Modelling variation in sound change: social setting as a determinant of process application

Karolina Broś

University of Warsaw

June 29th, 2018

1 Introduction

1.1 Sound change in phonological analysis

- diachronic variation (change in real time)
- diachronic variation (change in apparent time)
- cross-dialectal variation caused by social, political and/or geographical factors
- intra-dialectal inter-speaker variation pointing to change in progress
- intra-speaker variation pointing to change in progress

Labov (1972, 1990) points to the influence of external factors on even a single grammar of a dialect user.

Recent approaches to variation include arguments for incorporating it into the theory of grammar instead of just relegating it to the phonetic implementation component or late phonology (e.g. Coetzee 2009, 2016).

1.2 Grammar-external factors in phonological analysis

- speech rate (Coetzee 2017)
- register (van Oostendorp 1997)
- style (Boersma & Hayes 2001)
- lexical frequency (Bybee 2000, 2001; Coetzee & Kawahara 2013)
- morphological status (Coetzee 2009)
- lexical idiosyncrasy (Goeman 1999, Coetzee 2009)
- social setting

All factors that contribute to grammatical structure should be included in formal representation (without detriment to purely phonological computation / grammar dominance).

1.3 Purpose of the study

- Take different stages of sound change presented by the same speakers
- Moderate vs. more radical weakening depending on the situational context and the associated speech 'modality'
- Domain of application effects and between-modality systematicity

1.4 Speech modalities and generalisations

Language/variety: Spanish from Gran Canaria

Speakers: 6 natives recorded on 2 occasions

Data 1: read and repeated speech

Data 2: spontaneous speech

2 Data

2.1 3 interacting weakening processes

Modality 1

/s/ -> [h/H] /_V	*prensa[h]idráulicas* 'hydraulic presses'
/s/ -> [h] /_k	*chocolate[h]con* 'chocolates with'
/s/ -> [∅] /_d	*pane[∅]de* 'breads from'
/b d g/ -> [b d g] /V(C)_	*pane(s)[d]e* 'breads from'
/b d g/ -> [B D G] /V_	*cinco[D]ulces* 'five sweets'
/p t k/ -> [b d g] /V_	*cinco[b]anes* 'five breads'

- *s* debuccalisation before voiceless consonants and vowels
- *s* deletion before voiced consonants (and pauses)
- *b d g* spirantisation after vowels, **blocked** in derived vocalic environments
- *p t k* voicing after vowels

Modality 2

/s/ -> [h/H]/_V	*prensa[H]idráulicas* 'hydraulic press'
/s/ -> [∅] /_C	*chocolate[∅]con* 'chocolates with'
/b d g/ -> [B D G] /V(C)_	*pane(s)[D]e* 'breads from'
/b d g/ -> [B D G] /V_	*cinco[D]ulces* 'five sweets'
/p t k/ -> [b d g] /V_	*cinco[b]anes* 'five breads'
/p t k/ -> [p t k] /V(C)_	*chocolate(s)[k]on* 'chocolates with'

- *s* debuccalisation before voiceless consonants and vowels
- *s* deletion before **all** consonants
- *b d g* spirantisation in **all vocalic environments** (including derived)
- *p t k* voicing after vowels, **blocked** in derived vocalic environments

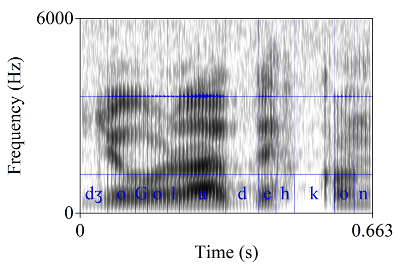
2.2 Processes and domains of application

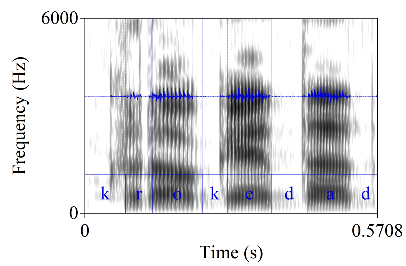
1. Coda weakening (debuccalisation, voicing, elision). In spontaneous speech it also includes other consonants: /d/, /r/, /l/ (variation: optional).

2. Voiced stop weakening also applies (variably) after a non-deleted sonorant, and always after a non-deleted /s/ in spontaneous speech. Intervocalically very strong, incl. deletion. => Domain extension

3. Voiceless stop weakening applies both inside words and across word boundaries, but strictly after a vowel. It can be accompanied by approximantisation and occasionally occurs after deletion.

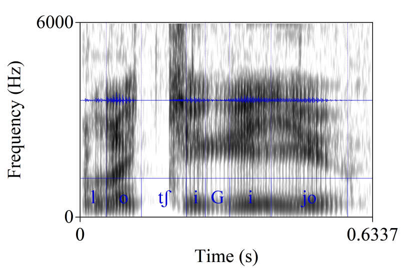
2.3 Spectrograms

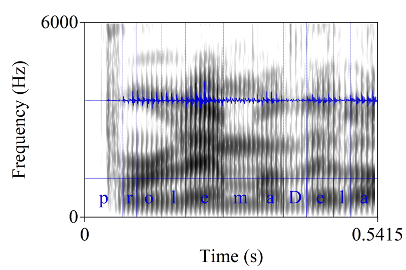




Controlled speech.

- Left: *chocolates con* 'chocolate with' presents no /s/ deletion before a voiceless stop and no voicing.
- Right: *croquetas de* 'croquettes with' presents deletion before a voiced segment but no spirantisation.

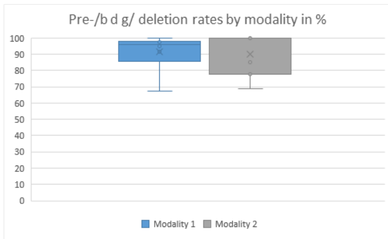


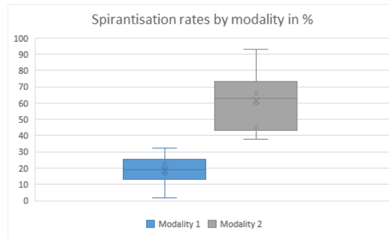


Spontaneous speech.

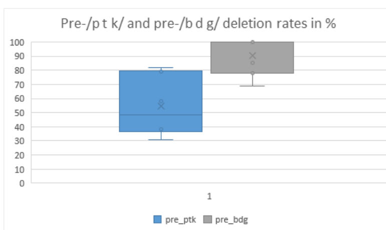
- Left: *los chiquillos* 'the guys' presents deletion before a voiceless sound and no voicing.
- Right: *problemas de la* 'problems with/about' presents deletion before a voiced sound and spirantisation.

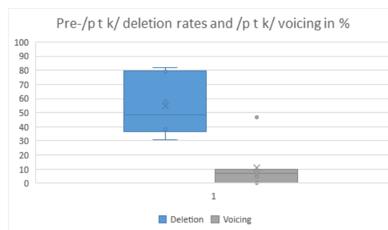
2.4 Frequency of occurrence – distribution graphs





- Deletion rates before voiced segments the same across modalities.
- Occasional spirantisation in controlled speech, with gender differences.
- Twofold rise in spirantisation across speakers.
- Number of tokens uneven (rate expected to rise with more tokens).





- The data are for Modality 2.
- In Modality 1, the pre-/ptk/ deletion rate is 0% hence no voicing.
- Rates of deletion before voiceless sounds include all consonants.

3 Theoretical framework

3.1 Theoretical assumptions and formal account

The data require a variationist approach:

- speaker productions are highly dependent on speech modality
- competition between two co-phonologies
- variation is a reflection of a change in progress: transition from one system to another

The lifecycle of phonological processes:

- the domains of application are gradually extended
- spirantisation inside words is now phonologised, across a word boundary the transition is not complete
- new rules alongside old rules
- the same trajectory applies to post-vocalic voicing, a much younger process at a different advancement stage

3.2 Problems

1. Different weakening stages

Positional and general markedness constraints : *ptk, *V_ptk, *bdg, *V_bdg

2. Turbidity for selective blocking

Deleted segments leave a trace/block processes.
Positional markedness constraints are not violated as the segment is not erased from the phonological representation.

/paneh+de/	*h	*V_bdg	MAX(C)	IDENT(cont)
a. paneh.de	*!			
☞ b. pane(h).de			*	
c. pane(h).De			*	*!

3. Cross-modality variation

Coetzee's (2009, 2016) model incorporating external factors in computation

Stochastic approach:
Noisy Harmonic Grammar (Coetzee & Pater 2011)

Constraints:
h, *H, MAX(C), *s]CODA, AGREECC voice, IDENT(cont), IDENT(voice), IDENT(Place)
First model: non-noisy Harmonic Grammar (weights instead of rankings)

4 HG analysis

4.1 Grammar 1 (Modality 1)

(4.0) *V ptk > (3.0) Max(C), *s]Coda >

(2.0) AgreeCC voice, *H, *V bdg, Ident(cont) >

(1.0) *h, Ident(voice), Ident(Place), *bdg, *ptk

- Deletion is banned unless other important constraints are violated
- Positional markedness outranks general markedness

Evaluation of the phrase *panes de* 'breads of' WITH deletion but NO spirantisation

Weights	3.0	3.0	2.0	2.0	1.0	1.0	1.0	
Input: panes de	Max(c)	*s]Coda	AgreeCC voice	Ident(cont)	*h	Ident(Pl)	*bdg	
panes de	0	-1	-1	0	0	0	-1	WT: -6.0
panes De	0	-1	-1	-1	0	0	0	WT: -7.0
paneh de	0	0	-1	0	-1	-1	-1	WT: -5.0
paneh De	0	0	-1	-1	-1	-1	0	WT: -6.0
☞ pane(s) de	-1	0	0	0	0	0	-1	WT: -4.0
pane(s) De	-1	0	0	-1	0	0	0	WT: -5.0

Evaluation of the phrase *panes con* 'breads with' with NO deletion and NO voicing

Weights	3.0	3.0	1.0	1.0	1.0	1.0	1.0	
Input: panes kon	Max(c)	*s]Coda	*h	Ident(voice)	Ident(Pl)	*bdg	*ptk	
panes kon	0	-1	0	0	0	0	-1	WT: -4.0
☞ paneh kon	0	0	-1	0	-1	0	-1	WT: -3.0
pane(s) kon	-1	0	0	0	0	0	-1	WT: -4.0
paneh gon	0	0	-1	-1	-1	-1	0	WT: -4.0
pane(s) gon	-1	0	0	-1	0	-1	0	WT: -5.0

Evaluation of the phrase *cena de* 'dinner of' with post-vocalic spirantisation

Weights	2.0	2.0	1.0	
Input: sena de	*V_bdg	Ident(cont)	*bdg	
sena de	-1	0	-1	WT: -3.0
☞ sena De	0	-1	0	WT: -2.0

Evaluation of the phrase *cena con* 'dinner with' with post-vocalic voicing

Weights	4.0	2.0	1.0	1.0	1.0	
Input: sena kon	*V_ptk	*V_bdg	Ident(voice)	*bdg	*ptk	
sena kon	-1	0	0	0	-1	WT: -5.0
☞ sena gon	0	-1	-1	-1	0	WT: -4.0

- It is enough to change the weight of one constraint for the ranking to flip in favour of the other candidate

- I assume that if voicing were present exactly 50% of the time, the weighted total of each candidate should be the same.

4.2 Grammar 2 (Modality 2)

(3.0) *V ptk > (2.0) *s]CODA, *bdg, *ptk >
(1.0) AGREECC voice, *H, MAX(C), *h, *V bdg,
IDENT(voice), IDENT(cont), IDENT(Place)

- MAX(C) is demoted to allow deletion more freely across the voiced/voiceless environments
- General markedness constraints ascend; *bdg is ranked above *V bdg to allow across-the-board spirantisation (domain of applicaton extended)

Evaluation of the phrase *panes de* 'breads of' WITH deletion AND spirantisation

Weights	2.0	2.0	1.0	1.0	1.0	1.0	1.0	
Input: panes de	*s]Coda	*bdg	AgreeCC voice	Max(c)	*h	Ident(cont)	Ident(Pl)	
panes de	-1	-1	-1	0	0	0	0	WT: -5.0
panes De	-1	0	-1	0	0	-1	0	WT: -4.0
paneh de	0	-1	-1	0	-1	0	-1	WT: -5.0
paneh De	0	0	-1	0	-1	-1	-1	WT: -4.0
pane(s) de	0	-1	0	-1	0	0	0	WT: -3.0
☞ pane(s) De	0	0	0	-1	0	-1	0	WT: -2.0

Evaluation of the phrase *panes con* 'breads with' WITH deletion and NO voicing

Weights	2.0	2.0	2.0	1.0	1.0	1.0	1.0	
Input: panes kon	*s]Coda	*bdg	*ptk	Max(c)	*h	Ident(voice)	Ident(Pl)	
panes kon	-1	0	-1	0	0	0	0	WT: -4.0
paneh kon	0	0	-1	0	-1	0	-1	WT: -4.0
☞ pane(s) kon	0	0	-1	-1	0	0	0	WT: -3.0
paneh gon	0	-1	0	0	-1	-1	-1	WT: -5.0
pane(s) gon	0	-1	0	-1	0	-1	0	WT: -4.0

5 GLA for variation modelling

5.1 Is it therefore possible to model variation based on predefined input parameters?

- Let's submit this grammar to the HG-GLA (Praat) with noise set at the default 2.0

- The addition of noise allows for random variation in the grammar

- The addition of predefined pattern frequencies helps the algorithm find the best fit as per the observed data

5.2 Stochastic parameters for the GLA

- **pre-*b d g* deletion: 92%**
- **pre-*p t k* deletion: 24%**
  - general deletion: 58%

- **post-deletion spirantisation: 44%**
- **post-vocalic spirantisation: 100%**
  - general spirantisation: 72%

- **post-deletion voicing: 5%**
- **post-vocalic voicing: 64%**
  - general voicing: 34%

- **Input weights calculated following the chain rule of probability.**
- **Deletion and voicing/spirantisation patterns are in a dependency relation.**

panes = x, *de* = y:

$$x_1y_1 = [panehde] \qquad 8\% * 100\%$$

$$x_2y_1 = [panede] \qquad 92\% * 56\%$$

$$x_1y_2 = [panehDe] \qquad 8\% * 0\%$$

$$x_2y_2 = [paneDe] \qquad 92\% * 44\%$$

Pair Distributions

pairs [1]:
string1 = "/Vs#d/"
string2 = "[V(s)#d]"
weight = 52
pairs [2]:
string1 = "/Vs#d/"
string2 = "[V(s)#D]"
weight = 40
pairs [3]:
string1 = "/Vs#d/"
string2 = "[Vh#d]"
weight = 8
pairs [4]:
string1 = "/Vs#k/"
string2 = "[Vh#k]"
weight = 76
pairs [5]:
string1 = "/Vs#k/"
string2 = "[V(s)#k]"
weight = 23
pairs [6]:
string1 = "/Vs#k/"
string2 = "[V(s)#g]"
weight = 1
pairs [7]:
string1 = "/V#k/"
string2 = "[V#g]"
weight = 64
pairs [8]:
string1 = "/V#k/"
string2 = "[V#k]"
weight = 36
pairs [9]:
string1 = "/V#g/"
string2 = "[V#G]"
weight = 100
pairs [10]:
string1 = "/V#g/"
string2 = "[V#g]"
weight = 0

5.3 Procedure

- A grammar file with input/output pairs and their distributions averaged between the two modalities
- Noise and plasticity set to defaults
- LinearOT parameter chosen to model Noisy HG
- Initial constraint weights set at 100.00
- Given the noise factor, the simulation was repeated 10 times
- 'To output distributions' function used each time
- Constraint weights and output distributions averaged to give a final estimate

What are the learned constraint weights?

Constraint	Weight
MAX(C)	139.696
ptk	137.259
V ptk	104.209
s]CODA	100.000
V bdg	95.791
AGREECC-voice	83.633
IDENT(cont)	81.7226
bdg	81.017
IDENT(Place)	76.669
IDENT(voice)	62.740
h	60.303

Does the learned grammar correctly predict variation?

Deletion	Spirantisation	Voicing	Observed	Predicted
✓	–		52	53.02
✓	✓		40	38.32
–	–		8	8.65
–		–	76	77.71
✓		–	23	21,15
✓		✓	1	1.12
		✓	64	66,46
		–	36	33,54
	✓		100	100
	–		0	0

How are the different options modelled?

/Vs**d**/	Max-C	*ptk	*$V_p$tk	*s-Coda	*bdg	AgreeCC-Voice	Ident-cont	*$V_b$dg	Ident-Place	*h	Ident-voice	
[V(s)**d**]	*				*							-239.44
☞ [V(s)**D**]	*						*					-239.34
[Vh**d**]					*	*				*		-244.94

/Vs**d**/	Max-C	*ptk	*$V_p$tk	*s-Coda	Ident-cont	*bdg	AgreeCC-Voice	*$V_b$dg	Ident-Place	Ident-voice	*h	
☞ [V(s)**d**]	*					*						-241.80
[V(s)**D**]	*				*							-242.84
[Vh**d**]						*	*				*	-246.21

/Vs**d**/	Max-C	*ptk	*s-Coda	*$V_p$tk	Ident-cont	*bdg	AgreeCC-Voice	*$V_b$dg	Ident-Place	*h	Ident-voice	
[V(s)**d**]	*					*						-238.175
[V(s)**D**]	*				*							-238.177
☞ [Vh**d**]						*	*			*		-237.508

How are the different options modelled?

/Vs**k**/	Max-C	*ptk	*$V_p$tk	*s-Coda	Ident-cont	AgreeCC-Voice	*$V_b$dg	*bdg	Ident-Place	Ident-voice	*h	
☞ [Vh**k**]		*							*		*	-282.789
[V(s)**k**]	*	*										-288.279
[V(s)**g**]	*							*		*		-291.490

/Vs**k**/	Max-C	*ptk	*$V_p$tk	*s-Coda	*$V_b$dg	*bdg	Ident-cont	AgreeCC-Voice	Ident-Place	*h	Ident-voice	
[Vh**k**]		*							*	*		-289.53
☞ [V(s)**k**]	*	*										-287.70
[V(s)**g**]	*					*					*	-294.28

6 Weight scaling

6.1 Random vs. parameter-based variation

Fitting the model to the data

- The baseline model accounting for average variation across modalities can be generated
- But Noisy HG-GLA assumes random noise added at each evaluation, whereas the actual data present a clear pattern
- What is the exact role/contribution of the 'modality' factor to the output distributions?
- As the setting or communication situation is external to grammar, it does not influence grammar-internal relations
- Following Coetzee and Kawahara's (2013) model, I use my Noisy HG grammar as baseline and apply weight scaling – a simpler grammar (6 pairs, 8 constraints)
- The task is difficult: mutliple interdependent processes and constraints

Fitting the model: Generated weights comparison

Con	Gen. distrib.	Mod. 1	Diff.	Mod. 2	Diff.
MAX-C	139.656	143.621	**3.965**	139.397	**-0.259**
*ptk	137.219	100.000	**-37.219**	138.156	**0.937**
AGREE-CC	83.609	92.184	**8.575**	82.169	**-1.440**
IDENT(cont)	81.616	100.996	**19.380**	80.654	**-0.962**
*bdg	81.165	99.004	**17.839**	81.190	**0.025**
IDENT(Place)	76.735	64.195	**-12.540**	78.434	**1.699**
IDENT(voice)	62.781	100.000	**37.219**	61.844	**-0.937**
*h	60.344	56.379	**-3.965**	60.603	**0.259**

General	52	40	8	76	23	1
Mod.1	69	23	7	100	0	0
Mod. 2	35	55	10	52	43	5

Fitting the model: Modality 1

Pattern	Baseline	Fit 1	Fit 2	Fit 3	Fit 4	Desired fit
V(s)d	50.4	58.6	69	71.8	69.1	**69**
V(s)D	41.7	26.3	20	20.4	23.3	**23**
Vhd	7.9	15.1	11	7.8	7.6	**7**
Vhk	76.7	77	82.6	82.7	82.7	100
V(s)k	22.3	22.6	17	17	17	0
V(s)g	1	0.4	0.4	0.3	0.3	0

Fit 1: Add 1.0 weight to all faithfulness contraints

Fit 2: Add 1.7 to Max and Ident(cont), and 1.5 to AGREE

Fit 3: Add 1.7 to Max and Ident(cont), and 2.2 to AGREE

Fit 4: Add 1.7 to Max, 1.4 to Ident(cont) and 2.2 to AGREE

Fitting the model: Modality 1

Pattern	Baseline	Fit 1	Fit 2	Fit 3	Desired fit
V(s)d	50.4	66.8	62	39.5	69
V(s)D	41.7	23.2	22.2	16.9	23
Vhd	7.9	9.9	15.7	43.6	7
Vhk	76.7	86.4	91.4	98.7	**100**
V(s)k	22.3	13.3	8.3	1.2	**0**
V(s)g	1	0.3	0.2	0.04	**0**

Fit 1: Add 2.2 to Max, 1.4 to Ident(cont) and 2.2 to AGREE

Fit 2: Add 3.2 to Max, 1.4 to Ident(cont) and 2.2 to AGREE

Fit 3: Add 5.2 to Max, 1.4 to Ident(cont) and 2.2 to AGREE

Fitting the model: Modality 2

Pattern	Baseline	Fit 1	Fit 2	Fit 3	Fit 4	Fit 5	Desired
V(s)d	51.8	38.9	41.9	36.4	34.9	34.4	**35**
V(s)D	40.8	54.7	56.2	53.1	56.3	55.9	**55**
Vhd	7.4	6.3	1.9	10.5	8.8	9.6	**10**
Vhk	76.7	76.1	55.7	55.7	51.7	52.2	**52**
V(s)k	22.2	21.1	41.1	41.2	44	43.5	**43**
V(s)g	1	1.9	3.1	3.1	4.3	4.3	**5**

Fit 1: Subtract 1.0 from IDENT(cont) and IDENT(voice)

Fit 2: Additionally, subtract 2.0 from MAX(C) – more deletions

Fit 3: Additionally, subtract 3.0 from AGREE to balance Vhd cases

Fit 4: Additionally, subtract 0.2 from IDENT(cont) and 0.3 from MAX(C) to better match voicing cases

Fit 5: Additionally, subtract 0.2 from AGREE to rebalance Vhd cases

Fitting the model: comparison

Con	Gen. distrib.	Modality 2	Scaled	Diff.
MAX-C	139.621	139.397	137.321	-2.076
*ptk	137.248	138.156	137.248	-0.908
AGREE-CC	83.601	82.169	80.401	-1.768
IDENT(cont)	81.642	80.654	80.442	-0.212
*bdg	81.111	81.190	81.111	-0.079
IDENT(Place)	76.678	78.434	76.678	-1.756
IDENT(voice)	62.752	61.844	61.252	-0.592
*h	60.379	60.603	60.379	-0.224

7 Conclusions

- Individual speaker choices can be systematic across different social settings
- Intra-speaker variation can be a reflection of sound change in progress

- Variation should be modelled by incorporating external factors into the grammar but assuming grammar-dominance
- Noisy HG can correctly predict/learn variation based on input data
- The treatment of external factors as independent variables added in the form of single scaling factors may be insufficient
- Boersma & Hayes (2001) model should be revisited vis a vis Coetzee & Kawahara's (2013) proposal

Thank you!

References