RUSE 2019

Booklet of Abstracts (Regular Talks)

Symposium on Representations, Usage and Social Embedding in Language Change

20–21 August 2019

MANCHESTER 1824
The University of Manchester

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

LAGB
LINGUISTICS ASSOCIATION OF GREAT BRITAIN

James Burridge, Tam Blaxter and Bert Vaux
University of Portsmouth and University of Cambridge

RUSE 2019

Testing the evolutionary paths of grammar

Abstract

We introduce a stochastic model of language change in a population of speakers who are divided into social or geographical groups. We assume that the evolution of language is driven by the inference of grammatical rules from memorised linguistic patterns. These paths of inference are controlled by an inferability matrix which can be structured to model a wide range of linguistic change processes. The extent to which speakers are able to determine the dominant grammar in their speech community is captured by a temperature-like parameter [1]. This can induce symmetry breaking phase transitions, where communities select one of two or more possible branches in the evolutionary tree of language. We use the model to test three hypotheses regarding the rise of the phrasal possessive in English and Continental North Germanic [2, 3].

References

[1] Burridge J 2017 Spatial Evolution of Human Dialects. *Physical Review X*,**7** 031008

[2] Allen C L 2008 Genitives in Early English: Typology and Evidence. *Oxford University Press*

[3] Norde M 1997 The History of the Genitive in Swedish: A Case Study in Degrammaticalization. *University of Amsterdam*

Andrea Ceolin

University of Pennsylvania

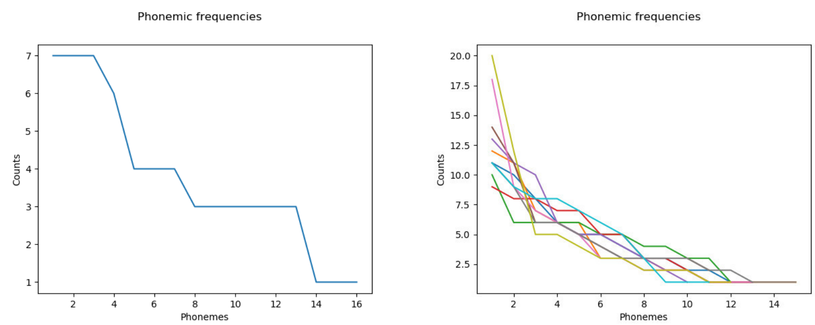
RUSE 2019

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

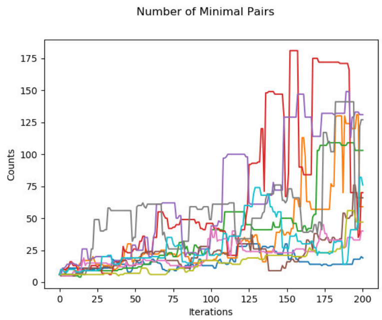


Figure 2: Minimal Pairs over time after 200 sound changes

(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*

Thais Cristofaro Silva, Christina Gomes and Daniela Guimaraes
UFMG, CNPq and UCL

Frequency effects in -ão plural formation in Brazilian Portuguese

This paper addresses the role of frequency effects in plural formation in nouns ending in -ão in Brazilian Portuguese (henceforth BP). It intends to address the role of frequency effects as suggested by Exemplar Models. Nouns that end in (ão) in the singular may have three different plural formation: (mão+s) > (mãos) ''hands'; (pão+s) > (pães) 'bread (pl)' and (leão+s) > (leões) 'lions'. The plural form (mão+s) > (mãos) would fall into the regular plural formation which adds (-s) to the nouns in the plural: (kaza+s) > (kazas) 'house'. The other two alternatives (ão) > (ães) and (ão+s) > (ões) are considered irregular plural formation. There are etymological reasons for the plural choices in ão-words, however in current language use speakers have no indication as to which form should be chosen for the plural. There is evidence that plural formation for -ão words is lexical (HUBACK 2011). As some words are more recurrent than others one expects that speakers will know the plural form for frequently used words. However, in rarely used words speakers have no clue to rely on and will have to make a choice between the three possible plural formation for -ão words. A question to be posited is which pattern is favoured in the plural formation in -ão words in BP: the general regular pattern which adds (s) to a noun or the most frequent pattern for plural formation amongst -ão words which is (ão+s) > (ões)? In order to address this question, we considered cross-sectional data from 40 children aged 3-7 years old. We designed an experiment that comprised BP nouns and nonce words. Nouns were grouped as low and high frequency ones. We posited the hypothesis that frequency effect is crucial for speakers' choice in the plural of ão-words in BP.-Results showed that the expected plural and the realized plural were realized with 76.5% for (ão+s) > (ões); 50.9% for (ão+s) > (ães) and 39.1% for (ão+s) > (ães). This shows that the most frequent pattern for plural formation amongst -ão words which is (ão+s) > (ões) presented the expected result at higher rates (76.5%) when compared with the other two options for plural formation (cf. Figure 1). We suggest that this result reflects the preference of speakers for the higher frequency count words amongst all the three options for plural formation. When plural formation is grouped by low and high frequency count one observes that low frequency words tend to receive (ão+s) > (ões) as the choice for plural (58.2%) (cf. Figure 2). We suggest that this result indicates that children prefer the pattern of plural for -ão words which is the most frequent one, i.e., which has the higher frequency: (-ões).–Finally, when the plural formation is considered for nonce words results show that the pattern (ão+s) > (ões) is the most favoured choice for plural (70.8%) (cf. Figure 3).-Our results indicate that the highest frequency count for plural formation in -ão words in BP, i.e. (ão+s) > (ões), is favoured amongst the three possible options. It is also observed that children presented the same behavior observed for adults (HUBACK 2011). When speakers know the plural form, usually for high frequency words, they tend to use the expected one. Our results show that for low frequency words and nonce words speakers tend to favour the most frequent pattern for plural formation amongst -ão words which is (ão+s) > (ões) rather than the general regular pattern which adds (s) to a noun. Our results are in accordance with Exemplar Models proposal that assume the role of frequency effects in shapping representations (BYBEE 2001, 2010; JOHNSON 1997, PIERREHUMBERT 2001). Besides providing evidence for frequency effects in plural formation in BP our results contribute to the debate on how irregular morphology is processed (MARCUS 2000; STEMBERG, MACWHINNEY 1986).

Figure 1.Plural forms per expected plural

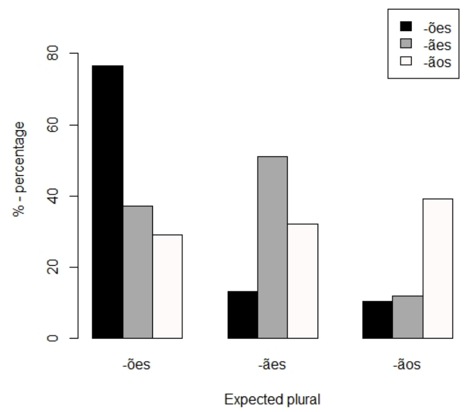


Figure 2.Plural forms per word frequency

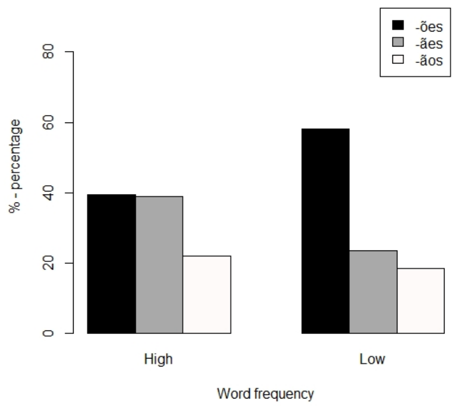
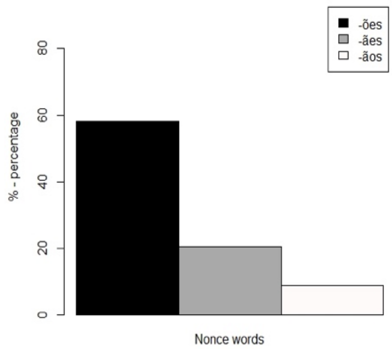


Figure 3.Plural forms in Nonce word test



References

BYBEE, J. The phonology of the lexicon: evidence from Lexical Diffusion. In: KEMMER, S.; BARLOW, M.. (Ed.) *Usage-based models of language*. Stanford California: CSLI Publications, 2000. p.65-85.

______. *Language Usage and Cognition*. Cambridge: Cambridge University Press, 2010.

HUBACK, A. P. (2011) Irregular Plurals in Brazilian Portuguese: An Exemplar Model Approach. *Language Variation and Change* (Print), v. 23, p. 1-12.

JOHNSON, K. Speech perception without speaker normalisation. In: JOHNSON, Keith; MULLENIX, John W. (Ed.) *Talker variability without in speech perception*. San Diego: Academic Press, 1997. p.145-165.

MARCUS, G. F. (2000). Children's Overregularization and Its Implications for Cognition. In P. Broeder, & J. Murre (eds). Models of Language Acquisition: Inductive and deductive approaches. Oxford: Oxford University Press, pp 154-176. [not currently available on-line]

PIERREHUMBERT, J. Exemplar dynamics: Word frequency, lenition and contrast. In: BYBEE, J.; HOPPER, P. (Ed.) *Frequency effects and the emergence of linguistic structure*. Amsterdam: John Benjamins, 2001. p. 1-19.

STEMBERG, MACWHINNEY (1986). Frequency and the lexical storage of regularly inflected forms. *MemoryandCognition*, 14 (1):17-26.

Daniel Duncan
Newcastle University

Competing representations in sociophonetic variation?

An implicit assumption in the analysis of sociophonetic variation is that speakers have one mental representation (a phoneme or set of allophones) and one phonological system (knowledge of which variant surfaces in which environments). This assumption may be seen, for example, in the typical use of linear regression analysis of acoustic data: the intercept reflects the mean production for a given baseline (the single representation), and language-internal and -external factors then have effects represented as a change from the baseline (the single phonological system). Recent empirical evidence problematizes the latter part of this assumption; Sneller (2018) suggests that speakers in Philadelphia show intraspeaker variation between two competing phonological systems with respect to the TRAP vowel (traditional short-a split vs. nasal system). If the assumption of a single phonological system does not appear to hold, we may ask whether the same is true of the assumption of a single representation. For example, could speakers have competing phonemic representations for a given vowel? This paper draws on empirical data from Greater St. Louis to suggest that this is indeed the case for some speakers.

I test this question by examining intraspeaker variation in production of the LOT vowel among 52 white middle-class females, born 1896-1992. Approximately 15 minutes of informal conversation with the women were transcribed and forced-aligned with FAVE (Rosenfelder et al. 2014), and vowel formants were extracted at 40% of the vowel duration using a Praat script. Vowels were Lobanov normalized and rescaled using the overall speaker mean and standard deviation, and then pre-obstruent tokens of F2 of the LOT vowel were extracted from the data (n=2337, 19-89 per speaker). F2 of LOT was selected for emphasis because the St. Louis area participated in the Northern Cities Shift, which involves fronting of the vowel (Duncan 2018, *inter alia*).

I model speaker representations with two hidden Markov models (Visser and Speekenbrink 2010): one which assumes two states and one which assumes one. These models iterate through the data to attempt to find the mean value for each of a specified number of states. When there is more than one, the model predicts transition probabilities that represent how likely the system is to switch between states. In the present data, we may interpret states as representations/targets: one state will have a higher F2 value as a mean than the other (representing fronter/backer vowels), and the transition probabilities thus model the speaker selecting between fronter and backer representations. For each speaker, I assess model fit using the Bayesian Information Criterion; the model with the lower BIC is selected. I interpret the model with two states fitting the data better as indicating that the data supports a view of competing representations, and that speakers have both a fronter and backer target for the vowel. I interpret the model with one state fitting the data better as indicating that intraspeaker variation is based on production targeting one representation, even as it may include fronter and backer tokens.

BIC was lower for the two-state model in 14/52 model pairs (26.9%). Given that the implicit assumption is that no two-state models would be selected, this is surprisingly high. The two states appear to represent quite distinct targets; the average 'front' state was 1638.8 Hz, while the average 'back' state was 1324.7 Hz. The average range of 314.1 Hz suggests that speakers for whom the two-state model was selected seem to in fact have two targets. The targets additionally appear to largely not be due to speech errors or lexical items lagging/leading change. Were this the case, we would expect to find that transition probabilities predict that one state is quite likely to be followed by the other, but not vice versa. This would indicate that one state is the speaker's representation of the LOT vowel, but that occasional outliers (whether speech error or lexical exception) are uttered before speakers almost certainly return to their primary target. While this was the case for two speakers, it was not the case for 12/14 speakers, who were quite likely to continue in one state upon having uttered a token in that state (p at least .686). These 12 speakers appear to truly vary between two representations.

The question, of course, is why these speakers and no others would do so. Having multiple representations does not appear to be conditioned by the social factors of age, education level, or location (urban vs. suburban) within Greater St. Louis. The speakers' ages additionally do not appear to cluster around a period of change (whether advance/retreat of fronting) in F2 of LOT. Nevertheless, I hypothesize that having multiple representations may play some role in sound change, and may perhaps arise due to as yet unknown social or cognitive factors. For now, we may say that individuals may have competing representations, and that whether an individual does is itself variable. I call for additional research into this problem, as replication of this result should lead us to more deeply consider the origin and role of competing representations.

References:

Duncan, Daniel. 2018. Language variation and change in the geographies of suburbs. Doctoral dissertation, New York University.

Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini, Scott Seyfarth, Kyle Gorman, Hilary Prichard, and Jiahong Yuan. 2014. FAVE (Forced Alignment and Vowel Extraction) Program Suite v1.2.2 10.5281/zenodo.22281.

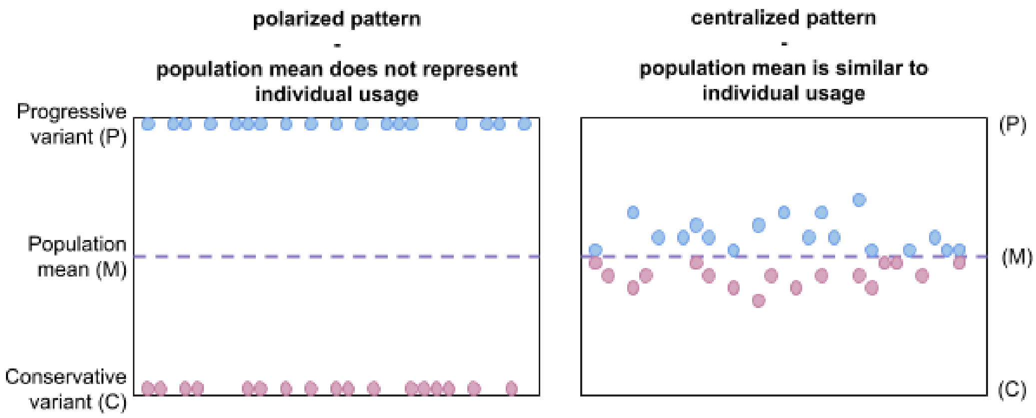
Sneller, Betsy. 2018. Mechanisms of phonological change. Doctoral dissertation, University of Pennsylvania.

Visser, I, and M. Speekenbrink. 2010. depmixS4: An R Package for Hidden Markov Models. *Journal of Statistical Software*, 36(7): 1-21. URL http://www.jstatsoft.org/v36/i07/.

Lauren Fonteyn and Andrea Nini
Leiden University Centre for Linguistics and University of Manchester

Individuality in syntactic variation: an investigation of the 17th-century gerund alternation

Despite the longevity of interest in the role individuals play in language change, it appears that historical (corpus) linguistics has predominantly focussed on studying change in the 'grammar-generated variation' in aggregated, population-level data. However, in recent years, scholars have been placing more explicit emphasis on the relation between the linguistic behaviour of individuals and the changes we observe in such population-level language (see, among many others, Baxter & Croft 2016; Hundt *et al.* 2017; Petré & Van de Velde 2018; Mackenzie 2019). One notable example is a study by Nevalainen et al. (2011), who use real-time historical corpus data from the Early Modern English period to document how individual language users 'participate' in different types of morpho-syntactic change. From their study, Nevalainen et al. (2011) concluded that, when confronted with the fact there are "different ways to say the same thing" (Labov 1972:188), individual language users are more likely to actually participate in the variability that is observable in population-level language if the process is protracted and involves a change in an abstract structural pattern. In other words, if two variant forms are used at a 50-50 proportion in an aggregate data pool, the behaviour of individuals tends to be more centralized (rather than polarized) when the two variants concern abstract syntactic patterns in competition.



Given such figures, it is tempting to conclude that slow, gradual changes of abstract patterns are essentially changes whereby the behaviour of individual language users aligns with the mean observed in aggregate language. However, such conclusions cannot and should not be drawn, as such figures simply represent how often individuals opt for a progressive or conservative variant, and thus, they do not reveal anything regarding the potential differences in 'grammatical constraints' these individuals apply to condition the observed variation.

The aim of this study, then, is to investigate (i) whether individuals who use 'alternate ways of saying the same thing' employ shared or idiosyncratic rules to condition that variation, and (ii) at what level of specificity such possible idiosyncrasies emerge. The variation pair under scrutiny is illustrated in (1):

(1) a. *...the dishonour of Gods Name should affect us more then* [*the shedding **of** the warmest blood in our veins*] (John Flavell, 1668)

b. *he... made an end of...* [*Shedding **ø** the Blood of Rams, Lambs, Heifers, Goats and other Creatures*] (George Fox, 1686)

The structures in (1a)—the nominal gerund [NG]—and (1b)—the verbal gerund [VG]—illustrate two types of *ing*-nominals that were used interchangeably during the Modern English period (see, among many others: Fanego 2004, Nevalainen et al. 2011, De Smet 2013, Fonteyn 2019). The study presents a quantitative analysis of 14,000 NGs and VGs (taken from the *Early Modern Multiloquent Authors* corpus) found in the writings of 19 authors born in three subsequent generations (between 1599 and 1640), all of whom have all been proven to be connected to the same social circles (Petré *et al.* 2019). The data set has been examined by means of two complementary statistical models that have recently been added to the variationist toolkit: Conditional Inference Trees and Random Forests (Tagliamonte & Baayen 2012). Each observation in the data set has been coded for language-internal (e.g. *determiner, function in clause, verb type*) and external factors (e.g. *individual, age, generation, genre*). By homing in on how historical individuals used variable syntactic structures such as the English gerund, this study takes an important step towards documenting "the extent and nature of individual variance for linguistic features at all levels of grammar" (Tagliamonte & Baayen 2012: 24) by means of real-time historical data.

The results of the analysis indicate that individuality is an important predictor of the variation, which trumps the predictive power of some of the proposed language-internal factors as well as the higher-order sociolinguistic factors such as the age or generation of the author. Furthermore, by offering a comparison between author-specific models, we were able to determine that authors do share some 'constraints', but none of the models are exactly alike in terms of breadth, depth, and the specific order and importance of predictors. In light of the model of language proposed by usage-based theories, these results can be explained by the fact that different individuals can come across different exemplars, and consequently will build slightly different models of a construction.

Baxter, G. & W. Croft. 2016. Modeling language change across the lifespan: Individual trajectories in community change. *Language Variation and Change* 28.2, 129-173.

De Smet, H. 2013. *Spreading patterns: Diffusional change in the English system of complementation.* Oxford: Oxford University Press.

Fanego, T. 2004. On reanalysis and actualization in syntactic change: The rise and development of English verbal gerund. *Diachronica* 21.1, 5–55.

Hundt, M., Mollin, S. & S. E. Pfenninger. 2017. *The Changing English Language: Psycholinguistic Perspectives*. Cambridge: CUP.

Labov, W. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

MacKenzie, L. 2019. Perturbing the community grammar: Individual differences and community-level constraints on sociolinguistic variation. *Glossa: A Journal of General Linguistics*, 4(1).

Nevalainen, T., H. Raumolin-Brunberg & H. Mannila. 2011. The diffusion of language change in real time: Progressive and conservative individuals and the time depth of change. *Language Variation and Change* 23.1, 1–43.

Petré, P., L. Anthonissen, S. Budts, E. Manjavacas, E.-L. Silva, W. Standing & O. A.O. Strik. 2019. *Early Modern Multiloquent Authors (EMMA),* release 1.0. University of Antwerp, Linguistics Department.

Petré, P. 2017. The extravagant progressive: an experimental corpus study on the history of emphatic [*be* V*ing*]. *English Language and Linguistics* 21.2, 227-250.

Petré, P. & F. Van de Velde. 2018. The real-time dynamics of the individual and the community in grammaticalization. *Language* 94.4.

Tagliamonte, S. & H. Baayen. 2012. Models, forests and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24.2, 135-178.

Frederik Hartmann
University of Konstanz

RUSE 2019

Investigating the phonological predictability of sound change using deep neural networks

Abstract submission RUSE 2019

Abstract

The traditional view of sound changes being quasi-arbitrary events that do not adhere much to rules and patterns that could be predictable has been frequently challenged and although directionality and long-term conditioning of sound change have been studied and successfully demonstrated [1, 2, 3, 4, 5, 6], not much research has been devoted to analyzing if the phonological conditions of sound changes is trans-temporally applicable, i.e. that the circumstances leading to past sound changes can be extrapolated to correctly predict future changes. In this study, I train deep neural networks on sound changes that occurred from Old High German (OHG) to Middle High German (MHG) and task them afterwards with predicting the sound changes that occurred from MHG to New High German (NHG) which can, in turn, be checked whether they are accurate. The data for this study are 1482 German lemmas that appear in all three language stages considered here which were extracted from the English Wiktionary .xml dump on 20.10.2018 [7]. After processing, the data consisted of 22230 feature matrices encoding each phonetic segment of these lemmas according to 28 phonetic features for each sound in each word and a label vector where binary values indicate whether the particular sound belonging to the respective feature matrix has changed from OHG to MHG. Similarly, the same data matrices and label vectors were created for the change from MHG to NHG. That the features of the sound in question are themselves informative about the occurrence of a change is nothing surprising given that there is a raw statistical likelihood of change by the frequency distribution of changes alone. In the data used in this analysis, OHG /r/ appears in 539 instances and only in 16 of these, it was either reduced or underwent a sound change. This, while being in accordance with previous findings of the changes /r/ did or did not undergo, yields a change probability of P = 0.03 for OHG /r/. These probabilities are detected by a neural network (model I)[1], leading to a prediction accuracy of 0.755 (ROC-AUC score)[2] for changes from MHG to NHG. Yet this finding is unremarkable since it only shows a deep neural network can detect conditional probabilities for sound-specific change events and that these probabilities are temporally consistent enough to map on later changes in the language's history fairly well. To investigate whether sound changes are predictable via a word's particular phonological circumstances, it is necessary to train a network to detect whether a change occurs from OHG to MHG given *only* the phonological environment of the sound in OHG. If this network is, in turn, able to predict changes from MHG to NHG correctly on the basis of the OHG-MHG data, it would be evidence for trans-temporal phonological conditioning of sound changes. This can be done with two different approaches: (1) I trained a model on the full OHG sound environment data to detect change-prone environments independent of the respective sound's features (model II). The results show a prediction accuracy of 0.760 (ROC-AUC score) for changes from MHG to NHG. The problem with this approach is that it is not sound-specific and thus only considering *general* features of change-prone environments. Moreover, we encounter the risk of the network, in reality, inferring the target sound features and then estimating the probabilities similar to the network mentioned above. (2) A solution to this problem would be to train separate networks on single sounds where we can be sure the change probability is the same for all samples during training since it is the same for all samples. In doing this, we could force the network to actually 'learn' conditions for sound changes. Yet to analyze every sound individually, the sample sizes for individual sounds is too small with too few changes to safely train a neural network on. In preliminary tests on some of these few samples, the networks performed between 5 and 10 percent above the random baseline, however to obtain reliable results, further research is needed. The overall results show that sound changes are at least somewhat predictable from the conditions of sound changes present in the previous language stage. This observation supports the claim that sound changes are not singular events but can be predicted with some accuracy from phonological conditions only.

[1]For further details on all trained models, please refer to the appendix.
[2]With 50 percent expected by a random baseline.

Appendix

Table 1: Network architecture and evaluation for model I

Network type	Dense neural network	
Output type	Binary	
Optimizer	Adam	
Batch size	250	
Layer	Layer size	Activation
Dense layer 1	64	ReLU
Dense layer 2	32	ReLU
Output layer	2	softmax

Pred. results test data	Model evaluation	Random baseline
Precision	0.388	0.110
Recall	0.685	0.491
F1 score	0.495	0.181
ROC-AUC score	0.773	0.481
Pred. results MHG-NHG		
Precision	0.388	0.1
Recall	0.658	0.483
F1 score	0.447	0.165
ROC-AUC score	0.755	0.491

Table 2: Network architecture and evaluation for model II

Network type	Multi-input CNN	
Output type	Binary, weighted	
Optimizer	Adam	
Batch size	250	
Layer	Layer size	Activation
Conv. layer 1a and 1b	256	ReLU
Conv. layer 2a and 2b	128	ReLU
Dense layer 1a and 1b	64	ReLU
Dense layer 2	256	ReLU
Dense layer 3	128	ReLU
Output layer	2	softmax

Pred. results test data	Model evaluation	Random baseline
Precision	0.316	0.114
Recall	0.882	0.496
F1 score	0.466	0.185
ROC-AUC score	0.818	0.498
Pred. results MHG-NHG		
Precision	0.259	0.101
Recall	0.775	0.489
F1 score	0.388	0.168
ROC-AUC score	0.760	0.496

References

[1] Patricia J. Donegan and Geoffrey S. Nathan. Natural phonology and sound change. In Patrick Honeybone and Joseph Curtis Salmons, editors, *The Oxford handbook of historical phonology*, Oxford handbooks in linguistics, pages 431–449. Oxford University Press, Oxford, 2015.

[2] Juliette Blevins. Evolutionary phonology: A holistic approach to sound change typology. In Patrick Honeybone and Joseph Curtis Salmons, editors, *The Oxford handbook of historical phonology*, Oxford handbooks in linguistics, pages 485–500. Oxford University Press, Oxford, 2015.

[3] J. J. Ohala. Coarticulation and phonology. *Language and speech*, 36 (Pt 2-3):155–170, 1993.

[4] J. J. Ohala. The phonetics of sound change. In Charles Jones, editor, *Historical linguistics*, Longman linguistics library, pages 237–278. Longman, London, 1993.

[5] Mark Hale. Neogrammarian sound change. In Brian D. Joseph, editor, *The handbook of historical linguistics*, Blackwell handbooks in linguistics, pages 343–368. Blackwell, Malden, MA, 2003.

[6] Ricardo Bermúdez-Otero. Individual differences and the explanation of sound change, 2017.

[7] Wiktionary, the free dictionary. https://en.wiktionary.org. Accessed: 2019-03-13.

Andres Karjus, Richard A. Blythe, Simon Kirby and Kenny Smith
University of Edinburgh

Changing communicative need predicts lexical competition and contributes to language change

Large diachronic text corpora, as samples of utterances produced by populations over time, enable a usage-based approach to the study of evolutionary dynamics in languages, while advances in machine learning allow for automatic inference of semantic proximity and meaning change (cf. Hamilton et al. 2016, Xu and Kemp 2015, Schlechtweg et al. 2017, Turney et al 2019). We present work on identifying and quantifying interactions between words and test the hypothesis that increased communicative need in a semantic subspace supports co-existence of similar words, while low communicative need leads to a survival of the fittest situation.

'Competition' is used here to refer to any processes where the usage dynamics of some word or words affect the usage of other words. An example of direct competition would be a word that goes out of usage due to being replaced by a new borrowing, or by another native word that has undergone semantic change and is being used in an overlapping sense. An example of less direct competition would be a word that goes out of usage because its entire discourse topic is going out of usage, in turn due to the rise of new topics.

Our approach to detect competitive dynamics is based on the following components. Semantic proximity is estimated using diachronic word embeddings (cf. Yao et al. 2018). Distributionally similar words are more likely to be in direct competition to be used in an utterance than unrelated ones. Frequency change is the obvious indicator of change in usage. The strength of competition is quantified using a metric based on the idea that as the usage of a word increases, some other word(s) must decrease for the probability mass to be equalized (as occurrence probabilities in a corpus segment sum up to 1). If the top synonym(s) of a target have decreased as much as the target increased, then this indicates likely competition between them.

Often entire clusters of semantically similar words increase (or decrease) together, exhibiting no sign of competition between them. We have previously quantified this effect in the topical-cultural advection model ([anonymized]), which measures the extent that a change in the frequency of a word can be attributed to the mean frequency change in its topic. We use this measure as a proxy to changing communicative need (cf. Regier et al. 2016, Gibson et al. 2017) in a semantic subspace.

We test our approach on subsets of data drawn from large diachronic corpora of multiple languages: English, German, Estonian, and Scottish Twitter English. In addition to the advection variable, we also control for a number of lexico-statistical measures, including commonality of the form, formal similarity to nearest neighbours, semantic subspace density, semantic change, frequency, momentum, and dissemination in the corpus segment. We also include a simple measure of polysemy, to control for the distorted signal of polysemous words that are typically not well captured by simple word embeddings. Semantic change is also how competition could manifest: the 'loser' could acquire a new meaning and continue to be used.

These variables are subsequently modelled in a standard regression framework. The technical complexity underlying the data collection process requires careful control for confounds and making sure the observed effects are not artefacts of the machine learning models (Dubossarsky et al. 2016). To that end, the models are compared with randomized baselines, based on randomized embeddings. Increasing communicative need in a semantic subspace, operationalized by the advection model, does predict decreased competition, while high semantic similarity in subspaces with neutral or negative advection often leads to competition. In addition to presenting completed work, we will discuss ongoing work on testing these findings experimentally.

References

Dubossarsky, H., Weinshall, D., & Grossman, E. (2017). Outta Control: Laws of Semantic Change and Inherent Biases in Word Representation Models. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 1147–1156.

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., Gibson, M., Piantadosi, S., Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*.

Hamilton, W.L., Leskovec, J., Jurafsky, D. (2016). Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016. Volume 1: Long Papers*.

[anonymized]

Regier, T., Carstensen, A., Kemp, C. (2016). Languages Support Efficient Communication about the Environment: Words for Snow Revisited. *PLOS ONE 11, 1–17*.

Schlechtweg, D., Eckmann, S., Santus, E., im Walde, S. S., & Hole, D. (2017). German in Flux: Detecting Metaphoric Change via Word Entropy. *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, 354–367.

Stewart, I., Eisenstein, J. (2018). Making "fetch" happen: The influence of social and linguistic context on nonstandard word growth and decline. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Brussels, Belgium*, pp. 4360–4370.

Turney, P.D., Mohammad S.M. (2019) The natural selection of words: Finding the features of fitness. *PLOS ONE 14(1): e0211512*.

Xu, Y., Kemp, C., 2015. A Computational Evaluation of Two Laws of Semantic Change. *CogSci 37*.

Yao, Z., Sun, Y., Ding, W., Rao, N., Xiong, H. (2018). Dynamic Word Embeddings for Evolving Semantic Discovery. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM '18*, pp. 673–681.

Henri Kauhanen and George Walkden
University of Manchester and University of Konstanz

RUSE 2019

When is a constant rate truly constant? A Monte Carlo power analysis of the logistic operationalization of constant rate effects

The Constant Rate Hypothesis (CRH) maintains that when a change targets multiple linguistic contexts simultaneously, the rate of change is the same across contexts (Kroch, 1989). Assuming that linguistic changes are S-shaped (see Denison, 2003; Blythe and Croft, 2012; Nevalainen, 2015), most scholars working on Constant Rate Effects (CREs) have followed Kroch (1989) in operationalizing this hypothesis using the simple logistic

$$p_c(t) = \frac{1}{1+e^{-s_c(t-k_c)}}. \tag{1}$$

Here $p_c(t)$ is the frequency of the innovation in context c at time t, s_c gives the rate of change, and k_c serves to translate the entire curve along the time axis. The Constant Rate Hypothesis is the statement that $s_c = s$ for some unique s, for all contexts c. Variation in the k_c parameter, on the other hand, is allowed.

Since its formulation 30 years ago, a number of studies have sought to establish the CRH by fitting (1) to various historical datasets (e.g. Kroch, 1989; Santorini, 1993; Pintzuk, 1995; Kallel, 2007; Wallage, 2008; Fruehwald, Gress-Wright, and Wallenberg, 2013; Wallage, 2013; Zimmermann, 2017). This pursuit has, however, been criticized on two grounds. Firstly, it has been pointed out that the currently available statistical methodology casts the CRH as the null, not the alternative, hypothesis (Paolillo, 2011). This has the unwelcome consequence that the rate at which the methods report false positives is unknown, a problem which is compounded by the typically small size of the datasets encountered in historical linguistic work. Secondly, it has been argued that operationalizing the CRH on the logistic (1) is ill-founded, as the constancy of rates of change across contexts is not actually being derived from any underlying model of speaker–speaker interactions (Kauhanen & Walkden, 2018).

In this contribution, we provide a systematic *a priori* power analysis of the logistic operationalization of the CRH using a Monte Carlo set-up. We consider two models: (A) one in which the s_c (rate of change) parameter is allowed to vary across contexts, and (B) one in which it is tied to a constant value: $s_c \equiv s$. The two models were used to generate synthetic data in a binomial experiment in which V tokens of the innovative linguistic form were drawn for each of H time points across the length of the logistic, for L contexts. The two models were then cross-fitted to each other's data using a logistic regression model involving a time–context interaction term (cf. Fruehwald et al., 2013); whenever the interaction term proved statistically significant at the $\alpha = 0.05$ level, model A (no CRE) was selected, otherwise model B (CRE) was selected. The procedure was repeated 1,000 times with each model as the generator and a confusion matrix established. Effect size was estimated as

$$E = 1 - \frac{|\min_c(s_c)|}{|\max_c(s_c)|}. \tag{2}$$

In other words, the closer the shallowest and steepest slopes s_c are to each other, the smaller the effect of s_c on the probability of the new linguistic form; E ranges from 0 to 1.

The Monte Carlo simulations indicate that the type II error rate β of the method may be approximated to a good degree by a combination of the effect size E and an overall data resolution quantity R, which is simply expressed as the product of the three data size parameters: $R = VHL$. The dependence of β on E and R is

$$\beta = \frac{1-\alpha}{1+\left(\frac{E}{11.60R^{-0.45}}\right)^{4.44}} \tag{3}$$

for a type I error rate α (see Figure 1).

Equation (3) can be used to estimate the data size required to reach a desired level of statistical power when modelling CREs using the logistic operationalization. It can also be used to evaluate to what extent previous studies have managed to reach reasonable standards of statistical rigour. To this end, we compiled a database of 27 studies from existing literature on CREs, and measured the data resolution ($R = VHL$) and effect size (E) quantities for each study. Of the 27 studies, 9 reach a type II error rate of 0.05 or less (a power of 0.95 or more), but nearly half of the studies (13) attest an error rate in excess of 0.5. The significance of this finding for the methodology of quantitative corpus-based diachronic linguistics will be discussed. We will also comment on whether the adoption of alternative models of the CRE (Kauhanen & Walkden, 2018) can be expected to ameliorate the situation.

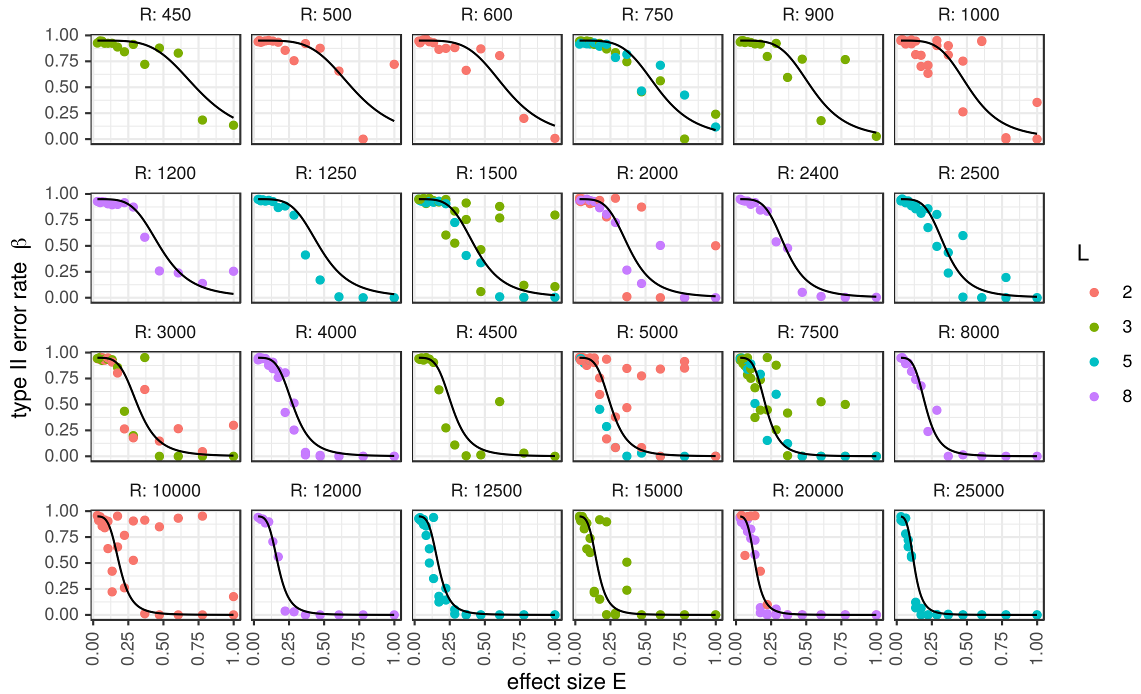


Figure 1. Type II error rate β of the logistic operationalization of CREs as a function of effect size E, overall data resolution R and number of contexts L, for type I error rate $\alpha = 0.05$. Points from simulations, curves from equation (3).

References

Blythe, R. & Croft, W. (2012). S-curves and the mechanisms of propagation in language change. *Language*, *88*(2), 269–304. Retrieved from https://www.jstor.org/stable/23251832

Denison, D. (2003). Log(ist)ic and simplistic s-curves. In R. Hickey (Ed.), *Motives for language change* (pp. 54–70). Cambridge: Cambridge University Press.

Fruehwald, J., Gress-Wright, J., & Wallenberg, J. C. (2013). Phonological rule change: The constant rate effect. In S. Kan, C. Moore-Cantwell, & R. Staubs (Eds.), *NELS 40: Proceedings of the 40th Annual Meeting of the North East Linguistic Society* (Vol. 1, pp. 219–240). GLSA Publications. Retrieved from https://www.research.ed.ac.uk/portal/files/14416788/Fruewald_Gress_Wright_Wallenberg_Phonological_Rule_Change.pdf

Kallel, A. (2007). The loss of negative concord in Standard English: Internal factors. *Language Variation and Change*, *19*(1), 27–49. doi:10.1017/S0954394507070019

Kauhanen, H. & Walkden, G. (2018). Deriving the Constant Rate Effect. *Natural Language & Linguistic Theory*, *36*(2), 483–521. doi:10.1007/s11049-017-9380-1

Kroch, A. S. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change*, *1*(3), 199–244. doi:10.1017/S0954394500000168

Nevalainen, T. (2015). Descriptive adequacy of the s-curve model in diachronic studies of language change. In C. Sanchez-Stockhammer (Ed.), *Can we predict linguistic change?* (Vol. 16). Studies in Variation, Contacts and Change in English. Helsinki: VARIENG. Retrieved from http://www.helsinki.fi/varieng/series/volumes/16/nevalainen/

Paolillo, J. C. (2011). Independence claims in linguistics. *Language Variation and Change*, *23*, 257–274.

Pintzuk, S. (1995). Variation and change in Old English clause structure. *Language Variation and Change*, *7*(2), 229–260. doi:10.1017/S0954394500001009

Santorini, B. (1993). The rate of phrase structure change in the history of Yiddish. *Language Variation and Change*, *5*(3), 257–283. doi:10.1017/S0954394500001502

Wallage, P. (2008). Jespersen's Cycle in Middle English: Parametric variation and grammatical competition. *Lingua*, *118*(5), 643–674. doi:10.1016/j.lingua.2007.09.001

Wallage, P. (2013). Functional differentiation and grammatical competition in the English Jespersen Cycle. *Journal of Historical Syntax*, *2*(1), 1–25. doi:10.18148/hs/2013.v2i1.4

Zimmermann, R. (2017). *Formal and quantitative approaches to the study of syntactic change: Three case studies from the history of English* (Doctoral dissertation, University of Geneva). Retrieved from http://www.old-engli.sh/myres%20documents/Zimmermann_R.(2017)PhDDissertation.pdf

Li Lei, Limor Raviv and Phillip Alday
Radboud University Nijmegen and Max Planck Institute for Psycholinguistics, Nijmegen

Language Change in Online Social Networks

Over the past few decades, computational methods have emerged as a powerful tool in studying language change in social networks, given their ability to test hypotheses and easily manipulate social properties in large scale.

For example, a previous study (Fagyal, Swarup, Escobar, Gasser, & Lakkaraju, 2010) has proposed a degree-biased voter model (DBVM) to simulate language change in social networks, focusing on the role of hubs and loners in the community. There were eight linguistic variants which could spread in a bi-directional closed network and each agent in the network would choose a neighbour to update its linguistic variant. The selection rule was that a higher in-degree leads to a higher chosen probability. In the DBVM, the establishment of novel variants could be observed after language diffusion. This model revealed that loners play a key role in language change as variant-keepers.

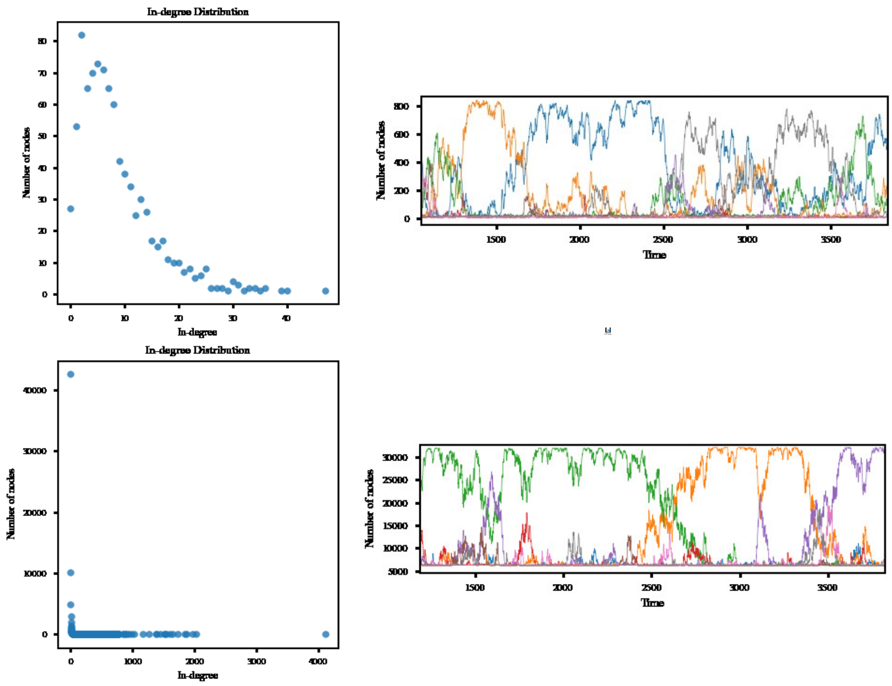
However, the social network structure used in Fagyal et al. (2010) was simulated and based on "an artificial but socially realistic influence network", which relies on previous theories on social science and does not necessarily adhere to all properties of real-world networks. Here, we extend the original model to use network structure inspired by real-world online communities, and explore the model’s application to such a network.

First, we successfully replicated the DBVM model presented in Fagyal et al. (2010), which included 900 agents and 7561 edges. We obtained a similar in-degree distribution (Figure 1a) and similar diffusion results over 40,000 iterations (Figure 1b), following the same pattern identified in Fagyal et al. (2010). The result showed that competing variants which were represented by different colours in Figure 1b could establish the norm alternately over time on the basis of different chosen probability.

Next, we tested this model using a real-world "who-trust-whom" online social network with 75,879 agents and 508,837 edges, which was extracted from a general review website Epinions (Richardson, Agrawal & Domingos, 2003). While its in-degree distribution was different from that of the original network (Figure 1c), the process of language change indicated a highly similar diffusion dynamics between two networks (Figure 1d). This result shows that the DBVM can model a real-world network.

Finally, we scaled down the Epinions network to include 900 agents with the same in-degree distribution and ratio and ran several simulations with different numbers of edges (5000, 7561, 27000). The results showed similar trends in language change: alternate norms of linguistic variants could be observed. Interestingly, we found that more edges lead to a longer and smoother fixation of linguistic variants. This result suggests that language change is more difficult in a high-density network, in line with an empirical study in Belfast’s Protestant enclaves (Milroy, 1987): a norm is easier to maintain in a high-density community. In ongoing work, we plan to change additional variables of the online social network, such as a more "socially realistic" update rule instead of simple in-degree-bisad communication.

(a) (b)



(c) (d)

Figure 1: (a) In-degree distribution in the replicated DBVM; (b) Changes in the prominence of linguistic variants over an extended time series in the replicated DBVM; (c) in-degree distribution in the real-world Epinions network; (d) Changes in the prominence of linguistic variants over an extended time series in the real-world Epinions network.

References

Fagyal, Z., Swarup, S., Escobar, A. M., Gasser, L., & Lakkaraju, K. (2010). Centers and peripheries: Network roles in language change. *Lingua, 120*(8), 2061-2079.

Milroy, L. (1987). *Language and social networks*. Oxford: Basil Blackwell.

Richardson, M., Agrawal, R., & Domingos, P. (2003). Trust management for the semantic web. *Proceedings of the 2nd International Semantic Web Conference*. 351–368.

Marcelo Melo and Christina Gomes
Federal University of Rio de Janeiro

RUSE 2019

Frequency and social effects in the lenition of coda fricative in Brazilian Portuguese

This presentation addresses the competing effects of word frequency and social group in the directionality of sound change in the variable use of fricative coda (s) in Brazilian Portuguese as in me[ʒ]mu, me[z]mo, me[ɦ]mo e meØmo ‘mesmo’ *same*, in the speech community of Rio de Janeiro, Brazil, specifically comparing data from 8 speakers from EJLA Sample and 8 speakers from Censo 2000 Sample. The former is formed by socially excluded adolescents without regular schooling from carioca slums and the later are middle-class speaker with Elementary and High school. Evidence is provided for the competing effect of structural conditioning, regarding following context, word frequency and social group and also that there are different patterns of variation being developed in the speech community. According to Bybee (2002, 2010), the dual view of the unit of a sound change, whether the lexical item or the segment, is rooted in a view of phonology that states phonemic underlying representations based on discrete segments (phonemes). Bybee (2002, 2010) also suggests that the resolution of this dichotomy, also known as neogrammarian controversy (LABOV, 1981, 1994), relies on the propositions of an exemplar model of phonology which provides a unifying treatment for the simultaneous effect of both phonetic conditioning and the gradual implementation of sound change in the lexicon, since actual instances of words, as they are produced and perceived by the speaker/hearer, are part of the representations of the wordforms in the lexicon. Logistic Regression using R-brul (JOHNSON 2014) found significant fixed effects of following context and morphological status (morpheme vs. non-morpheme), in both samples. For EJLA Sample data, it was also found significant random effects for speaker and word. In relation to internal constraints, lenition of the fricative to a back articulation – [x ɣ h ɦ] – is favoured when the coda (s) is followed by voiced consonants (Table 1), and when it has no morphological status. The level of lenition depends on the individual speaker, although these 8 lower-class speakers from EJLA Sample (30%) used more lenition than the 8 speakers from Censo 2000 (5%) and other social groups in the community as well (SCHERRE and MACEDO 2000; CALLOU and BRANDÃO 2009). As lenition depended on the identity of the word, a continuous numeric predictor was added, representing the frequency of the words in the EJLA corpus. An Rbrul run retained the significant effects observed above and also included the effect of word frequency: the lenited back fricative variant occurs more often in frequent words, as might be expected (BYBEE 2010). However, the significant random effect of word remains, prompting the search for other relevant word-level predictors while leaving open the possibility that lexical items may individually favour or disfavour the lenition process. Chart 1 also presents evidence that, besides word frequency, social constraint plays a role in the implementation of the back variant. Although the highest frequent words favors the velar/glottal variant for EJLA data, the same words present the opposite pattern in Censo Sample, with the prevalence of the post-alveolar fricative in coda. We argue that, although the effect of following context is similar in the data from both samples, the speakers from each sample present different representational patterns for some words (for instance, *mesmo*, *nós*, *às vezes*), according to the Exemplar Model hypothesis, probably due to the social evaluation of the back fricative (MELO 2017). These results provide further evidence of token frequency as a mechanism of change implementation but also that social values related to the variants can play an important role in the directionality of language change, consisting a key aspect to be considered in the issues related to the actuation problem (WEINREICH, LABOV, HERZOG 1968).

Table1. Effect of following context in the variable use of glottal in coda in BP

	CENSO 2000			EJLA		
	Apl/N	%	peso	Apl/N	%	peso
consoante soante	157/735	21,4	0,961	413/472	87,5	0,988
obstruinte sonora	71/781	9,1	0,899	271/336	83,3	0,965
obstruinte surda	12/2160	0,6	0,290	153/1245	11,7	0,381
vogal	1/972	0,1	0,045	5/339	1,5	0,029
pausa	2/561	0,4	0,191	5/300	1,3	0,023

Chart 1. Lexical conditioning of the glottal in coda in BP

		CENSO 2000		EJLA	
ITENS	Meaning	Apl/N	%	Apl/N	%
às vezes	sometimes	0/141	0	81/85	95,3
mesmo	same	41/279	14,7	113/127	89
desde	since	3/23	13	10/12	83,0
vários	many	0/12	0	17/26	65
nós	we	7/34	20,6	137/235	58,59
dois	two	2/56	3	37/74	50
mais	more	44/332	13,3	52/108	48,14
mas	but	32/461	6,9	65/144	45,13
as	as	6/104	4	20/58	34,5
duas	two (fem.)	2/40	5	32	31
eles	they	8/234	3,4	23/124	21,30
os	the (det. masc.	5/167	3	263	21
existe	exist	0/26	0	0/10	0

References

BYBEE, Joan. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14 , pp 261-290, 2002.

______ . *Language, usage and cognition*. Cambridge: Cambridge University Press, 2010.

CALLOU, D. M. I.; BRANDÃO, S. F. "Sobre o /S/ em coda silábica no Rio de Janeiro: falas culta e popular". In: Salgado, Ana Claudia Peters; Barretto, Mônica M. Guimarães Savedra. (Org.). *Sociolinguística no Brasil*: uma contribuição dos estudos sobre línguas em/de contato: homenagem ao Prof. Jürgen Heye. Rio de Janeiro: 7 Letras, p. 27-34, 2009.

Johnson, D. E. (2014) *Progress in regression: Whysociolinguistic data calls for mixed-effects models*. Self published manuscript. < http://www.danielezrajohnson.com/johnson_2014.pdf >

LABOV, William . "Resolving the neogrammarian controversy". In Language, 57: 267-308, 1981.

______ . Principles of linguistic change: Internal factors. Oxford: Blackwell, 1994.

MELO, M. A. S. L. *Direcionalidade da mudança sonora: o papel do item lexical e da avaliação social*. Tese (Doutorado em Linguística) – UFRJ, Faculdade de Letras, Rio de Janeiro, 2017.

SCHERRE, Maria Marta Pereira e MACEDO, A. V. T. (2000) Restrições fonético-fonológicas e lexicais: o -S pós-vocálico no Rio de Janeiro. In: Maria Cecília Mollica; Mário Eduardo Martelotta (org.). *Análises linguísticas*: a contribuição de Alzira Macedo. Rio de Janeiro: Serviço de Publicações - FL/UFRJ, p. 52-64.

WEINREICH, W., LABOV, W., HERZOG, M. *Empirical Foundations of a Theory of Language Change*. In: Lehmann, W. P.; Malkiel, Y. (eds) Directions for Historical Linguistics. Austin: University of Texas, 1968, p.97-195.

Ilja Serzant
Leipzig University

Evolutionary Typology of Bound Verbal Person-Number Indexes

Intro: The present paper seeks to identify universal trends in the lengths of person-number affixes of verbs. Instead of *bound pronouns* or *agreement markers* and following Lazard (1998) and Haspelmath (2013) and refer to these as to *bound person-number indexes*. The precise morphophonological realization of indexes is subject to cross-linguistic variation, which, however, has no bearing on the claims to be made here (cf. Haspelmath 2013). For example, I gloss over the morphological differences between affixes and clitics. Given the diachronic perspective of the paper it does make sense to treat these as variants of the same phenomenon. In order to identify universal trends I adopt the *dynamic approach* to universals (since Greenberg 1969) which is in contrast to the more traditional, *static approach.* The *dynamic approach* relies on the comparison of two subsequent historical stages, i.e. a proto-stage and its modern descendant, in order to investigate the *transition* between these two. The crucial question here is whether the relevant mechanisms of change provide evidence for the alleged universal (Bybee 2008) or whether the changes bring about a higher degree of adherence to the alleged universal pattern than before these changes (Bickel et al. 2014). The dynamic approach overcomes a number of shortcomings of the static approach.
Methods and the data: I rely on a database with obligatory intransitive subject indexes from (a) 290 modern languages from 14 unrelated (sub)families covering all macroareas and (b) their proto-forms as reconstructed by the Historical-Comparative Method in the authoritative literature: Indo-European, Uralic, Mayan, Dravidian, Semitic, Oceanic (a subfamily of Austronesian), Bantu (Niger-Congo), Sogeram (Trans-New-Guinea stock), Awyu-Dumut (Trans-New-Guinea stock), Rgyalrongic-Kiranti (Tibeto-Burman), Worrorran, Muskogean, Athabaskan and Turkic. The transition is revealed by comparing the lengths of the proto-forms and the respective averaged lengths in the modern languages (approx.. 10-50 modern languages per (sub)family).
Claims and discussion: I will make three major claims as regards the lengths of the indexes: (i) there is an attractor state in terms of “the ideal length” of the indexes which modern languages tend to arrive at; (ii) modern languages tend to develop shorter indexes for the third person as opposed to the first and second person but there is no trend for zeros (cf. Bickel et al. 2015; pace Benveniste 1971; Koch 1995; Bybee 1985: 53; Cysouw 2003: 61-2; Siewierska 2010); (iii) modern languages tend to keep plural indexes longer than their respective singular counterparts. Crucially, these universal tendencies are observed despite very diverse prerequisites in the respective proto-languages (some (sub)families such as Proto-Indo-European or Proto-Bantu start with proto-forms that do not adhere to these generalizations) and, what is more, despite very diverse historical changes that the particular indexes undergo (reduction by sound law, analogical restructuring, lengthening, retention, etc.).

Moreover, I will show that index systems considerably violating the lengths predictions (i)-(iii) are either recent innovations (predicted to subsequently develop towards adherence) or the whole indexing system undergoes a functional change into a non-referring and thus largely semantically redundant system (gram-indexing) as, for example, in German *geh-e, geh-st, geh-t* where the indexes are not referring anymore.

The mechanism that licences these lengths proportions and the attractor state in the long run is crucially based on frequency that drives the adaptation of the more efficient variant from those produced by various historical accidents (shortening, lengthening, analogical replacement, etc.). The lengths asymmetries in (ii) and (iii) fully align with the predictions by Zipf (1939): the more frequent index is coded with less material while less frequent indexes need more material.

References:

Benvéniste, Emil 1971: Problems in general Linguistics. Translated by Mary Elisabeth Meek. [Miami Linguistics Series 8]. Coral Gables (Florida): University of Miami Press.

Bickel, Balthasar, Alena Witzlack-Makarevich & Taras Zakharko 2014: Typological evidence against universal effects of referential scales on case alignment. In: Bornkessel-Schlesewsky, Ina, Andrej Malchukov, Marc Richards (eds.), *Scales and Hierarchies: a cross-disciplinary perspective on referential hierarchies*. Berlin: De Gruyter Mouton, 7-44.

Bickel, Balthasar; Witzlack-Makarevich, Alena; Zakharko, Taras; Iemmolo, Giorgio 2015: Exploring diachronic universals of agreement: alignment patterns and zero marking across person categories. In: Fleischer, Jürg; Rieken, Elisabeth; Widmer, Paul. Agreement from a diachronic perspective. Berlin: De Gruyter Mouton, 29 - 51.

Bybee, Joan 1985: *Morphology: A Study of the Relations between Meaning and Form*. Amsterdam/Philadelphia: John Benjamins.

Bybee, Joan L. & Mary Alexandra Brewer 1980: Explanation in morphophonemics: changes in Provençal and Spanish preterite forms, *Lingua* 52, 201–242.

Cysouw, Michael 2003: *The paradigmatic structure of person marking*. Oxford: Oxford University Press.

Greenberg, Joseph H. 1966: *Language universals, with special reference to feature hierarchies*. The Hague: Mouton.

Gutman, Eynat 2004: Third person null subjects in Hebrew, Finnish and Rumanian: an accessibility-theoretic account, *Journal of Linguistics* 40, 463-490.

Haspelmath, Martin. 2008: Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19(1), 1-33.

Horn, W. 1921. Sprachkörper and Sprachfunktion. Berlin: Mayer and Müller.

Koch, Harold 1995: The creation of morphological zeros. In Geert Booij (ed.), *Yearbook of morphology*, 31–71. Dordrecht: Kluwer.

Siewierska, Anna. 2010. Person asymmetries in zero expression and grammatical functions. In Franc Floricic (ed.), *Essais de typologie et de linguistique générale. Mélanges offerts à Denis Creissels*, 471-485. Paris: Presses de L'Ecole Normale Supérieure.

Zipf, George. 1935. The Psychobiology of Language: An Introduction to Dynamic Philology. Cambridge, Mass.: M.I.T. Press.

Cesko Voeten
Leiden University

RUSE 2019

On-going developments in the analysis of on-going sound change: answering new questions using new methods

This paper reports on the results of a four-year project investigating language variation and change in Dutch, focusing on how methodological innovations can help us linguists answer new questions. The phenomenon under study is an on-going vowel shift whereby /eː,øː,oː/ are becoming diphthongs [ei,øy,ou], resulting in a push chain which subsequently moves /ɛi,œy,ɔu/ towards [ai,ɒy,ɑu]. These two phonetic shifts are blocked before coda /l/, giving rise to new allophone pairs such as [ei#~eːɫ]. In addition, this vowel shift remains confined to the variety of Dutch spoken in the Netherlands, whereas the (mutually-intelligible) Dutch spoken in Flanders (the northern half of Belgium) has not been affected by these changes.

The project investigated the diffusion of these sound changes throughout both social structure, i.e. the regional differences between and within the Netherlands and Flanders, and linguistic structure, i.e. individuals' phonological and phonetic grammars. Regional differences were assessed by means of the 'teacher corpus', a comprehensive data set collected by Adank (2003), sampling 160 individuals from four regions in the Netherlands and four regions in Flanders. A comparison is made of the phonological behavior of the vowels in question before /l/ versus before nonapproximant consonants. This is a phonetically challenging enterprise, because it turned out impossible to segment the vowel-[ɫ] sequences. I will show that (and how) the formant trajectories can nonetheless be compared by modeling the vowels' whole formant trajectories within a generalized additive model (Wood 2017) and comparing trajectories and peaks of trajectories.

Following the results of the corpus analysis, the perception and production of these vowels was investigated in lab experiments. This time, the object of investigation was the diffusion of the sound changes throughout the linguistic systems of individuals, particularly of Flemish individuals who have lived in the Netherlands for a long time (years -- decades), and who therefore might have accommodated to these sound changes (cf. Evans & Iverson 2007). Participants were 45 speakers of Netherlandic Dutch, 45 speakers of Flemish Dutch, and 18 speakers of Flemish Dutch who had migrated to the Netherlands up to 30 years ago. Production (word-list) and perception data (phoneme decision along a continuum) were collected. Analyses were performed not at the group level, but at the individual level, by looking for clusters in the random effects obtained from a mixed-effects model. The production data show that eight of the eighteen Flemish Dutch speakers have accommodated to the sound changes in Netherlandic Dutch. In perception, no clear results are found at the individual level, but at the group level, significant effects emerge. In addition, significant correlations are found between production and perception.

The results from the production data give rise to a further question: what factors separate the individuals who have acquired the sound changes from those who have not? This is investigated by means of a background questionnaire; results from a linear regression analysis on the responses reveal significant effects of length of stay, physical factors such as concentration, and the number of friendships. These results expand on previous findings by Yu (2013) on the role of socio-cognitive processing in sound change and Lev-Ari (2018) on social network size.

In sum, the present study shows how recent innovations in linguistic and statistical methodology make it possible to answer new questions. This has implications for our analysis of linguistic data, but also for our knowledge of the processes underlying language variation and change.

References

Adank, P.M. (2003). *Vowel normalization: A perceptual-acoustic study of Dutch vowels*. PhD dissertation, Radboud University Nijmegen.

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America*, 121(6), 3814-3826.

Lev-Ari, S. (2018). Social network size can influence linguistic malleability and the propagation of linguistic change. *Cognition*, 176, 31-39.

Wood, S. N. (2017). Generalized additive models: an introduction with R. Chapman and Hall/CRC.

Yu, A.C.L. (2013). Individual differences in socio-cognitive processing and the actuation of sound change. In: Yu, A.C.L. (Ed). *Origins of Sound Change: Approaches to Phonologization* (pp 201-227). Oxford, OUP.