Andrea Ceolin
University of Pennsylvania

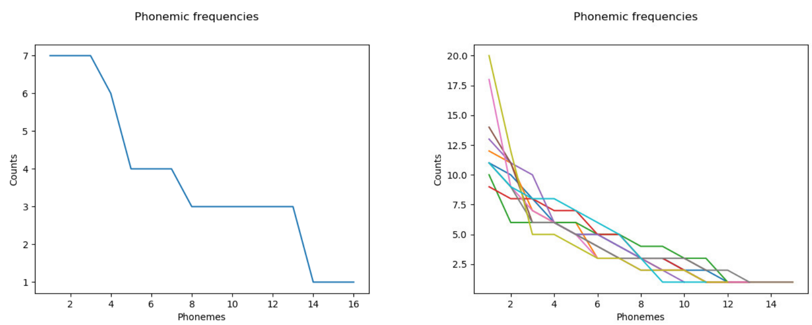
RUSE 2019

A null model of sound change

Introduction Languages do not use their sound inventories in a uniform way: each language contains phonemes which are very frequent and phonemes which are uncommon. This extends to phonemic sequences: some of them are more productive than others in the lexicon. One of the most common explanations for these facts is that these patterns result from a functional pressure towards communication efficiency. In this work, we develop a null model of sound change, to study the emergence of sound patterns in the vocabulary, and we show that these patterns can all be derived from regular sound change.

Model We use an artificial lexicon of 20 CVC words, with an alphabet that contains 10 possible vowels and 18 possible consonants. We apply three types of sound change to the lexicon: mergers, splits, and contractions (by using an epenthetic segment to keep the size of the words constant), estimating their frequency from historical data from IE, Uralic, and Altaic.

Results Tambovtsev and Martindale (2007) show that within-language phonemic frequencies are distributed according to the Yule-Simon distribution (Simon 1955). Figure 1 summarizes a first experiment using our artificial model of sound change. In Figure 1a, we plot the phonemes of the lexicon in the x-axis (represented by their rank) and their counts on the y-axis. Figure 1b shows that after the occurrence of 50 simulated sound changes (in ten parallel runs), the distribution converges to a power law distribution. This is a first example that shows how a null model of sound change can derive patterns like Yule's distribution.



(a) At start (b) After 50 sound changes

Figure 1: Phoneme frequencies

Another case study is the change of phonemic dispersion in the lexicon over time. Dautriche et al. (2017) show that natural languages favor certain phonemic frequencies over others. They do this by training a language model over a real lexicon and making the model generate pseudo-words. They show that if we simply count the number of minimal pairs, real language vocabularies display a higher amount of them than artificial vocabularies, even after controlling for factors like derivational morphology, phonotactics and word length. They interpret this result as evidence for the fact that the lexicon might be optimized for production and learnability, at the expense of perception and disambiguation. We test Dautriche et al.'s hypothesis that diachronic change is what causes languages to become clumpier over time.

The simulations exhibit the pattern in Figure 2. We see that the number of minimal pairs is indeed increasing over time. Even though on average there is an increase of minimal pairs, in several runs we see some drastic reduction of minimal pairs at certain specific points. This is associated with the creation

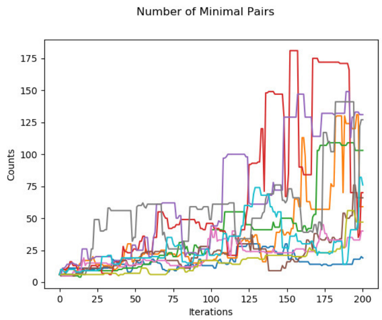


Figure 2: Minimal Pairs over time after 200 sound changes

(A)	→	(B)
sad		sad
set		sat
far		far
her		har
(A)	←	(B)

Table 1: Minimal pair creation is not always reversible

of homophones: if a merger targets a contrast which is supported by many minimal pairs, and massive homonymy is created, then the number of minimal pairs is drastically reduced.

We did not expect minimal pairs to increase over time, since sound changes with the effect of increasing the number of minimal pairs are well represented in the model. Inspecting the intermediate states of the lexicon, we saw that the reason why minimal pairs increase is due to the fact that unconditioned mergers are *irreversible*. This is clearly visible in Table 1: there are cases in which after a merger creates minimal pairs, it is not possible to go back to the previous stage, because splits are always conditioned on the environment. In this example, once the two nuclei merge and create some minimal pairs where the contrast is in the coda and others where the contrast is in the onset, it is impossible by means of a regular sound change to eliminate both of them in a single step. The opposite is not true: for every case of split, it is possible to return to the previous stage through a merger. Note that in the same way, while it is possible to create homonymy through a sound change, it is not possible to revert back to the previous stage once the merger goes to completion.

Conclusion We have shown that our model provides evidence for two claims: i) Yule's distribution can be derived from the interaction between mergers, splits and contractions over time and ii) phonemic dispersion decreases over time because of the irreversibility of mergers. From the theoretical viewpoint, these claims show that we do need to motivate the phonemic distributions in natural language vocabularies with reference to communication efficiency and functional pressures, but they can be derived from a purely stochastic model of sound change.

References Dautriche et al. (2017), Words cluster phonetically beyond phonotactic regularities. *Cognition*; Simon (1955), On a class of skew distribution functions. *Biometrika*; Tambovtsev and Martindale (2007), Phoneme frequencies follow a Yule distribution. *SKASE Journal of Theoretical Linguistic*