

Phonemic dispersion in the lexicon decreases over time because of sound change

Andrea Ceolin - University of Pennsylvania
ESHPP - December 9th, 2019

1. Introduction

1.1 Sound change and phonemic dispersion in the lexicon

1.2 This talk

2. Methods

2.1 A model of sound change

3. Results

3.1 Addressing Labov's paradox

3.2 Phonetics

3.3 Other sound changes

3.4 Anti-homonymy

3.5 Lexical Replacement

4. Conclusions

Section 1

1. Introduction

1.1 Sound change and phonemic dispersion in the lexicon

1.2 This talk

2. Methods

2.1 A model of sound change

3. Results

3.1 Addressing Labov's paradox

3.2 Phonetics

3.3 Other sound changes

3.4 Anti-homonymy

3.5 Lexical Replacement

4. Conclusions

Are phonemic inventories shrinking?

- '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)
- Mergers = —
- Splits = +
- Mergers > Splits
- — > + ??

Phonemic dispersion in the lexicon is decreasing

- Dautriche I., Mahowald K., Gibson E., Christophe A., Piantadosi S. **Word cluster phonetically beyond phonotactic regularities.** Cognition 163 (2017) 128-145
- Model: 5-gram language model trained on CELEX phonological transcriptions.
- Used to generate pseudo-words and study phonemic dispersion in a real and a pseudo-lexicon.

Phonemic dispersion in the lexicon is decreasing

- '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 [...]' (Dautriche et al. 2017: 143)
- Potential explanations: ease of production, ease of learnability, memory...
- '[...] 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)

Goal of this talk

1. Investigate Labov's paradox
2. Understand how sound change relates to clumpiness/phonemic dispersion in the lexicon

Section 2

1. Introduction

1.1 Sound change and phonemic dispersion in the lexicon

1.2 This talk

2. Methods

2.1 A model of sound change

3. Results

3.1 Addressing Labov's paradox

3.2 Phonetics

3.3 Other sound changes

3.4 Anti-homonymy

3.5 Lexical Replacement

4. Conclusions

Methods

1. Technically one could take a lot of cases where we have diachronic data (for instance, Latin, Old English, Ancient Greek,...) and compare dispersion between classical and contemporary languages
2. Alternatively, we can build different **models of sound change**, and see how phonemic dispersion in the lexicon over time changes as a function of our parameters via simulations -> **This is what we do today**

A model of sound change

RECIPE

1. **Alphabet**: 25 consonants, 13 vowels
English orthography used as a convention
2. **Mini lexicon**: \approx 150 words that easily map to CVC representation
(e.g., *bad, big, book, but, can, dad, for, get, god, him, head,...*)
3. **Mergers and Splits applied to the mini-lexicon**
4. **Feature representation** (later)

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 [k] and one outside of the inventory [s]
3. Select conditioning environment [e, i]
4. k becomes s in the conditioning environment.

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

Example

Conditioned merger of **a** in **o** after [b m p f d l n s c g j]

Conditioned merger of **n** in **s** in onsets before [i u e o]

INPUT -> **OUTPUT**

bad -> bod

big -> big

can -> con

cut -> cut

man -> mon

mom -> mom

not -> sot

son -> son

Section 3

1. Introduction

1.1 Sound change and phonemic dispersion in the lexicon

1.2 This talk

2. Methods

2.1 A model of sound change

3. Results

3.1 Addressing Labov's paradox

3.2 Phonetics

3.3 Other sound changes

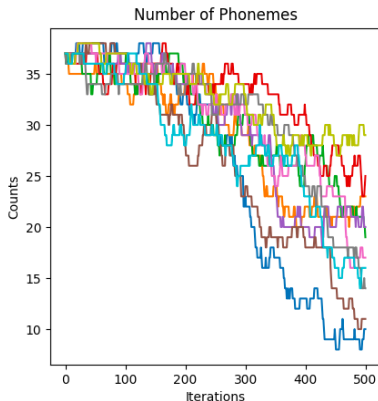
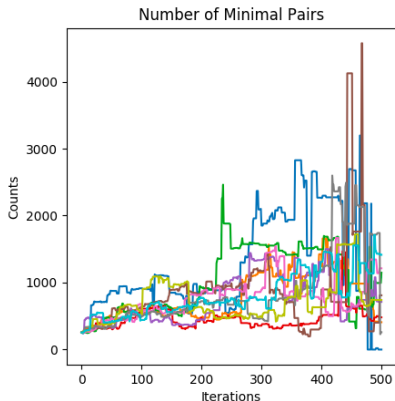
3.4 Anti-homonymy

3.5 Lexical Replacement

4. Conclusions

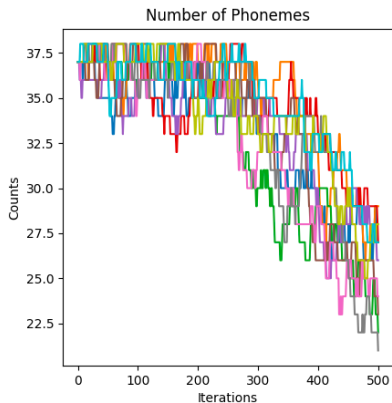
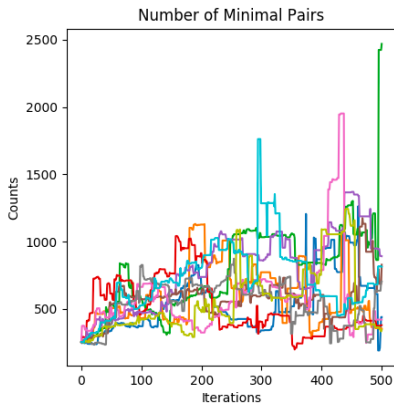
Example: Labov's paradox

Merge $p=2/3$, Split $p=1/3$, 10 parallel runs, 500 changes



Labov's paradox also true for equiprobability

Merge $p=1/2$, Split $p=1/2$, 10 parallel runs, 500 changes




Irreversibility of Mergers

EXPLANATION

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

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

sad	sad
se t	sa t
far	fa r
he r	ha r

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

Irreversibility of Mergers

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

(B)  (C)

sa**d**

sa**t**

sa**t**

sa**t**

(B)  (C)

Summary, so far

- Phonemic dispersion in the lexicon decreases because the irreversibility of mergers pushes the lexicon in **clumpy** spaces, and getting out of them is more difficult than getting in.
- We can try to make the model more realistic by introducing the notions of:
 1. Phonetics
 2. Additional sound changes
 3. Anti-homonymy
 4. Lexical replacement

1. Phonetics: Naive Feature Representation

b	c	ch	d	f	g	h	j	k	l	m	n	p	r	s	sh	t	th	tw	v	w	wh	x	'	'
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	

a	ai	e	ea	ee	ei	i	o	oa	oi	oo	ou	u
0	1	2	3	4	5	6	7	8	9	10	11	12

1. Phonetics: Merger

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

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

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

ONSET: **n**et, **n**ot

b	c	ch	d	f	g	h	j	k	l	m	n	p	r	s	sh	t	th	tw	v	w	wh	x	'	'
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	

a	ai	e	ea	ee	ei	i	o	oa	oi	oo	ou	u
0	1	2	3	4	5	6	7	8	9	10	11	12

1. Phonetics: 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

a	ai	e	ea	ee	ei	i	o	oa	oi	oo	ou	u
0	1	2	3	4	5	6	7	8	9	10	11	12

1. Phonetics: Merger

TARGET-1 becomes TARGET-2 in the conditioning environment

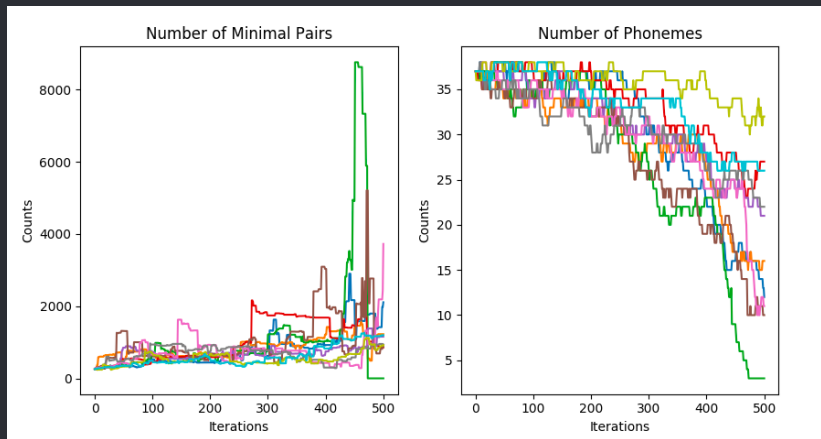
b	c	ch	d	f	g	h	j	k	l	m	n	p	r	s	sh	t	th	tw	v	w	wh	x	'
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23

a	ai	e	ea	ee	ei	i	o	oa	oi	oo	ou	u
0	1	2	3	4	5	6	7	8	9	10	11	12

net, nod -> met, nod

1. Phonetics

Merge $p=1/2$, Split $p=1/2$, 10 parallel runs, 500 changes



2. Additional sound changes

- Is focusing on Mergers and Splits an oversimplification?
Empirical data might be useful here.
- We collected a list of sources of phoneme creation/deletion in IE (Ringe 2011), Uralic (Sammallahti 1988) and Altaic (Robbeets 2003)

2. Additional sound changes

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	20	1	1		
Proto-Turkic to Turkish	1	2	8		8
TOTAL	71	9	28	14	72

Table 1: A summary of processes regulating phoneme creation and deletion.
Thanks to Bert Vaux for spotting an error.

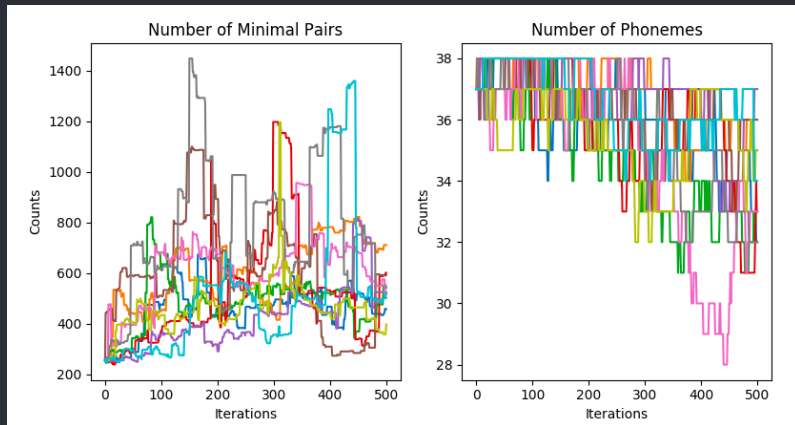
2. Additional sound changes: Pseudo-Contraction

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

2. Additional sound changes: Pseudo-Contraction

Merge $p=1/3$, Split $p=1/3$, Contract $p=1/3$, 10 parallel runs, 500 changes

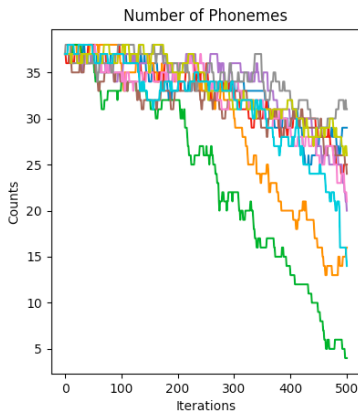
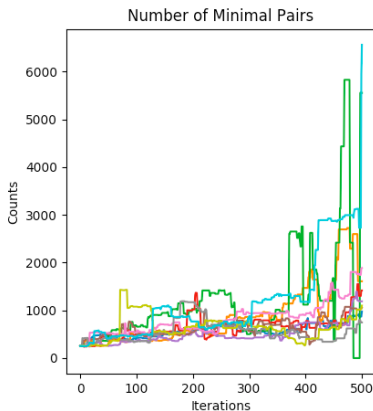


3. Anti-homonymy

- What if there is a bias against homonymy?
- We can modify the model as to only allow mergers that do not produce homonymy (unless ALL possible mergers do)

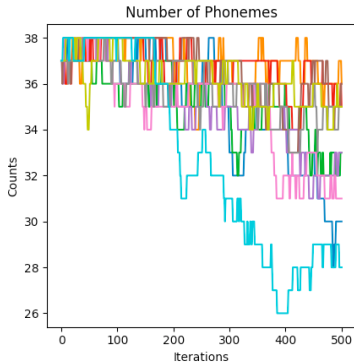
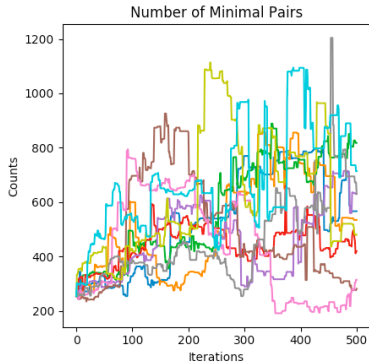
3. Anti-homonymy

Merge $p=1/2$, Split $p=1/2$, 10 parallel runs, 500 changes



3. Anti-homonymy

Merge $p=1/3$, Split $p=1/3$, Contract $p=1/3$, 10 parallel runs, 500 changes



4. Lexical Replacement

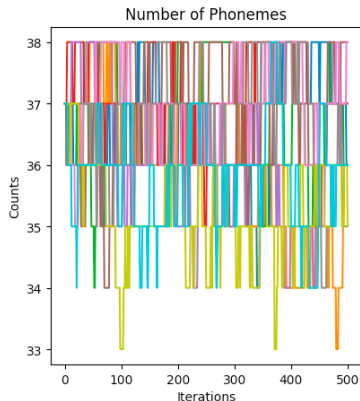
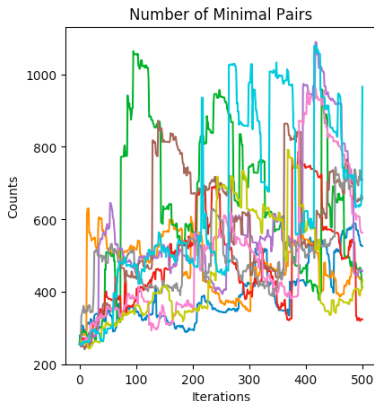
LEXICAL REPLACEMENT

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)

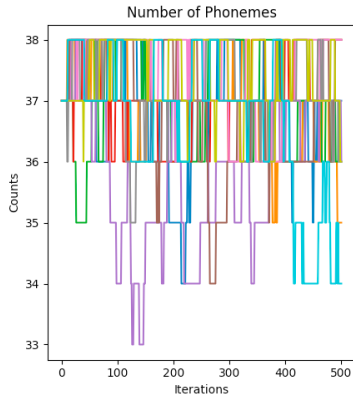
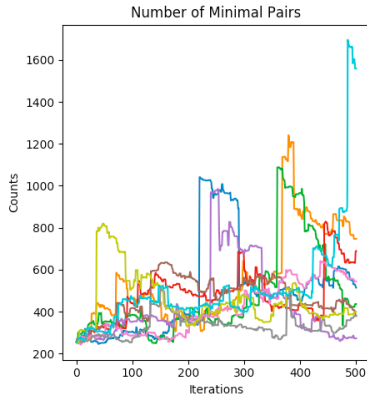
4. Lexical Replacement

Merge $p=1/3$, Split $p=1/3$, Borrowing $p=1/3$, 10 parallel runs, 500 changes



4. Lexical Replacement

Merge $p=1/4$, Split $p=1/4$, Contract $p=1/4$, Borrowing $p=1/4$, 10 parallel runs, 500 changes



Section 4

1. Introduction

1.1 Sound change and phonemic dispersion in the lexicon

1.2 This talk

2. Methods

2.1 A model of sound change

3. Results

3.1 Addressing Labov's paradox

3.2 Phonetics

3.3 Other sound changes

3.4 Anti-homonymy

3.5 Lexical Replacement

4. Conclusions

Conclusions

- Phonemic dispersion in the lexicon decreased in all conditions.
- Languages tend to **decrease phonemic dispersion over time** regardless of functional considerations (explanation: **Irreversibility of Mergers**)
- This does not correlate with the fact that the size of phonemic inventory tends to decrease. Once a model includes contractions or lexical replacement, the size of the phonemic inventory appears **stable** (interestingly, this is not true for **anti-homonymy**)
- **No need for functional explanations for things that we can derive mechanically as a result of sound change!**

Thanks!

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

Thanks to the ESHP organizers!