Papers in Historical Phonology
http://journals.ed.ac.uk/pihph
ISSN 2399-6714
Volume 1, 1-4
DOI: 10.1010/10101010101010101010

Modelling frequency effects in sound change chains

Stefano Coretta
The University of Manchester

1 Methods

Four simulations were implemented in R (R Core Team 2017) to test the behaviour of the model described in ... The simulation starts with the creation of a lexicon of 100 lexical items. 50 words are assigned to the BART category, while the remaining 50 to the BAT category. Within each category, lexical frequency from 1 to 50 is assigned to each lexical item, so that there are a BART and a BAT word of frequency 1, a BART and a BAT word of frequency 2, etc. The lexicon is then populated using the specified formant values (6.5 and 5.5 Bark in this simulation). A loop produces 50,000 total tokens among the lexical items. The selection of the items is randomised and the weight of each item depends on its lexical frequency. Higher frequency items (words) have thus more tokens in their representation than lower frequency words.

After the lexicon is initialised, a simulated sound change is applied to such lexicon. The simulation goes through 200,000 iterations, in each of which a random lexical item is selected (lexical frequency weighing applies) and produced. To produce the selected lexical item, a random target among the exemplars is chosen and produced with a random level of noise (here between -0.2 and +0.2). If the selected word is a word containing the vowel which is undergoing change (the biased vowel), then a bias of -0.3 is applied to the formant value (on top of the noise).

Now that the word has been produced, an algorithm decides weather the word will be encoded back in memory or disregarded. This will depend on its lexical frequency and most importantly on whether the token falls in an area of ambiguity, as in (Hay et al. 2015). If the token is not ambiguous, i.e. if it is surrounded only by tokens of the same category, then the probability of being encoded back in memory is set to 1 (such a word is always sent back to memory). If the token is ambiguous, i.e. if it falls in an area of overlap containing tokens of both categories, then the probability of being encoded in memory is set as a function of the lexical frequency of the relative word: p(x|M) = frequency(x) / maximum(frequency).

Once a word has been encoded in memory (or not, depending on the outcome of the encoding function), a new cycle of the algorithm begins and another word is chosen, produced and encoded (or not).

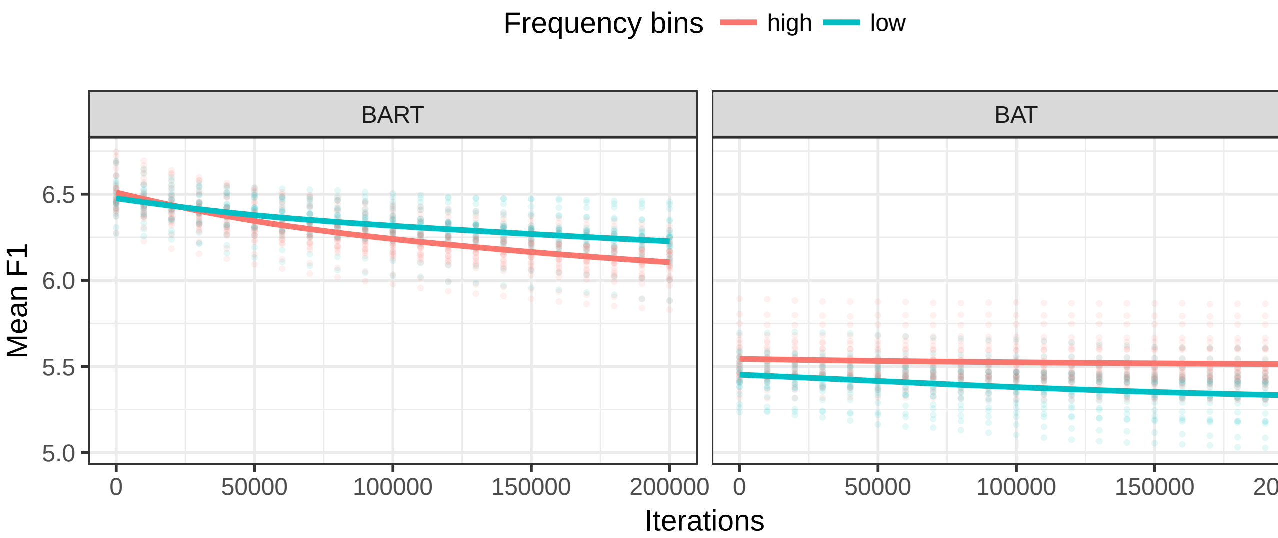
1.1 Statistical analysis

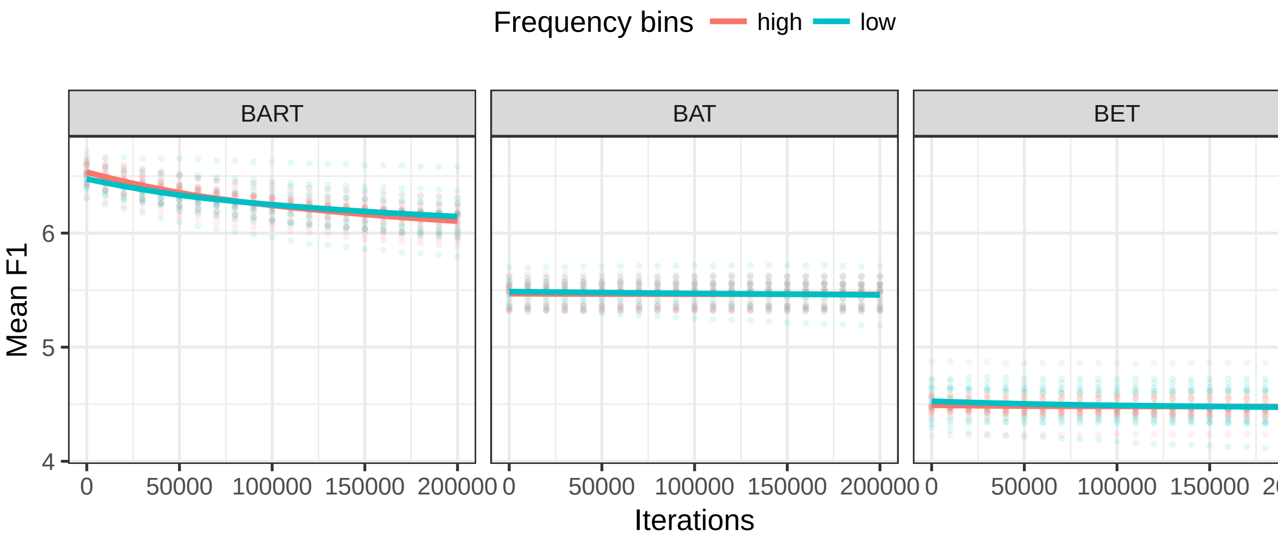
Generalised additive mixed model are used for statistical analysis.

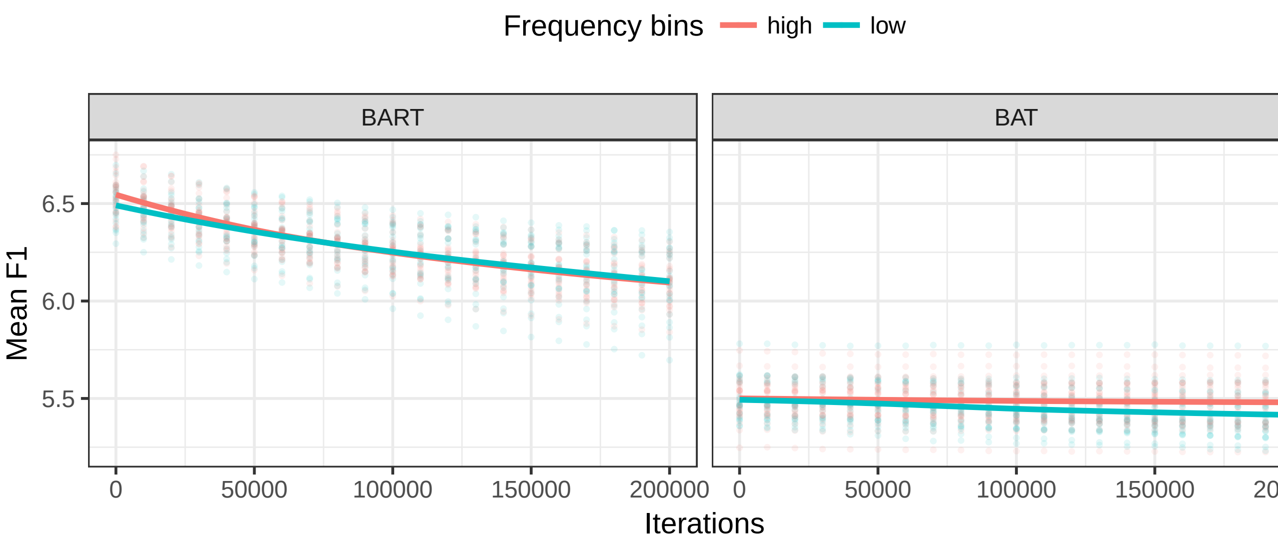
2 Results

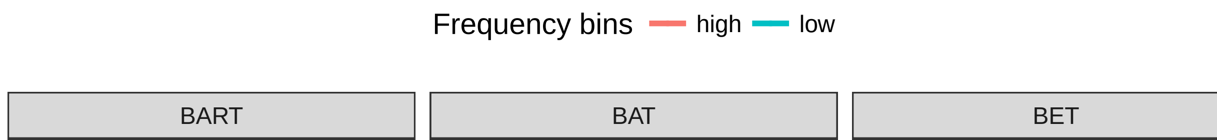
Starting from simulation 1, a general trend of decreasing mean F1 through time can be observed both in BART and BAT words, as shown in the top panel of Chapter 1. The graph also shows the decrease of mean F1 for low and high frequency words separately. The frequency bins were defined as the lower half (1–25) and upper half (26–50) of the frequency range (1–50). Mean F1 decreases at a higher rate in low frequency than in high frequency words in BAT words. The model in … thus correctly pushes low frequency words faster, as it was supposed to. Unexpectedly, though, it is high frequency words that change at a higher rate in BART words, not low frequency words. The model in its current state thus predicts that the effect of frequency on chain sound changes differs depending whether the observed phonetic category is the pushing category or the pushed category.

This pattern can be observed also in simulations 2 to 4. High frequency words change faster in the pushed category (BART), but low frequency ones do in the pushed category (BAT) or categories if there is more than one (BAT, BET). The main qualitative difference in these simulations if compared to simulation 1 is the magnitude of the effect of frequency, which


Frequency bins
high
low
BART
BAT
Mean F1
6.5
6.0
5.5
5.0
0
50000
100000
150000
200000
Iterations


Frequency bins
high
low
BART
BAT
BET
Mean F1
6
5
4
0
50000
100000
150000
200000
Iterations


Frequency bins
high
low
BART
BAT
Mean F1
6.5
6.0
5.5
0
50000
100000
150000
200000
Iterations


Frequency bins
high
low
BART
BAT
BET

is smaller. Simulations 2 and 4 also clearly show that both pushed categories (BAT, BET) undergo the effect of lexical frequency.

Author's contact details

Stefano Coretta
The University of Manchester

stefano.coretta@manchester.ac.uk

References

Hay, Jennifer B., Janet B. Pierrehumbert, Abby J. Walker & Patrick LaShell. 2015. Tracking word frequency effects through 130 years of sound change. *Cognition* 139. 83–91.

R Core Team. 2017. R: A language and environment for statistical computing. https://www.R-project.org/.