Neutral Models of Sound Change

Andrea Ceolin - University of Pennsylvania

Dissertation Proposal Defense, April 17th, 2019

Slides: https://www.ling.upenn.edu/∼ceolin/slides.pdf

Functionalism in 1994

MECHANICAL PRINCIPLE

'The relative progress of sound change is determined by phonetic factors alone, without regard to the preservation of meaning' (Labov 1994:603)

Functionalism in 2019

'Labov [...] comes to the conclusion that function plays no role in sound change and variation [...] In recent years the question has been reopened with new sophisticated statistical techniques' (Kiparsky 2016:14).

Functionalism in 2019

Piantadosi et al. (2011) - Word length are optimized for efficient communication

Graff (2012) - Communicative efficiency in the lexicon

Wedel et al. (2013) - High functional load inhibits phonological contrast loss

Cohen Priva (2017) - Informativity and the actuation of lenition

Dautriche et al. (2017) - Word cluster phonetically beyond phonotactic regularities

Bentz (2018) - Adaptive Languages: An information-theoretic account of linguistic diversity

Mahowald et al. (2018) - Word forms are structured for efficient use

Four Research Questions

1. Are Mergers more common than Splits?
2. Is phonemic dispersion in the lexicon stable over time?
3. Are Mergers constrained by functional considerations?
4. Do borrowing and word formation influence phonemic distributions?

Methods

1. Historical data (lots of them)
2. A neutral model of sound change

A Neutral Model of Sound Change

My neutral model for sound change

1. Mergers and Splits applied to a mini-lexicon
2. Alphabet:
   C: {*m*, *p*, *b*, *f*, *v*, *d*, *t*, *l*, *n*, *r*, *s*, *c*, *k*, *y*, *w*, *g*, *j*, *h*}
   V: {*i*, *I*, *e*, *E*, *a*, *ə*, *O*, *o*, *U*, *u*}
3. Mini lexicon: {*bad*, *big*, *but*, *can*, *dad*, *for*, *get*, *god*, *him*, *her*, *job*, *let*, *lot*, *man*, *mom*, *not*, *put*, *sir*, *son*, *yes*}
4. Feature representation (hard problem, cf. Dresher 2009)

Split

Algorithm for SPLITS

1. Pick one position in the syllable (onset, nucleus, coda) [**ONSET**]
2. Select one segment in the inventory available in that position [c] and one outside of the inventory [k]
3. Select conditioning environment [O, o, U, u]
4. c becomes k in the conditioning environment.
5. cet, cen, cot -> cet, cen, kot

MERGERS work the same way (only difference at point 2)

Example

Conditioned merger of /a/ in /o/ after [m p f d l n s c g j]
Conditioned split of /l/ in /t/ in onsets before before [i]
Conditioned merger of /n/ in /s/ in onsets before before [i u e o]

bad	->	bod
big	->	big
but	->	but
can	->	con
dad	->	dod
cut	->	cut
for	->	for
him	->	him
job	->	job
man	->	mon
mom	->	mom
not	->	sot
son	->	son
yes	->	yes

1. Are Mergers more common than Splits?

Mergers and Splits

'most reports of phonemic change involve mergers [...] [this fact] would lead to the odd conclusion that most languages are steadily reducing their vowel inventory [...] it stands to reason that just as many phonemic splits must take place as mergers' Labov (1994:331)

Rephrasing of the question: is phoneme loss via mergers more common than phoneme creation via splits?

Mergers and Splits

IE (Ringe 2011), Uralic (Sammallahti 1988) and Altaic (Robbeets 2003)

Lineage	Contractions	Borrowing	Splits	Resolutions	Mergers
Proto-Indo-European to Proto-Indo-Iranian	5		3		11
Proto-Indo-European to Proto-Germanic	3		2	4	8
Proto-Indo-European to Greek	14		1	2	12
Proto-Indo-European to Latin	6			6	8
Proto-Indo-Iranian to Sanskrit	1	5	3	2	4
Proto-Germanic to Old English	4		14		8
Proto-Uralic to Proto-Samoyed			2		3
Proto-Uralic to Proto-Finno-Ugric	5				
Proto-Samoyed to Proto-South-Samoyed					1
Proto-Samoyed to Proto-North-Samoyed					1
Proto-Finno-Ugric to Proto-Ugric			1		5
Proto-Finno-Ugric to Proto-Finno-Permic					1
Proto-Japanese to Old Japanese	8				
Proto-Korean to Middle Korean	4				
Proto-Tungusic to Manchu		1	1		2
Proto-Mongolic to Mongolian	13	1	1		
Proto-Turkic to Turkish	1	2			
TOTAL	**64**	**9**	**28**	**14**	**64**

Table 1: A summary of processes regulating phoneme creation and deletion.

Pseudo-Contraction

PSEUDO-CONTRACTION

1. Select a CV or a VC combination [an]
2. Select a C-segment NOT in the inventory [s]
3. an contracts to s.
   can, man -> cs, ms
4. **A nucleus, randomly chosen among those already present in the language, is added** [o]
   can, man -> cos, mos

Mergers and Splits

Q: Is phoneme loss via mergers more common than phoneme creation via splits?

A: **YES!**

But Contractions are the main source of phoneme creation.

2. Is phonemic dispersion in the lexicon stable over time?

Perception and Production Trade-off

Good perception		**Good production**
Many phonemes No Minimal Pairs	Minimal Pairs	Few phonemes Homonymy
High dispersion (A)	Medium dispersion (B)	Low dispersion (C)
sad set far her	sad sat far har	sat sat far far

Measures of dispersion: number of phonemes, average edit distance, number of minimal pairs

Dautriche et al. (2017)

Dautriche et al. (2017): **Word cluster phonetically beyond phonotactic regularities**

Model: 5-gram language model trained on CELEX phonological transcriptions (with Laplace smoothing). Used to generate pseudo-words

'Results for four languages (Dutch, English, German, French) show that the space of monomorphemic word forms is clumpier than what would be expected by the best chance model according to a wide variety of measures [...] The strongest evidence comes from minimal pairs'

Sources of Clumpiness: diachronic explanation?

'[...] it is likely that we may have a lot to learn from diachronic data to observe how clumpiness evolve in the lexicon' (Dautriche et al. 2017: 143)

Now that we have a proper model of sound change, we can track the change in minimal pairs over time under a neutral model

Phonemes tend to decrease

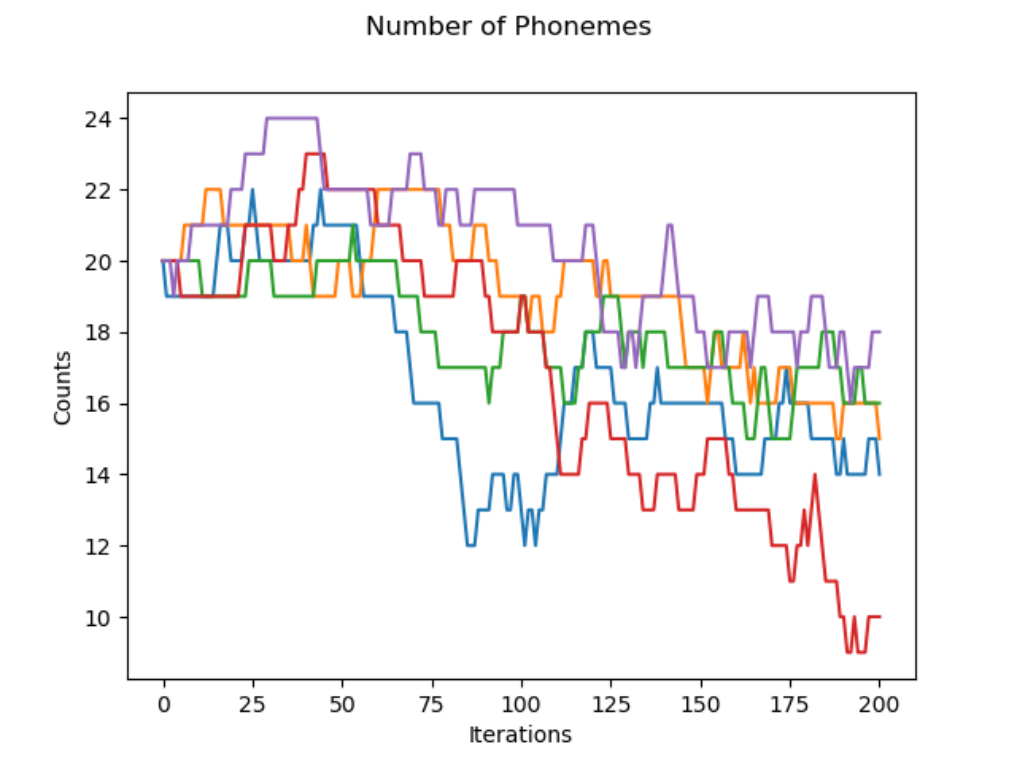


Figure 1: Phonemic counts after 200 sound changes, in 5 parallel runs

Mergers and Splits

IE (Ringe 2011), Uralic (Sammallahti 1988) and Altaic (Robbeets 2003)

Lineage	Contractions	Borrowing	Splits	Resolutions	Mergers
Proto-Indo-European to Proto-Indo-Iranian	5		3		11
Proto-Indo-European to Proto-Germanic	3		2	4	8
Proto-Indo-European to Greek	14		1	2	12
Proto-Indo-European to Latin	6			6	8
Proto-Indo-Iranian to Sanskrit	1	5	3	2	4
Proto-Germanic to Old English	4		14		8
Proto-Uralic to Proto-Samoyed			2		3
Proto-Uralic to Proto-Finno-Ugric	5				
Proto-Samoyed to Proto-South-Samoyed					1
Proto-Samoyed to Proto-North-Samoyed					1
Proto-Finno-Ugric to Proto-Ugric			1		5
Proto-Finno-Ugric to Proto-Finno-Permic					1
Proto-Japanese to Old Japanese	8				
Proto-Korean to Middle Korean	4				
Proto-Tungusic to Manchu		1	1		2
Proto-Mongolic to Mongolian	13	1	1		
Proto-Turkic to Turkish	1	2			
TOTAL	**64**	**9**	**28**	**14**	**64**

Table 2: A summary of processes regulating phoneme creation and deletion.

Minimal pairs tend to increase

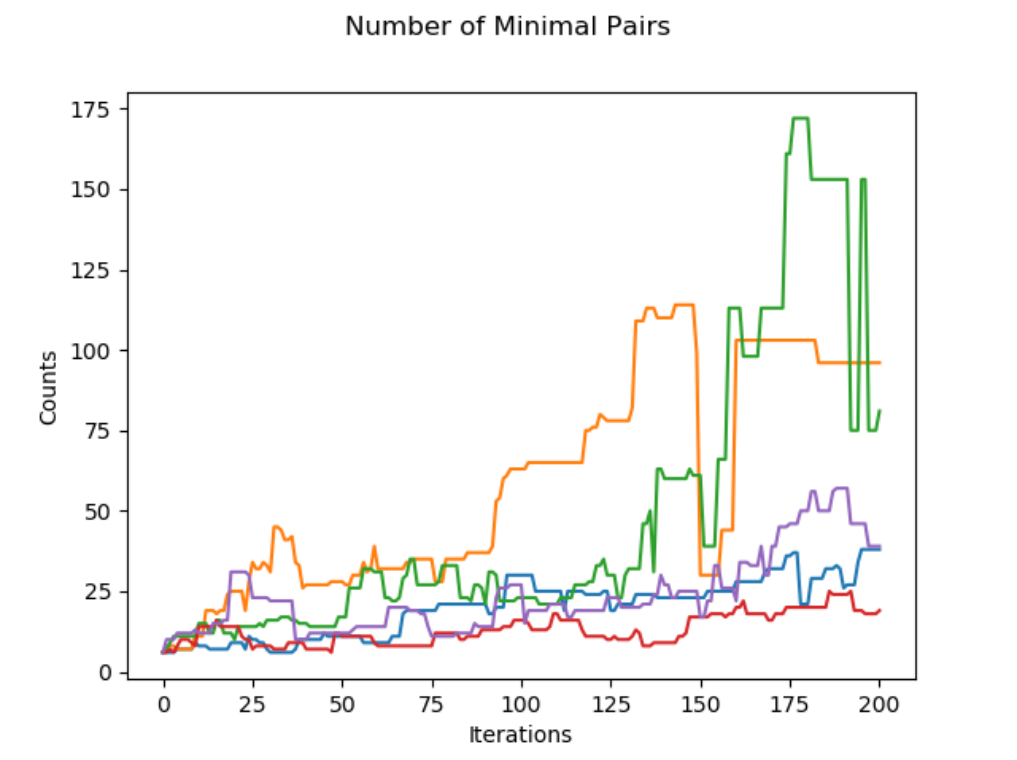


Figure 2: Minimal pair counts after 200 sound changes, in 5 parallel runs

Why?

Irreversibility of Mergers

Mergers are irreversible (Garde's principle, Labov 2010:121)

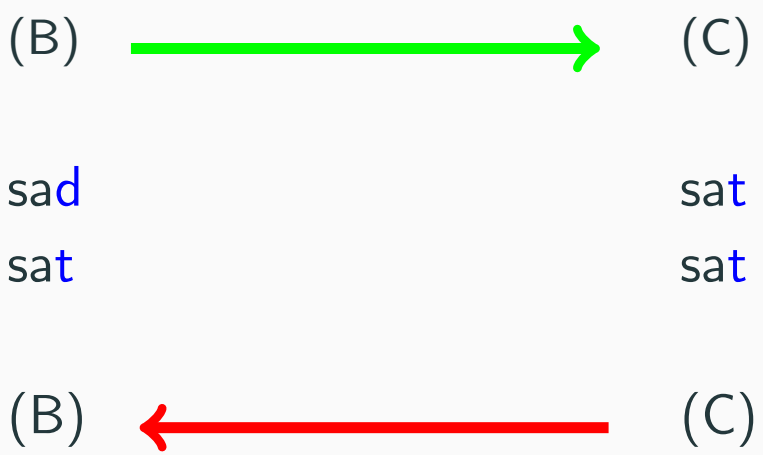
(A) → (B) /e/ > /a/ +2 Minimal Pairs (MP)

sad	sad
set	sat
far	far
her	har

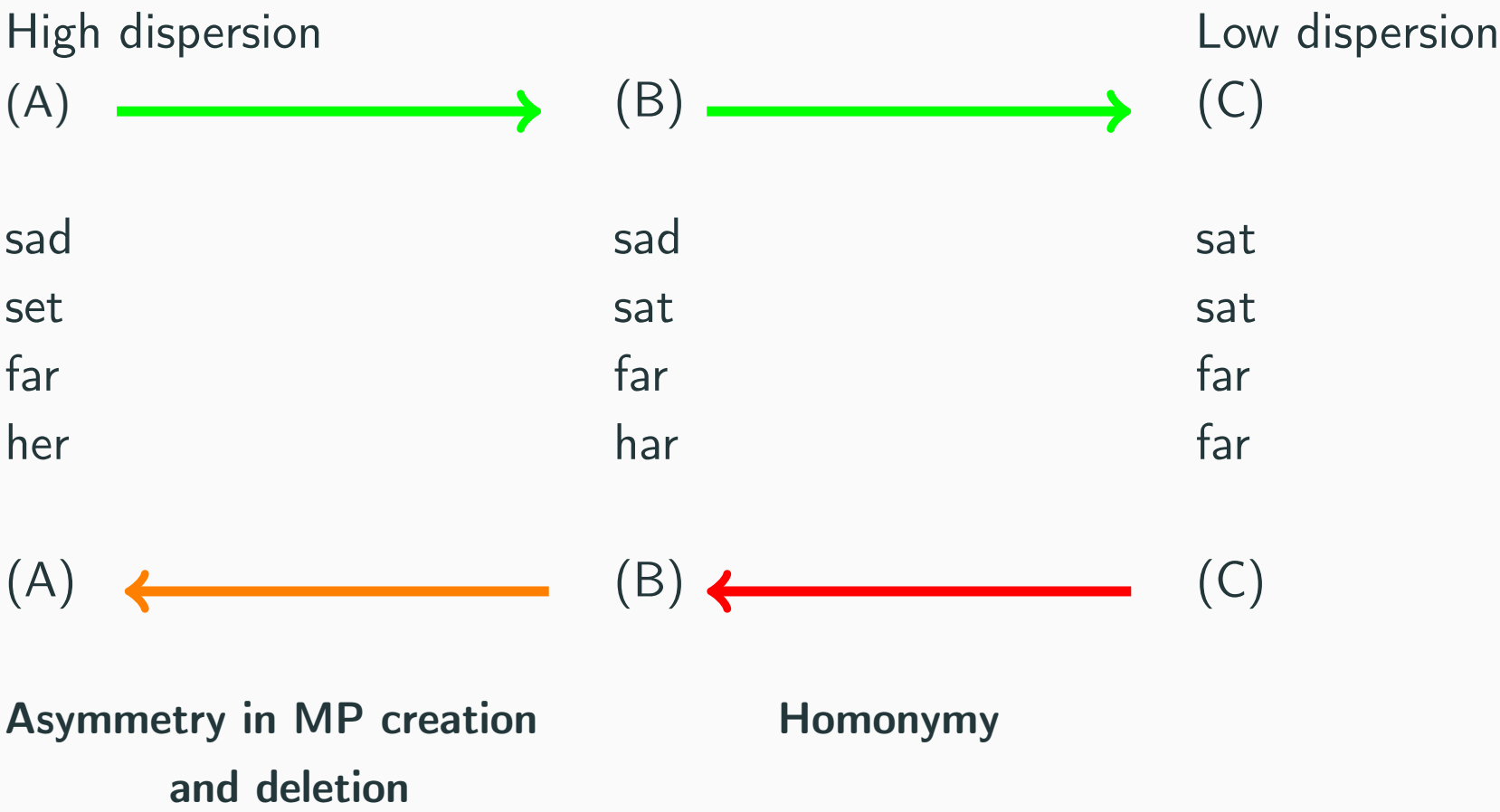
(A) ← (B) -2 MP complicated in a single step

Irreversibility of Mergers

Mergers are irreversible (Garde's principle, Labov 2010:121)



Phonemic dispersion



Is this model realistic? We can check historical data (Romance, PLEDS)

Proto-Romance > Italian

	Minimal pairs lost	Minimal pairs gained
Morphological change	34	40
Contraction	12	58
Metathesis	2	0
Secondary Split	25	0
Merger	3	102
h-deletion	6	13
TOTAL	83	213

Phoneme count is stable (27 -> 30) because of contractions.

Phonemic dispersion

Q: Is phonemic dispersion in the lexicon stable over time?

A: **NO**! Mergers and contractions cause languages to increase their number of minimal pairs. Only in languages with many splits, you get a decrease.

This does not require considerations about 'communication'. It derives from a neutral (mechanical) model.

We should turn our attention to potential balancing forces.

3. Are Mergers constrained by functional considerations?

The Functional Load Hypothesis

FUNCTIONAL LOAD

The likelihood of a merger between two phonemes is inversely proportional to the number of pairs distinguished by those two phonemes (Mathesius 1929, Martinet 1955, Jakobson 1975)

The Functional Load Hypothesis

Very few empirical investigations (King 1967, Surendran and Niyogi 2003, Wedel et al. 2013, Bouchard-Côté et al. 2013, Eychenne and Jang 2018)

Wedel et al. (2013) has been influential among linguists (Sóskuthy 2013, Kiparsky 2016, Babinsky and Bowern 2018)

Wedel et al. (2013)

'Big data' approach: they examine a large collection of mergers reported in British and American English, German, Dutch, French, Spanish, Slovak, Korean and Cantonese

They find a significant effect of Functional Load measured as the number of Minimal Pairs associated with a contrast

Frequency and Entropy are not (or barely) significant

Wedel et al. (2013)

CORPORA Minimal pairs are counted using the whole CELEX corpora. Not plausible from an acquisition viewpoint.

PHONETICS Merged pairs are evaluated against unmerged pairs with equal featural distance: e.g., /t/-/d/ is evaluated against /t/-/k/

MISTAKES Appendix of the paper reveals some mistakes

Replication of Wedel et al. (2013)

CORPORA Use CHILDES data or, in alternative, frequency threshold on CELEX

PHONETICS Add a confusability index as predictor (Wang and Bilger 1967, Hillenbrand et al. 1995, Weber and Smits 2003, Smits et al. 2003, Jouvet et al. 2015)

MISTAKES Cross-check every merger with the literature (Wells 1982, Labov et al. 2006, Wiese 2000, Kissine et al. 2003)

Data (roughly 50% of the dataset)

British English

θ-f
ð-v
θ-t
ð-d
*θ-s
*ð-z

aɪ-ɔɪ PRICE-CHOICE
ʊə-ɔ: CURE-THOUGHT
ɪə-ɛə NEAR-SQUARE
ɜ:-ɛə NURSE-SQUARE

American English

ɑ-ɔ LOT-THOUGHT
*æ-ɑ
*ɔɪ-ɚ

_l
i-ɪ HILL-HELL
ʊ-u PULL-POOL
ʊ-oʊ BULL-BOWL
ʌ-ɑ HULL-HALL (ʌ-ɔ)
ʊ-ʌ BULL-HULL

_n
ɪ-ɛ PIN-PEN

German

e-ɛ:

Dutch

s-z
f-v
χ-ʏ

Model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.48466	0.57321	-2.590	0.00959*
MinPairs	-0.35859	0.17520	-2.047	0.04068*
Phoneme Probability	0.37262	0.36551	1.019	0.30798
NoMP	-0.70144	0.70198	-0.999	0.31769
Phon. prob. by NoMP	-0.09863	0.61097	-0.161	0.87175

Table 3: Mixed-effect Logistic Regression, with Merger Set as Random effect. AIC 130.0

Model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.38577	0.65193	-3.660	0.000253*
MinPairs	-0.35312	0.16343	-2.161	0.030720*
Phoneme Probability	0.18824	0.46323	0.406	0.684473
NoMP	-0.64017	0.72233	-0.886	0.375479
ConfusionIndex	0.06586	0.01740	3.784	0.000154*
Phon. prob. by NoMP	-0.35478	0.69297	-0.512	0.608677

Table 4: Mixed-effect Logistic Regression, with Merger Set as Random effect. AIC 119.8

Model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.97224	0.52261	-5.687	<0.000001*
MinPairs	-0.15511	0.13727	-1.130	0.2585
ConfusionIndex	0.06617	0.02403	2.754	0.0059*

Table 5: Mixed-effect Logistic Regression, with Merger Set as Random effect, for Vowels

Model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.97430	0.84338	-3.527	0.000421*
MinPairs	-0.37884	0.18045	-2.099	0.035782*
ConfusionIndex	0.07983	0.03012	2.651	0.008028*

Table 6: Mixed-effect Logistic Regression, with Merger Set as Random effect, for Consonants

The Functional Load Hypothesis

Language	Merger	MP	CI (sum)	CI (max)
British English	θ-f	1	32.3	22.1
British English	ð-v	0	21.9	11.7
British English	θ-t	4	17.1	11.7
British English	ð-d	2	23.5	13.3
British English	θ-s	4	9.8	9.4
British English	ð-z	0	20.4	15.4
British English	θ-ð	0	20.4	11.0
British English	p-f	**6**	24.6	16.5
British English	b-v	0	20.8	10.6
Dutch	f-v	1	58.14	29.16
Dutch	s-z	0	39.96	20.44
Dutch	p-b	8	49.25	37.96
Dutch	t-d	**15**	90.34	85.40

Table 7: Consonant mergers

Functional Load and Sound Change

Same trend

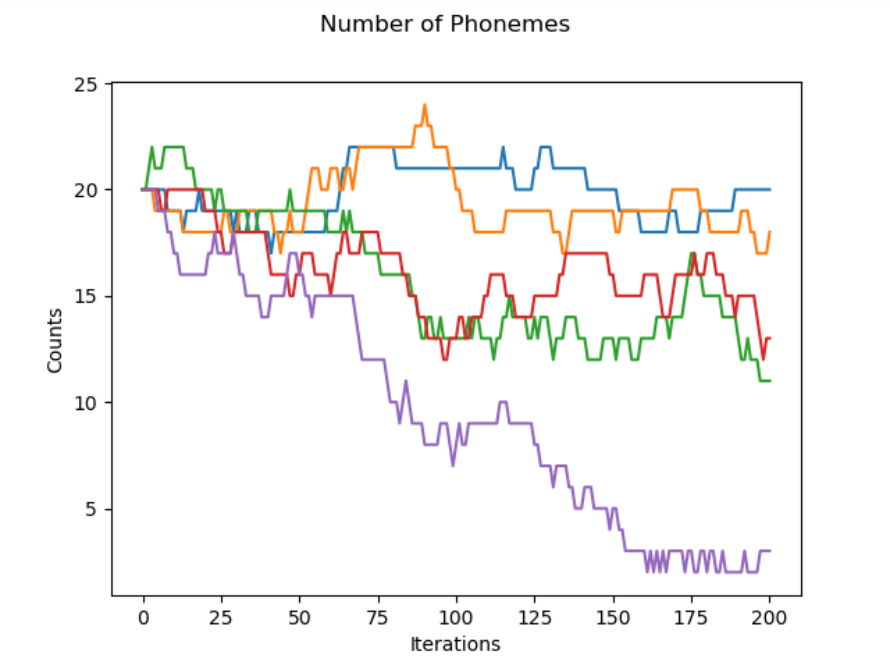


Figure 3: Phonemic counts after 200 sound changes, in 5 parallel runs

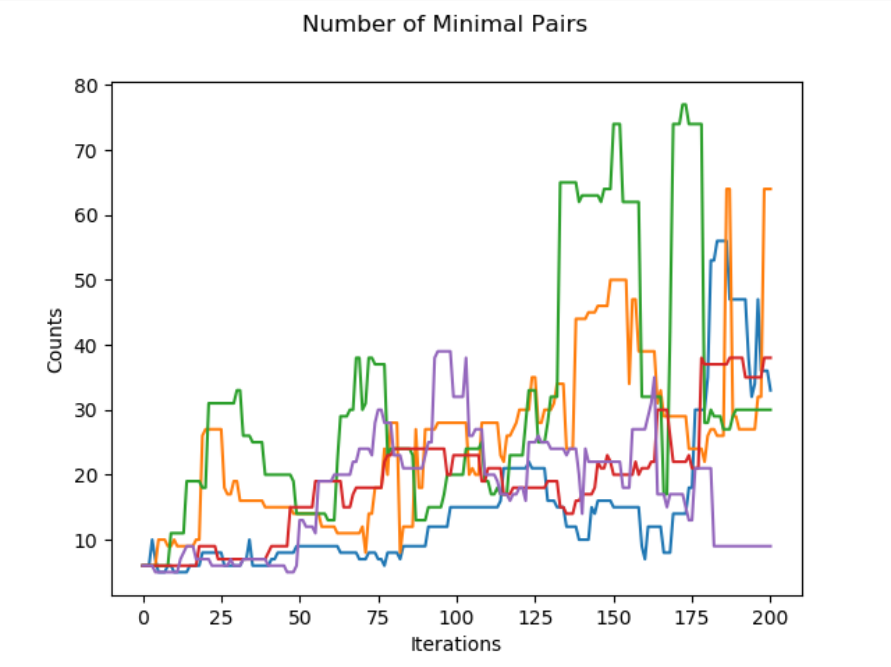


Figure 4: Minimal pair counts after 200 sound changes, in 5 parallel runs

Are Mergers constrained by functional considerations?

Q: Are Mergers constrained by functional considerations?

A: **MAYBE**...it's hard to tell.

But it is an important question, so we should investigate.

4. Do borrowing and word formation influence phonemic distributions?

Borrowing and Word Formation

Is borrowing an intractable problem? Borrowing is, essentially, gene flow.

Word formation: how much does compounding play a role compared to the creation of new words from scratch?

Borrowing and Word Formation

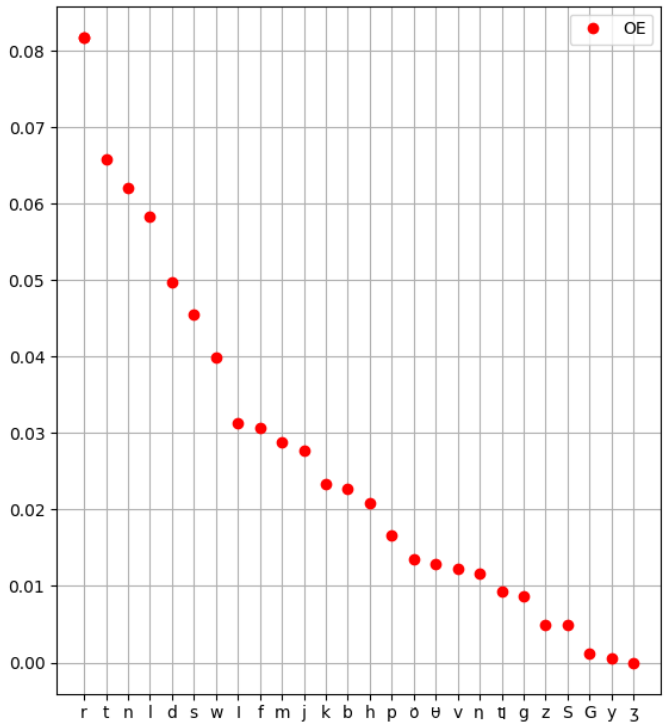
Only one model: Martin (2007), and then Graff (2012)

Borrowing and word formation are favored if they contain frequent phonemes (because of lexical access, Dell 1986) or if they contain unmarked phonemes.

Example: 'b' is uncommon in Latin, but Romance languages have many 'b'-words as effect of borrowing and word formation.

Let's check the data!

English COCA 1000 wordlist



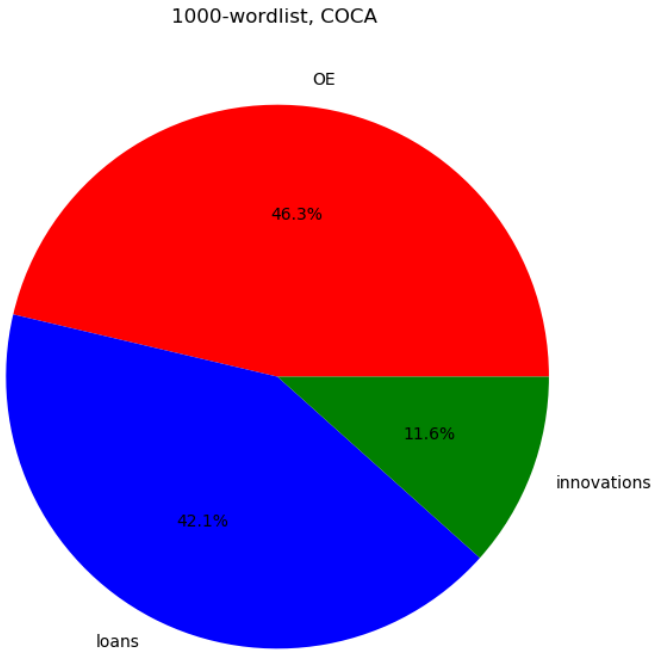


Figure 5: On the left, phoneme frequencies of English words that are native. On the right, a pie-chart of native words, loanwords and innovations in English.

English COCA 1000 wordlist

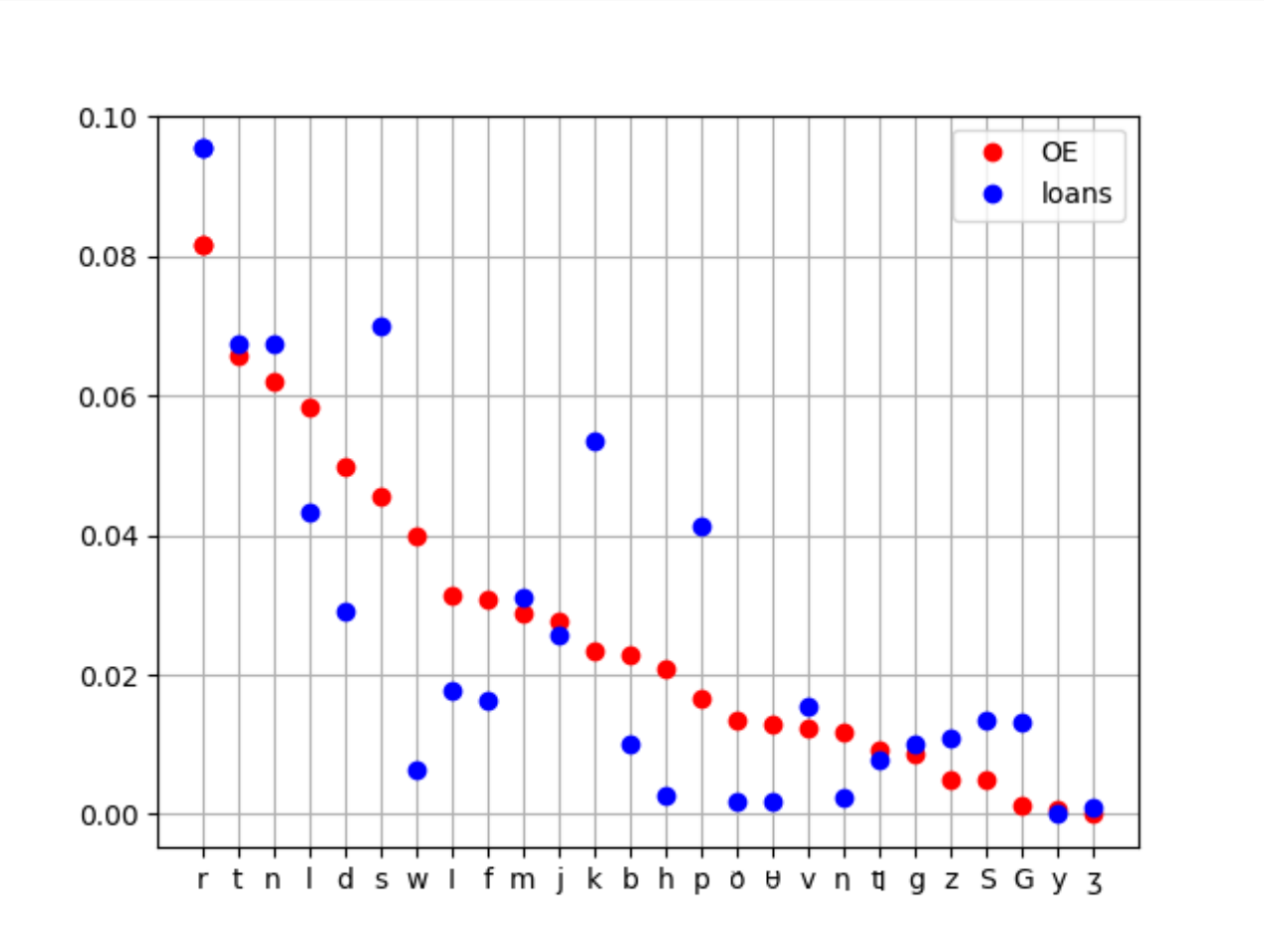
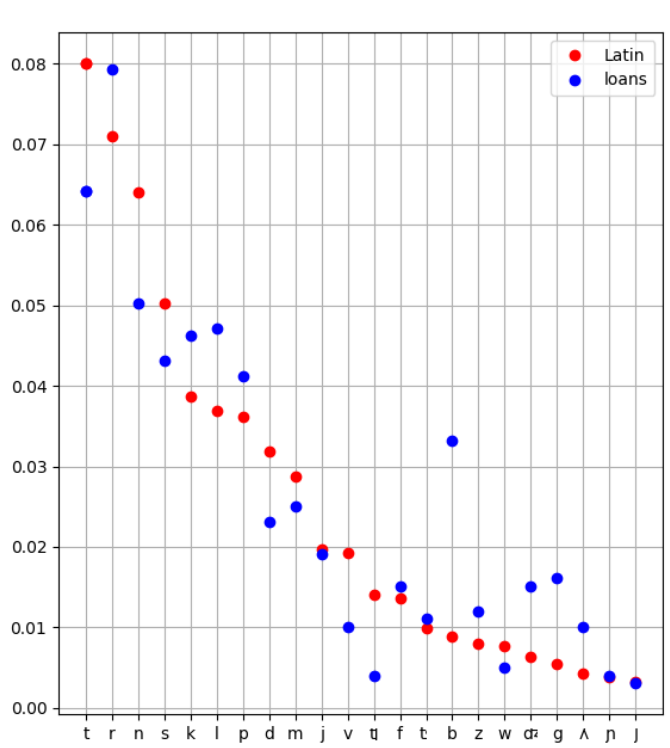


Figure 6: Phoneme frequencies of native words and loans in English.

Italian SUBTLEX 2000 wordlist



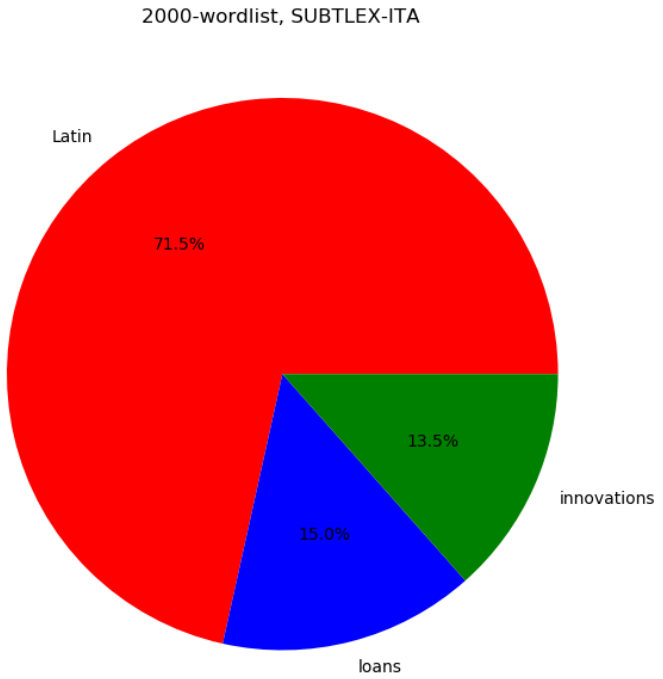


Figure 7: On the left, phoneme frequencies of native words and loans. On the right, a pie-chart of native words, loanwords and innovations in Italian.

English COCA 1000 wordlist

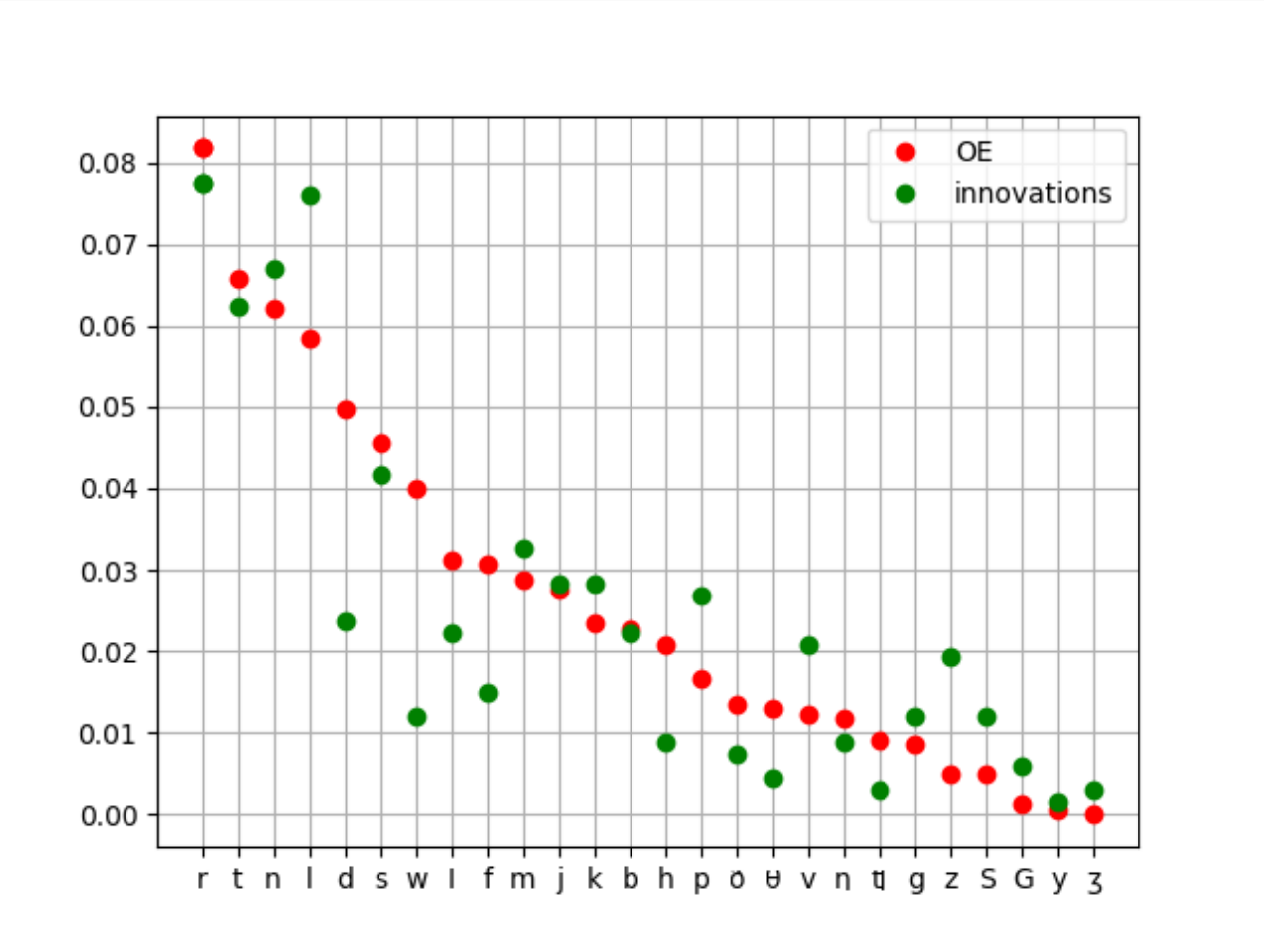


Figure 8: Phoneme frequencies of native words and innovations in English.

English COCA 1000 wordlist

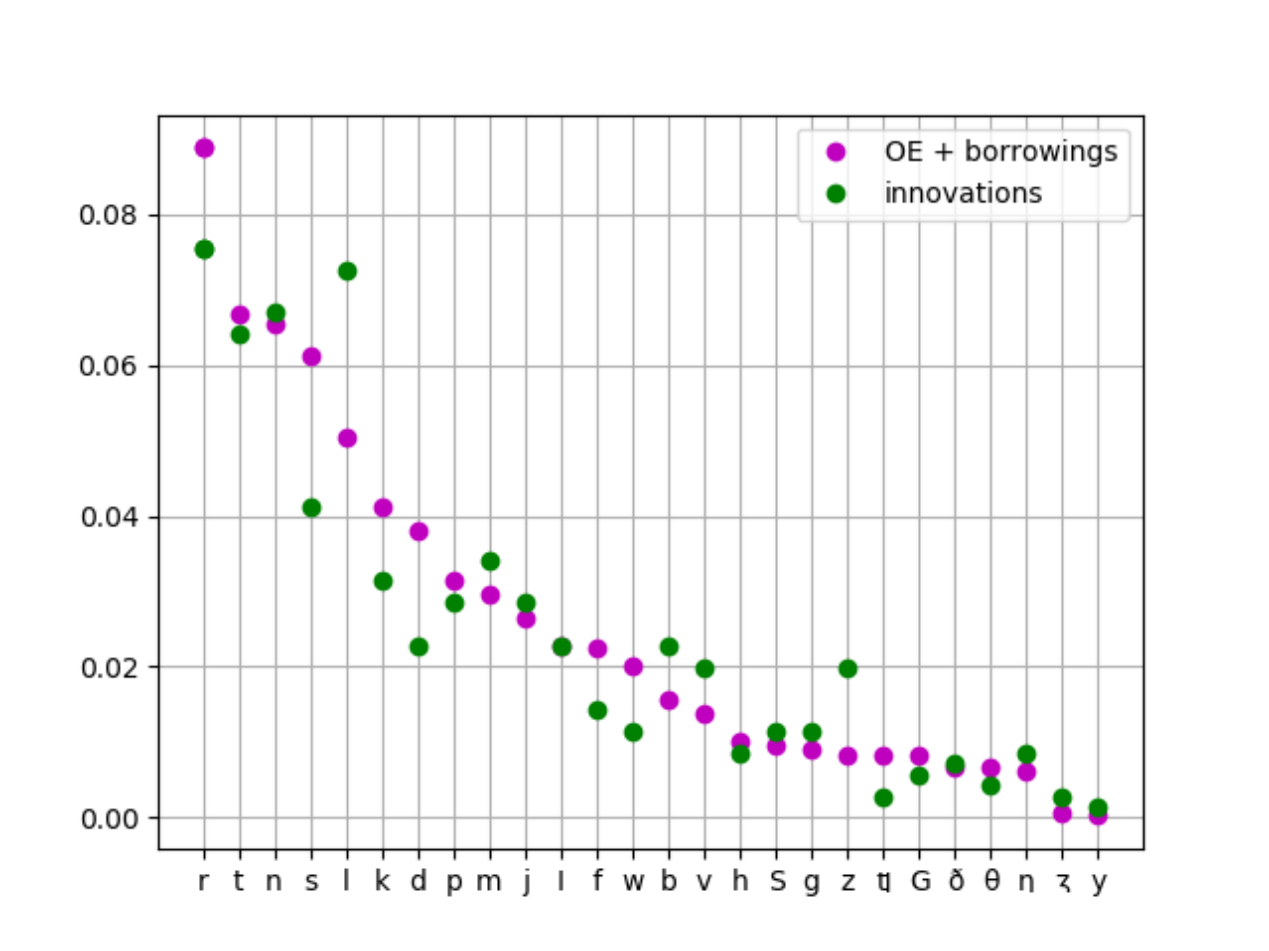


Figure 9: Phoneme frequencies of innovations in English plotted with the frequencies of the rest of the vocabulary.

Italian SUBTLEX 2000 wordlist

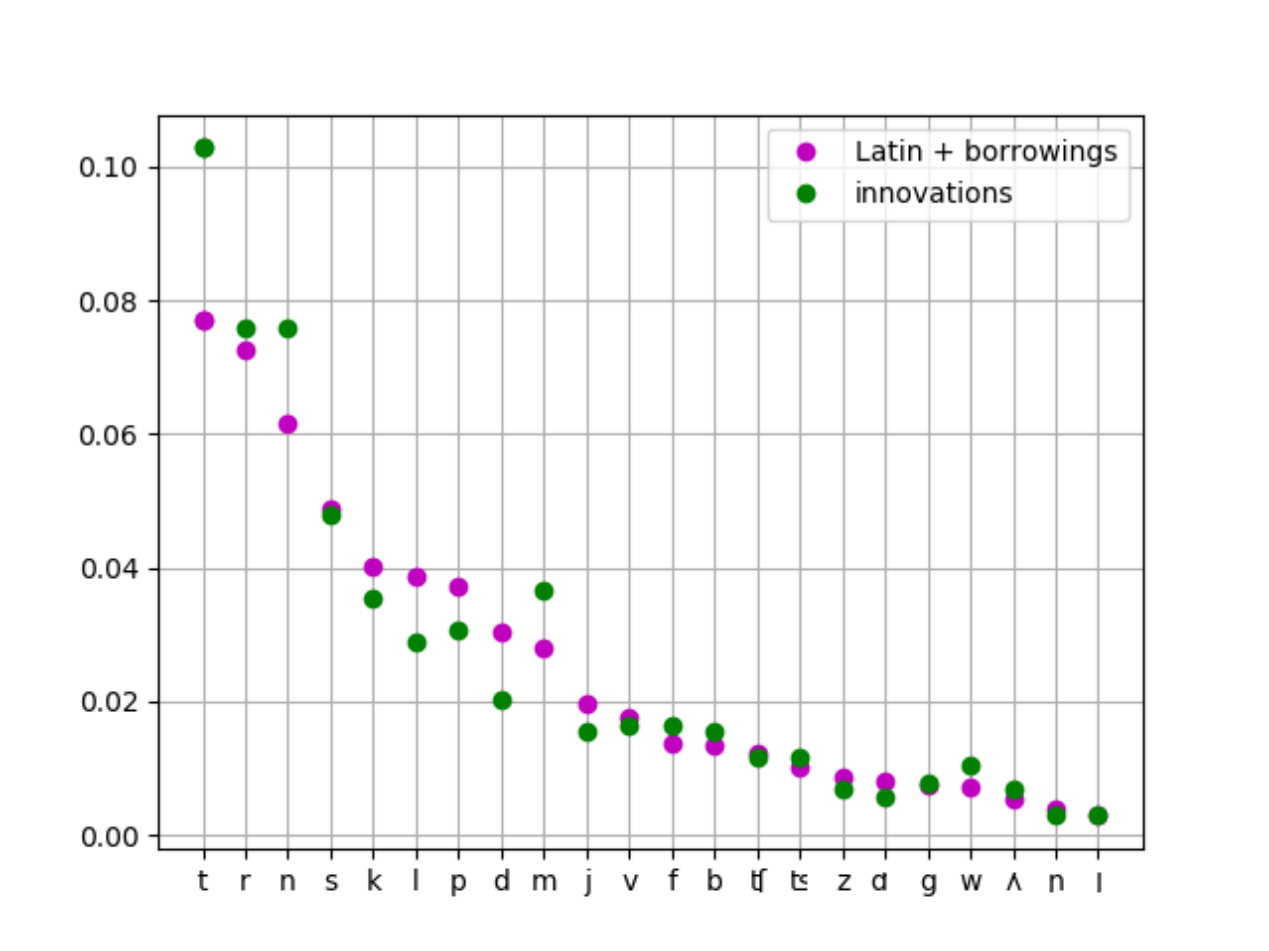


Figure 10: Phoneme frequencies of innovations in Italian plotted with the frequencies of the rest of the vocabulary.

Borrowing and Word Formation

Borrowing and word formation are two different processes

Borrowing can significantly alter the phonemic distribution of a language. Word formation does not.

Can we implement borrowing in a neutral model?

Implementing Borrowing

BORROWING

1. **memorize** one word at each generation
2. when the function is called, replace one word in the lexicon with one word drawn from the list of memorized words

Arbitrary, but close to how borrowing works (reasoning: if your grand-parents have /p/, your parents have /p/, and you have /p^h/, probably some of your cousins have /p/ too, and you can get /p/ by talking to them)

Borrowing

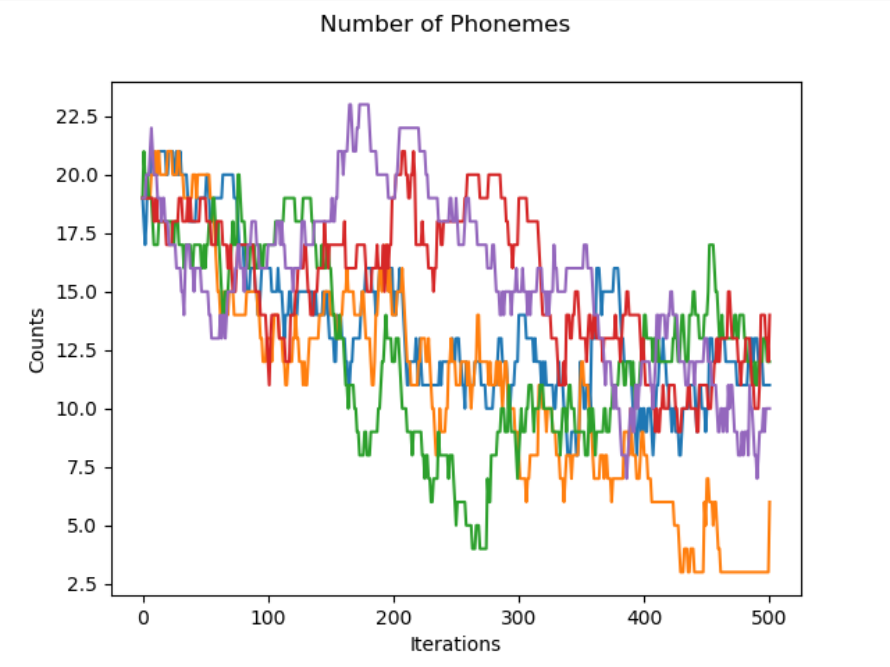


Figure 11: Phonemic counts after 500 sound changes, in 5 parallel runs

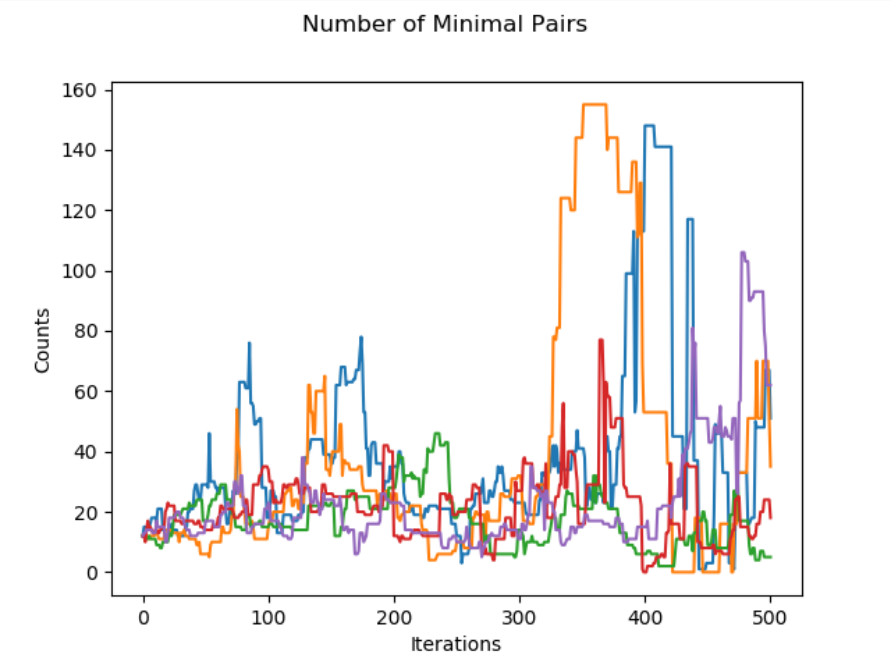


Figure 12: Minimal pair counts after 500 sound changes, in 5 parallel runs

Borrowing

The neutral model suggests two speculations.

- Speculation 1: more variation! Even if borrowing is neutral, it still has a 'functional' effect. No need for 'functional biases'
- Speculation 2: if you borrow words once in a while, no need to worry about ambiguity -> word formation less productive

Four Research Questions

FOUR TEMPORARY ANSWERS

1. Are Mergers more common than Splits? **YES**
2. Is phonemic dispersion in the lexicon stable over time? **NO**
3. Are Mergers constrained by functional considerations? **MAYBE**
4. Do borrowing and word formation influence phonemic distributions? **YES and NO**

Timetable

Date	Work
Summer 2019	Implement feature geometry for the neutral model
September 2019	Write Chapter 2 and 3
Fall-Winter 2019	Expand the analyses in Chapter 4, 5 and 6
February-March 2020	Write Chapter 4, 5 and 6
April 2020	Write Chapter 1 and Conclusions
May 2020	Defend Dissertation

Table 8: Timetable

Thank you!

Thanks to Ryan Budnick, Spencer Caplan, Aletheia Cui, Jordan Kodner, Caitlin Richter, Ollie Sayeed, David Wilson, Bob Berwick, Robin Clark, Rolf Noyer, Don Ringe, Meredith Tamminga, Charles Yang

Questions?

Naive Feature Representation

m	p	b	f	v	d	t	l	n	r	s	c	k	y	w	g	j	h
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

i	I	e	E	a	ə	O	o	U	u
0	1	2	3	4	5	6	7	8	9

Merger

Pick one position in the syllable (onset, nucleus, coda) [**ONSET**]

Select one segment in the inventory available in that position (TARGET-1) [g]

Select a second segment in the inventory between the two adjacent phonemes in terms of featural distance (TARGET-2) [w]

ONSET: get, got

m	p	b	f	v	d	t	l	n	r	s	c	k	y	w	g	j	h
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

Merger

Select a segment (TARGET-ENV) in the conditioning environment (nucleus for C, either onset or coda for V). Segments with a feature <TARGET-ENV are assigned to either the conservative or the conditioning environment depending on the directionality of the change.

E

i	I	e	E	a	ə	O	o	U	u
0	1	2	3	4	5	6	7	8	9

Merger

TARGET-1 becomes TARGET-2 in the conditioning environment.

m	p	b	f	v	d	t	l	n	r	s	c	k	y	w	g	j	h
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

i	I	e	E	a	ə	O	o	U	u
0	1	2	3	4	5	6	7	8	9

get, god -> wet, god

Samoyed

Proto-Uralic to Proto-Samoyed, consonant inventory: for Proto-Uralic 17 consonants can be reconstructed, but internal reconstruction from Samoyed languages points to only 13 segments. Two phonemes which were probably representing fricatives, /d/ and /d'/, merge in all positions with /r/ and /j/. The same happens with the phoneme /s'/, which merges to the unmarked /s/. The status of Proto-Uralic /x/ is less clear, and therefore we leave it out. Overall, it is clear that at least three phonemes were lost as a consequence of mergers.

Proto-Uralic to Proto-Samoyed, vowel inventory: Proto-Uralic had a eight vowel system, with four high vowels. Two new vowels, /ə́/ and /ɵ/, resulted from high vowels splits. So we have splits adding two phonemes to the inventory. The presence of a tentative phoneme /ö/ is also reported, but its sources are not clear.

Uralic

Proto-Uralic to Proto-Finno-Ugric, consonant inventory: the inventory is stable. Internal reconstruction point to three new phonemes, /š/, /c'/ and /l'/, whose presence is postulated only for words that have no cognates with other branches, and therefore their origin is not clear.

Proto-Uralic to Proto-Finno-Ugric, vowel inventory: contractions are reported from the interaction of Proto-Uralic /x/ with five different vowels, so that the vowel inventory of Proto-Finno-Ugric is reconstructed with 13 total vowels. In this case, contractions resulted in five new phonemes.

Proto-Samoyed to Proto-South-Samoyed, consonant inventory: no changes.

Uralic

Proto-Samoyed to Proto-South-Samoyed, vowel inventory: a merger affected the two back vowels /ä/ and /å/, therefore yielding only one back vowel for Proto-South-Samoyed. A second merger is reported for /ö/, but since the status of the vowel in Proto-Samoyed is not clear, this merger is discarded. We then consider only one merger for this branch.

Proto-Samoyed to Proto-North-Samoyed, consonant inventory: /c/ and /t/ merge in all environments.

Proto-Samoyed to Proto-North-Samoyed, vowel inventory: no changes.

Ugric and Finno-Permic

Proto-Finno-Ugric to Proto-Ugric, consonant inventory: /s/ and /š/ merge into /θ/.

Proto-Finno-Ugric to Proto-Ugric, vowel inventory: four of the long vowels (/uu/, /oo/, /ɨɨ/, /ee/) are merged. Interestingly, the merger between /uu/ and /u/ has the consequence that the allophonic split between /u/ and /ŭ/ is phonemicized. Here, we have four mergers and one split in total.

Proto-Finno-Ugric to Proto-Finno-Permic, consonant inventory: no changes.

Proto-Finno-Ugric to Proto-Finno-Permic, vowel inventory: /ɨɨ/ merge into /oo/.

Japanese

Proto-Japanese to Old Japanese, consonant inventory: The main difference between the consonant inventory of Proto-Japanese and Old Japanese is that internal comparison would not allow to reconstruct voiced obstruents for Proto-Japanese. The fact that in Old Japanese we find the voiced obstruents /b/, /d/, /z/ and /g/, but they do not appear in word-initial position, can be explained through a development of consonant clusters with nasals (*np, *nt, *ns, *ng) into voiced obstruents. In this case, contractions resulted in four new phonemes.

Proto-Japanese to Old-Japanese, vowel inventory: While Old Japanese has eight vowels, scholars agree that only four vowels (/a/, /o/, /i/, /u/) should be reconstructed for Proto-Japanese. Old Japanese developed one high central vowel and three new mid vowels as a consequence of vowel hiatus. Also in this case, contractions resulted in four new phonemes.

Korean

Proto-Korean to Middle Korean, consonant inventory The Middle Korean alphabet encodes many phonemes that cannot be reconstructed for Proto-Korean. While Robbeets acknowledges that the three phonemes associated to voiced fricatives might just be allophones of the obstruents, there is solid evidence that the aspirated obstruents /p^h/, /t^h/, /c^h/ and /k^h/ all developed from consonant clusters *pk, *tk, *ck and *kk. This would imply that also for Middle Korean, we have four new phonemes resulting from contraction.

Proto-Korean to Middle Korean, vowel inventory The question of the phonetic nature of Proto-Korean seven-vowel system is debated, but there is an agreement that the seven-vowel system of Middle Korean has a direct reflex in Proto-Korean, either as a direct descent or as a result of a vowel shift.

Manchu

Proto-Tungusic to Manchu, consonant inventory The phonemic inventories of Tungusic languages are quite similar, and therefore reconstruction is not controversial. The reference here is Benzing (1955). If we follow the development into Manchu, we would find a /p/ and /f/ contrast that is not reconstructed for Proto-Tungusic: since a change /p/ > /f/ is reported in the history of Manchu, modern /p/ can be explained through dialectal contact with other Tungusic languages in which the sound change did not occur. Another feature of modern Manchu is the contrast between /s/ and /ʃ/, which following Robbeet's report resulted from a secondary split.

Proto-Tungusic to Manchu, vowel inventory An eight-vowel system is reconstructed for Proto-Tungusic, but modern Manchu is described as a six-vowel system, as a result of the loss of a length contrast in /o/ and /u/.

Mongolian

Proto-Mongolic to Mongolian, consonant inventory The reconstruction follows Poppe (1954) and Svantesson (2005). A large number of contractions affected the consonant system. Proto-Mongolic lacked /w/ and /ʃ/, which entered the language as allophonic splits from *p and *s, and in the second case this might have been facilitated by the presence of loanwords. The majority of the phonemes are associated to the palatalized versions of the obstruents (eight), the nasals (two), the liquids (two) and the approximant /w/, which contrary to Proto-Mongolic exhibit contrast with their non palatalized version before pharyngeal vowels, and therefore add a total of 13 phonemes to the language.

Proto-Mongolic to Mongolian, vowel inventory The only point of debate is whether the modern Mongolian seven-vowel system can be traced back to Proto-Mongolic, or if the proto-language had a length contrast in /i/ that was lost. For this reason, no sound change is reported in this case.

Turkish

Proto-Turkic to Turkish, consonant inventory The reconstruction of Proto-Turkic is more unclear than the others. In particular, there is no agreement about whether Turkish /z/ and /ʃ/ result from secondary splits of reconstructed liquid consonants or from some contractions. The only patterns which are uncontroversial are the emergence of /h/ as a result from a secondary split of /p/ in word initial position, while /f/ and /ʒ/ entered the language through loanwords.

Proto-Turkic to Turkish, vowel inventory The eight-vowel Turkish system can be traced back to Proto-Turkic.

Proto-Romance to Italian

	Proto-Romance	Italian
Morphological Change		
Neuter -> Feminine	**RAPU** [rapu] 'turnip'	rapa [rapa]
	RAMU [ramu] 'branch'	ramo [ramo]
3rd declension -> 2nd declension	**LAVŌRE** [lavore] 'job'	lavoro [lavoro]
	LAVARE [lavare] 'to wash'	lavare [lavare]
4th declension -> 2nd declension	CORPUS [kɔrpus] 'body'	**corpo** [kɔrpo]
	CORNU [kɔrnu] 'horn'	**corno** [kɔrno]
Split		
/k/->/tʃ/	**CANTU** [kantu] 'song'	canto [kanto]
	CĒNTU [kɛntu] 'hundred'	cento [tʃɛnto]
/ɛ/-/je/ in V́	**VĒTARE** [vɛtare] 'to prohibit'	vietare [vjetare]
	VŌTARE [votare] 'to vote'	votare [votare]
/ɔ/->/wo/ in V́	**SANO** [sano] 'healthy'	sano [sano]
	SŌNO [sɔno] 'sound'	suono [swono]
Contraction		
/lj/ -> /ʎʎ/	**PALI̯A** [palja] 'straw/hay'	paglia [paʎʎa]
	PALMA [palma] 'palm tree'	palma [palma]
/gl/ -> /ʎʎ/	TALI̯A [talja] 'size'	**taglia** [taʎʎa]
	TEGLA [tɛgla] 'pan'	**teglia** [teʎʎa]
/dj/ and /j/ -> /ddʒ/	MAI̯US [majus] 'May'	**maggio** [maddʒo]
	RADI̯UM [radjum] 'ray'	**raggio** [raddʒo]
/aw/ -> /o/	PAUSARE [pawsare] 'to pause'	**posare** [pozare]
	PESARE [pesare] 'to weight'	**pesare** [pezare]

Table 9: Minimal pairs gains and losses between Proto-Romance and Italian.

Proto-Romance to Italian

	Proto-Romance	Italian
Merger		
/ɪ/ -> /e/	BELVA [belva] ‘beast’	**belva** [belva]
	SILVA [sɪlva] ‘forest’	**selva** [selva]
/ʊ/ -> /o/	COLPU [kolpu] ‘shot/blow’	**colpo** [kolpo]
	CULPA [kʊlpa] ‘fault’	**colpa** [kolpa]
unstressed /ɔ/ -> /o/	VŎLARE [volare] ‘to fly’	**volare** [volare]
	VŌTARE [votare] ‘to vote’	**votare** [votare]
unstressed /ɛ/ -> /e/	GĔLARE [dʒɛlare] ‘freeze’	**gelare** [dʒelare]
	BĒLARE [belare] ‘to baa’	**belare** [belare]
/ct/ -> /tt/	CATTU [kattu] ‘cat’	**gatto** [gatto]
	PACTU [paktu] ‘pact’	**patto** [patto]
/ps/ -> /ss/	CAPSA [kapsa] ‘box’	**cassa** [kassa]
	PASSA [passa] ‘withered/pass-3SG’	**passa** [passa]
h-deletion		
/h/ -> /∅/	**HORTU** [hɔrtu] ‘garden’	orto [orto]
	MORTU [mortu] ‘dead’	morto [morto]

Table 10: Minimal pairs gains and losses between Proto-Romance and Italian.

Minimal Pairs

British English

θ-f *three/free*

ai-oi *tie/toy*, *boy/bye*, *boy/by*, *boy/buy*

ɜ:-ɛə *her/hair*, *where/wear*, *where/were*

American English

i-ɪ *feel/fill*, *wheel/will*

ɑ-ɔ *talk/tock*

Borrowed words in English

Borrowed words, English: **ability**, **able**, **accept**, **across**, **act**, **action**, **activity**, **add**, **administration**, **adult**, **affect**, **age**, **agency**, **agree**, **air**, **allow**, **american**, **amount**, **analysis**, **animal**, **appear**, **apply**, **approach**, **area**, **argue**, **arrive**, **art**, **article**, **artist**, **attack**, **attention**, **author**, **authority**, **avoid**, **bank**, **bar**, **base**, **benefit**, **big**, **billion**, **blue**, **call**, **camera**, **campaign**, **candidate**, **car**, **career**, **carry**, **case**, **catch**, **cause**, **cell**, **center**, **central**, **century**, **certain**, **chair**, **challenge**, **chance**, **change**, **character**, **charge**, **check**, **choice**, **city**, **claim**, **class**, **clear**, **close**, **college**, **color**, **common**, **community**, **company**, **compare**, **concern**, **condition**, **conference**, **congress**, **consider**, **contain**, **continue**, **control**, **cost**, **country**, **couple**, **course**, **court**, **cover**, **create**, **crime**, **culture**, **current**, **data**, **decade**, **decide**, **decision**, **defense**, **degree**, **democrat**, **democratic**, **describe**, **design**, **detail**, **determine**, **develop**, **die**, **difference**, **different**, **direction**, **director**, **discover**, **discuss**, **disease**, **doctor**, **drug**, **easy**, **economic**, **economy**, **education**, **effect**, **effort**, **election**, **energy**, **enjoy**, **enter**, **entire**, **establish**, **event**, **evidence**, **example**, **executive**, **expect**, **experience**, **expert**, **explain**, **face**, **fact**, **factor**, **fail**, **family**, **federal**, **figure**, **final**, **fine**, **finish**, **firm**, **focus**, **force**, **foreign**, **form**, **fund**, **future**, **general**, **get**, **give**, **government**, **group**, **guy**, **history**, **hit**, **hospital**, **hour**, **huge**, **human**, **husband**, **idea**, **identify**, **image**, **imagine**, **impact**, **important**, **improve**, **include**, **including**, **increase**, **indicate**, **individual**, **industry**, **information**, **institution**, **interest**, **international**, **interview**, **involve**, **issue**, **join**, **just**, **kid**, **language**, **large**, **law**, **leg**, **legal**, **letter**, **level**, **list**, **local**, **low**, **main**, **maintain**, **major**, **manage**, **market**, **matter**, **measure**, **media**, **medical**, **member**, **memory**, **mention**, **message**, **military**, **million**, **minute**, **moment**, **money**, **move**, **movement**, **music**, **nation**, **national**, **natural**, **nature**, **necessary**, **nice**, **note**, **number**, **occur**, **office**, **officer**, **official**, **oil**, **operation**, **opportunity**, **order**, **page**, **pain**, **paper**, **parent**, **part**, **particular**, **party**, **pass**, **patient**, **pay**, **peace**, **people**, **per**, **perform**, **period**, **person**, **personal**, **physical**, **pick**, **picture**, **piece**, **place**, **plan**, **point**, **police**, **policy**, **political**, **poor**, **popular**, **population**, **position**, **possible**, **power**, **practice**, **prepare**, **present**, **president**, **pressure**, **price**, **private**, **problem**, **process**, **produce**, **product**, **production**, **professor**, **program**, **project**, **property**, **protect**, **prove**, **provide**, **public**, **push**, **quality**, **question**, **quite**, **race**, **raise**, **range**, **rate**, **real**, **reason**, **receive**, **recent**, **recognize**, **record**, **reduce**, **region**, **religious**, **remain**, **remember**, **remove**, **report**, **represent**, **require**, **research**, **resource**, **respond**, **response**, **rest**, **result**, **return**, **reveal**, **risk**, **rock**, **role**, **rule**, **same**, **save**, **scene**, **school**, **science**, **season**, **seat**, **second**, **section**, **security**, **seem**, **sense**, **series**, **serious**, **serve**, **service**, **several**, **sex**, **sign**, **significant**, **similar**, **simple**, **single**, **sister**, **site**, **situation**, **size**, **skill**, **social**, **society**, **sort**, **sound**, **source**, **space**, **special**, **specific**, **stage**, **standard**, **state**, **station**, **stay**, **store**, **story**, **structure**, **student**, **study**, **stuff**, **style**, **subject**, **success**, **suggest**, **support**, **sure**, **system**, **table**, **take**, **tax**, **term**, **test**, **their**, **them**, **theory**, **they**, **though**, **treat**, **trial**, **trip**, **trouble**, **try**, **type**, **use**, **value**, **very**, **view**, **visit**, **voice**, **wait**, **want**, **war**, **window**, **wrong**

Innovations in English

Innovations, English: according, actually, against, already, although, always, another, around, as, available, baby, bad, beautiful, because, behavior, bill, boy, card, certainly, clearly, computer, cultural, cut, despite, development, difficult, during, employee, environment, environmental, especially, evening, everybody, everything, exactly, finally, financial, girl, goal, growth, guess, gun, happy, herself, however, inside, its, itself, job, kill, later, lawyer, likely, management, manager, material, maybe, mister, morning, movie, myself, nearly, network, news, not, ok, or, organization, outside, particularly, past, performance, perhaps, phone, politics, pretty, probably, radio, realize, really, recently, relationship, republican, run, share, she, simply, since, someone, sometimes, sport, statement, suddenly, talk, technology, television, themselves, these, those, throughout, traditional, training, treatment, tv, unit, until, usually, various, whose, yourself

Borrowed words in Italian

Borrowed words, Italian: affare, angelo, appartamento, appuntamento, aria, arriva, arrivo, assassino, auto, autobus, bagno, balla, ballo, banca, bar, barca, base, basta, bastardo, bianco, bicchiere, biglietto, birra, bisogna, bisogno, blu, bordo, borsa, bottiglia, braccio, bravo, caffé, calma, camera, camion, carne, carriera, carta, cavallo, cavolo, cazzo, cinema, club, colazione, college, colpo, compleanno, computer, controlla, controllo, coraggio, detective, dettagli, diavolo, dipartimento, dna, dollari, drink, droga, energia, episodio, eroe, fantasma, fianco, film, foto, francese, funziona, furgone, gamba, gatto, gay, genio, giro, grandioso, gruppo, guai, guarda, guerra, guida, hotel, idiota, inglese, liceo, lista, macchina, magari, magia, mangia, metri, milioni, nord, okay, ore, orologio, palla, pantaloni, papa, papá, parco, passaggio, pausa, pensiero, periodo, pezzo, pista, pistola, politica, polizia, poliziotto, pratica, prigione, problema, progetto, programma, puttana, ragazzo, razza, reale, regalo, responsabile, rischio, ristorante, roba, sabato, sala, sbagliato, sbaglio, scarpe, scena, sceriffo, scherza, scherzo, schiena, schifo, scopo, scuola, sembra, sergente, sexy, show, sindaco, sistema, stanco, storia, stronzo, sud, taxi, telefono, teoria, tesoro, test, tipo, treno, troppo, turno, viaggio, villaggio, zio, zona

Innovations in Italian

Innovations, Italian: abbastanza, accanto, accusa, addosso, aereo, affatto, affronta, alcun, almeno, altrimenti, alzati, arrabbiato, arrivederci, assistente, attacco, attraverso, bambino, bellezza, benvenuto, bomba, buco, buonanotte, buonasera, buongiorno, campione, cellulare, ciao, cioé, colonnello, combatte, comincia, compagnia, comunque, conta, contratto, corsa, davanti, davvero, destino, doccia, emergenza, entrambi, eppure, fanculo, ferita, fidanzato, finché, finta, fretta, gara, ghiaccio, giornale, giornata, glielo, guardia, importanza, impronte, incinta, indietro, indirizzo, infatti, infermiera, innamorato, insomma, intorno, invece, laggiú, livello, malato, malattia, manca, mantenere, mentre, meraviglioso, mossa, neanche, nemmeno, nessun, nessuna, nonostante, occhiata, offerta, ognuno, oppure, ormai, ovunque, partita, passato, pazzo, perché, persino, piccolo, piuttosto, pomeriggio, posto, potere, presa, purtroppo, ricerca, ricordo, riesce, riguarda, riguardo, ringrazio, ritorno, riunione, riuscire, scelta, scommette, scomparsa, scoperto, scopre, semplicemente, sentimenti, serata, sguardo, sicurezza, sistema, situazione, smette, soldato, soltanto, soprattutto, sorpresa, spara, spiace, spiaggia, squadra, stamattina, stampa, stanotte, stasera, stavolta, sveglia, taglio, telefonata, tiro, tranne, troia, tv, video, visita, volo, zitto

Graff 2012

'Any mechanism that derives communicative efficiency from effects on individual sounds in phonological context only, will necessarily predict that distinctions between words will follow directly from the distributions of individual sounds' (Graff 2012:125)

This is not true if we assume that the likelihood of a merger is directly proportional to phonetic confusability. This should make minimal pairs between highly confusable segments more likely to disappear in time (because of homonymy production) regardless of frequency considerations.

Surendran and Niyogi 2003

Is local predictability different that overall predictability (=frequency)?

'Surprisingly, there has been no mention, let alone study, of this problem in the functional load literature' (Surendran and Niyogi 2003)

'We computed the functional load of all consonantal oppositions in English with k=0 and k=3 [...] The correlation is above 0.9 ($p << 0.001$), indicating that one is quite predictable from the other. This is surprising, since the k=0 model does not use any context at all, and is simply based on phoneme frequencies' (Surendran and Niyogi 2003)

Cohen Priva 2017

'Low informativity predicts word-final lenition of all obstruents in American English after phonetic and phonological factors were controlled for' (Cohen Priva 2017:569). What about frequencies?

'Segment local predictability was residualized using both segment probability and informativity. Thus, it would only be significant if it improved the model beyond the (unconstrained) effect that the variables it is residualized over have' (Cohen Priva 2017:585)

'A segment's probability has no residual effect on its likelihood to delete: it is not the case that frequent segments are necessarily under more pressure to be lenited than infrequent ones [...] The lack of effect is due to the inclusion of informativity: informativity explains all of the variance that the segment probability could explain, making segment probability redundant' (Cohen Priva 2017:586) What about the other way around?

Variational Model (Yang 2009)

It can be used to predict 'Impossible' mergers

Minor fixes to the equation to deal with homonymy

don-1 (3), *don-2* (2), *dawn* (1)

1. put the frequencies of the potential homonyms in a list X: [3, 2, 1]
2. H = sum(X)
3. c_+ = 1 - max(X) / H

Variational Model (Yang 2009)

- c_{-} = 1/H * $\sum$ p0[(1-e2)h1 + e2h2] + [q0*(e1h1+e2h2)]
- c_{-} = 1/H * $\sum$p0[(1-e2)h1 + e2h2] + [(1-p0)*(e1h1+e2h2)]
- c_{-} = 1/H * $\sum$p0[h1 - e2h1 + e2h2] + [e1h1+e2h2 - p0e1h1 - p0e2h2]
- c_{-} = 1/H * $\sum$p0[h1 - e2h1 + e2h2] + [e1h1+e2h2 - p0e1h1 - p0e2h2]
- c_{-} = 1/H * $\sum$p0h1 - p0e2h1 + **p0e2h2** + e1h1 + e2h2 - p0e1h1 - **p0e2h2**
- c_{-} = 1/H * $\sum$p0h1 - p0e2h1 + e1h1 + e2h2 - p0e1h1
- c_{-} = 1/H * $\sum$p0h1 - p0e2h1 - p0e1h1 + e1h1 + e2h2
- c_{-} = 1/H * $\sum$p0h1(1 - e2 - e1) + e1h1 + e2h2
- c_{-} = 1/H * (p0 $\sum$ h1(1 - e2 - e1) + $\sum$ e1h1 + e2h2)
- m = 1/H * $\sum$ h1(1 - e2 - e1)
- k = 1/H * $\sum$ e1h1 + e2h2

Variational Model (Yang 2009)

- put all the homonyms in vectors (X: [don-1, don-2], Y:[dawn])
- m = 1/H * $\sum$ h1(1 - e2 - e1)
- k = 1/H * $\sum$ e1h1 + e2h2
- in case of homonyms, the least frequent meanings are **never retrieved without misunderstanding**
- **f = 1/H * the sum of all the h of words which are not at the top of the vectors. h1 and h2 are now intended as the frequencies of the words at the top of the vectors for the minimal pair**
- c_{-} = m * p0 + k + **f**

Variational Model (Yang 2009)

- *don-1* (30), *don-2* (1), *dawn-1* (1), *dawn-2* (0) **p0=−0.036**
- *don-1* (6), *don-2* (5), *dawn-1* (1), *dawn-2* (0) **p0=0.107**
- *don-1* (6), *don-2* (5), *don-3* (4), *dawn-1* (1) **p0=0.107**
- *don-1* (3), *don-2* (2), *dawn-1* (1), *dawn-2* (0) **p0=0.286**
- *don-1* (6), *don-2* (5), *dawn-1* (4), *dawn-2* (3) **p0=0.643**
- *don-1* (6), *don-2* (5), *dawn-1* (6), *dawn-2* (5) **p0=0.999**
- *don-1* (1), *don-2* (1), *dawn-1* (1), *dawn-2* (1) **p0=0.999**