that tend to be dissimilar not just at the phoneme level, but also at the featural level, then it should produce Graff (2012)’s minimal pair finding, as well. Unfortunately, however, this analysis requires word-forms to have sequential, phonemic representation rather than the mere set-based phonemic representation. So, any real test of Graff (2012)’s minimal pair finding must await future work. Instead, the present model can lay some groundwork for such work.

Activation flows from node to node according to the rule from Dell (1986). During each time step *ti*, the activation level of node *j*, *A*(*j, ti*), is calculated with the following equation:

*A( j, ti ) = wj* [ *A*( *j, ti-1* ) *+* ](*1− q*) *+ noise*

In this equation, *c*1*, c*2 *... cn* are all of the nodes connected to *j*, *p*1*, p*2 *... pn* are the weights of the connections, and *q* is the decay rate. In Graff (2012) and the simulations presented here, *p* was set to 0.3 for excitatory connections and -0.3) for inhibitory connections, and *q* was set to 0.6. The noise added to the activation level is a value sampled from a Gaussian distribution with a mean of 0 and a standard deviation equal to 0.05 times the node’s previous activation level *A*(*j, ti-1*).

We now briefly discuss some preliminary simulations of this extended model.

Results

Parameters for the simulations (Table 1) described are identical to those of Graff (2012), except those parameters concerning our additions (italicized). With these parameter values, we do not appear to be replicating Graff (2012)’s finding that the average phonemic similarity between words drops over time. Figure 3 shows that, to the contrary, average phonemic similarity between words stays more or less constant over all 1000 generations, barring temporary fluctuations.

There are a couple other curious findings of the extended model with the above parameters. First, when a concept is activated, neither synonym ever really dominates the other – each tends to have a roughly comparable level of activation at every timestep. This of course seems unrealistic – during word production, perhaps initially each synonym is highly and comparably active, but we do eventually select one, implying that its activation. Similarly, and possibly relatedly, the longterm equilibrium following activation of a concept is not strong activation of *one* synonym and inhibition of all other words, but rather roughly equal activation of all words centered around 0. In other words, the model appears subject to a kind of ‘heat death’ in the longrun. It is less clear that this is problematic, because the words linked to the activated synonym *do* have the strongest activation initially, and only die out in the longrun. Perhaps this is psychologically realistic – we initially think of a concept to express, then the corresponding word is uttered, and then the system returns to a baseline (0) level of activation.

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| Number of concepts | 10 |
| Lengths of words | 5 |
| *Number of Phonemes* | *25* |
| *Number of Features* | *15* |
| Number of Generations | 1000 |
| Initial activation (of concept node) | 100 |
| Node weights | 1 |
| Edge weights | +/- 0.3 |
| Decay | 0.6 |
| Settling Time (number of timesteps activation is allowed to flow) | 5 |

Table 1. Parameter values for present simulations.

Discussion

Future work is needed with this model to really test its ability to explain Graff (2012)’s minimal pair finding. First, the model must replicate Graff (2012)’s findings. It is not clear at the moment whether this failure is due to the addition of the featural level, or of some other difference between Graff’s model and my own that has escaped my attention. Second, the bugs in the simulation noted in footnote 2 must be worked out to allow the model to produce more genuine minimal pairs. This will allow a more appropriate test of the current hypothesis regarding the origin of Graff’s minimal pairs finding. Third, and most importantly, the model needs to introduce some sequentiality into the phoneme representation, so that real minimal pairs can emerge and Graff’s finding can truly be tested. All these adjustments will need to be made so that the functional origins of Graff (2012)’s minimal pairs finding can truly be tested.

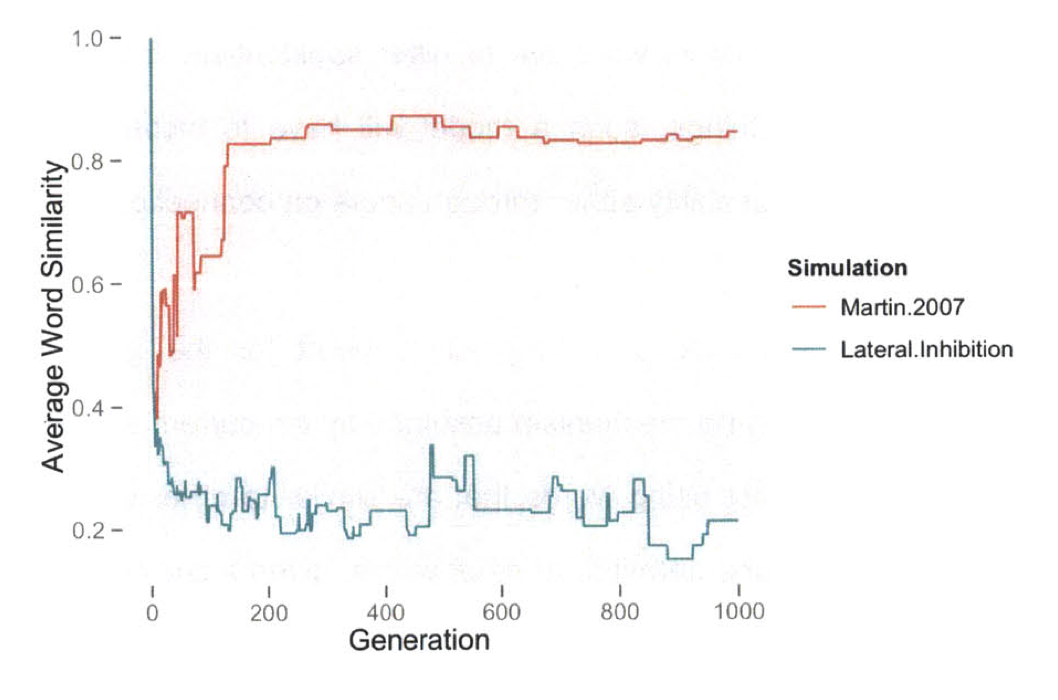
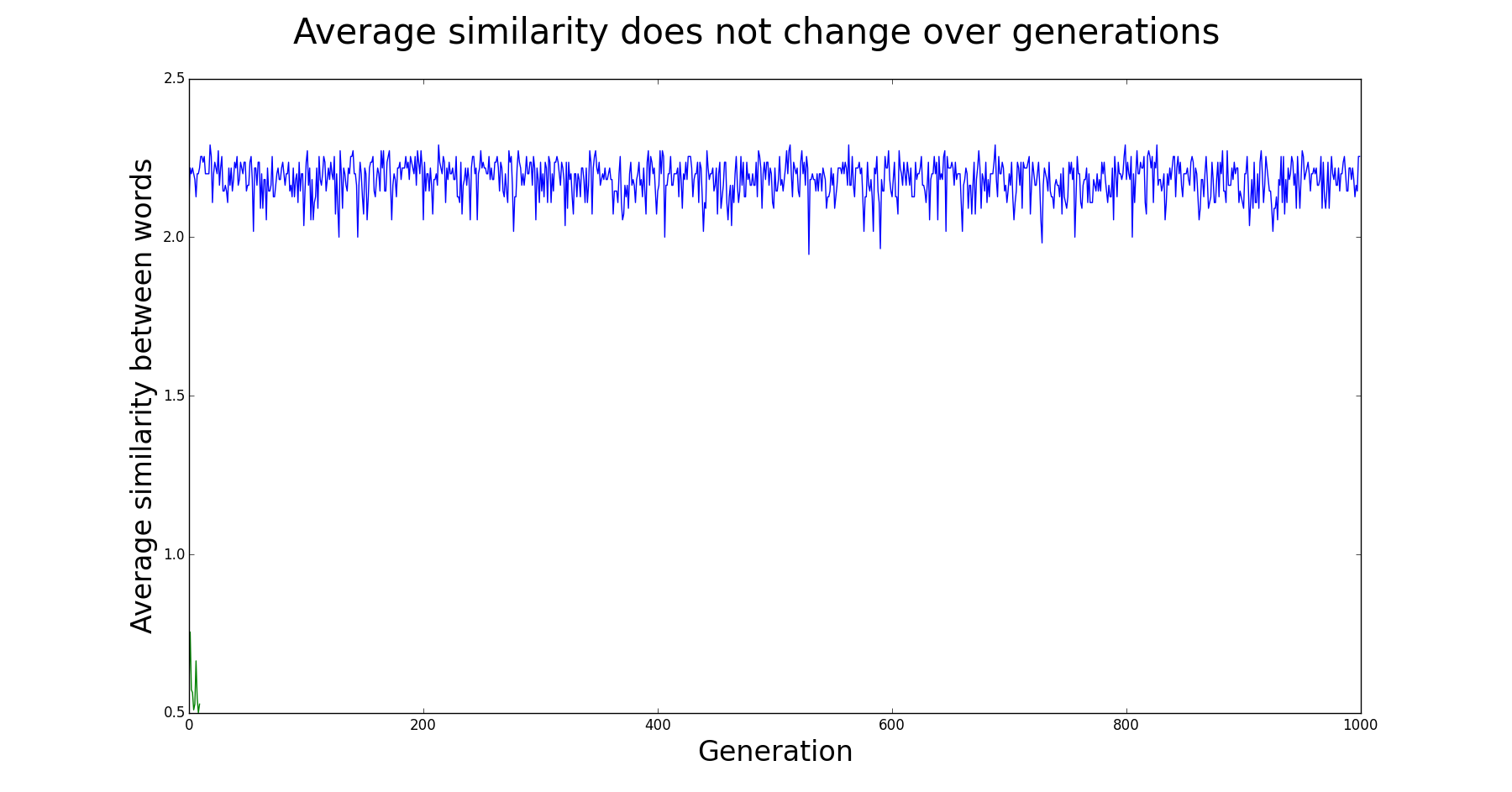


Figure 3. Average similarity over time. Contrary to Graff (2012)’s results (bottom, blue line), the current implementation with the addition of a feature layer (top) does not lead to increased average dissimilarity between words.

References

Dell, G.S. (1986). A spreading activation theory of retrieval and sentence production. *Psychological Review, 93*, 283-321.

Gibson, E., Piantadosi, S.T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2012). A noisy-channel account of crosslinguistic word order variation. *Psychological Science*, *24*, 1079-1088.

Mannell, R. (no date). Phonetics and Phonology: Distinctive Features. Retrieved from http://clas.mq.edu.au/speech/phonetics/phonology/features/

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(4), 1436–41. doi:10.1073/pnas.0610341104

1. Whether this is a property of all concepts or just those picked out by words is not important here. [↑](#footnote-ref-1)
2. The reason we did not conduct simulations with fewer phonemes and features is practical: they produce some bug in the simulation that causes it to crash. I believe the bug has something to do with the portion of the program that deletes synonyms after activation flow is complete, but I have not completely figured it out. [↑](#footnote-ref-2)