Class 19: Some more of what our colleagues are up to

1 Less is more: Becker, Clemens & Nevins 2011 on French plurals

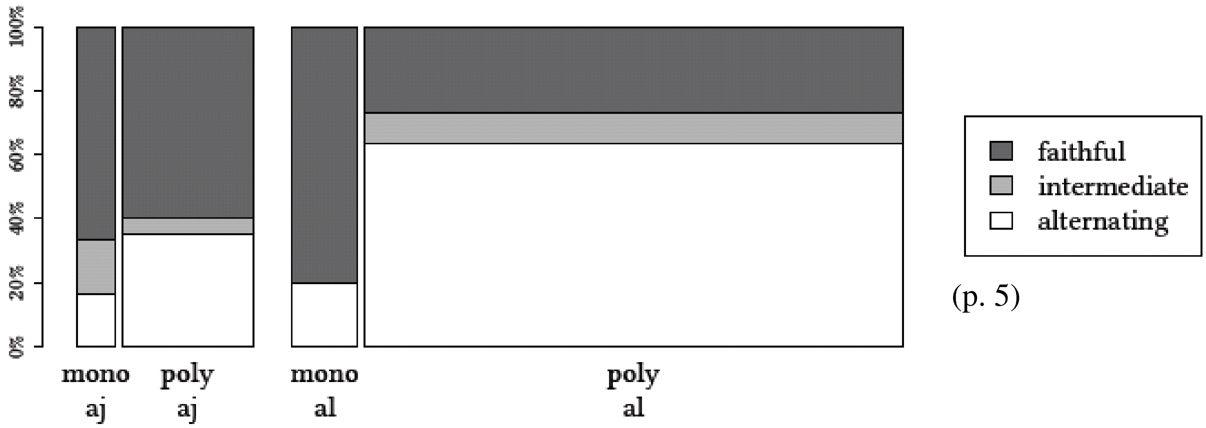
- French regular plural suffix is –Ø (with a “latent” [z]; shows up only in liaison contexts)
- But some masculines are irregular, changing the stem-final VC.
- Result is many-to-many relation of sing. and plural patterns—here’s the main relevant chunk:

<table>
<tr><th>singular</th><th>plural</th><th>singular/plural</th><th></th></tr>
<tr><td rowspan="3">Xal</td><td>Xal</td><td>bal/bal
ʃakal/ʃakal
kaʁnaval/kaʁnaval</td><td>‘ball’
‘jackal’
‘carnival’</td></tr>
<tr><td>Xal ~ Xo</td><td>val/val ~ vo
final/final ~ fino</td><td>‘valley’
‘final’</td></tr>
<tr><td rowspan="3">Xo</td><td>mal/mo
bokal/boko
ʒuʁnal/ʒuʁno
ʃəval/ʃəvo
ɛ̃tɛʁnasjɔnal/...no</td><td>‘evil’
‘jar’
‘newspaper’
‘horse’
‘international’</td></tr>
<tr><td>Xo</td><td>tɥijo/tɥijo
ʃɑto/ʃɑto
so/so</td><td>‘pipe’
‘castle’
‘pail’</td></tr>
<tr><td rowspan="3">Xaj</td><td>tʁavaj/tʁavo
supiʁaj/supiʁo</td><td>‘work’
‘basement window’</td></tr>
<tr><td>Xaj ~ Xo</td><td>baj/baj ~ bo
aj/aj ~ o[archaic]</td><td>‘lease’
‘garlic’</td></tr>
<tr><td>Xaj</td><td>maj/maj
detaj/detaj
evɑ̃taj/evɑ̃taj</td><td>‘hammer’
‘detail’
‘fan’</td></tr>
<tr><td rowspan="2">Xɛl</td><td>Xɛl</td><td>lɔʒisjɛl/lɔʒisjɛl
etɛʁnɛl/etɛʁnɛl</td><td>‘software’
‘eternal’</td></tr>
<tr><td rowspan="6">Xø</td><td>ʃəvɛl/ʃəvø
sjɛl/sjø</td><td>‘hair’
‘sky’</td></tr>
<tr><td>Xø</td><td>fø/fø
nəvø/nəvø</td><td>‘fire’
‘nephew’</td></tr>
<tr><td>Xœl</td><td>ajœl/ajø</td><td>‘ancestor’</td></tr>
<tr><td>Xœj</td><td>œj/(z)jø</td><td>‘eye’</td></tr>
<tr><td>Xœf</td><td>bœf/bø</td><td>‘steer’</td></tr>
<tr><td rowspan="2">Xɛj</td><td>vjɛj [liason]/vjø</td><td>‘old’</td></tr>
<tr><td>Xɛj</td><td>ɔʁtɛj/ɔʁtɛj
kõsɛj/kõsɛj</td><td>‘toe’
‘advice’</td></tr>
</table>

There are also regulars for *Xœl, Xœj, Xœf*

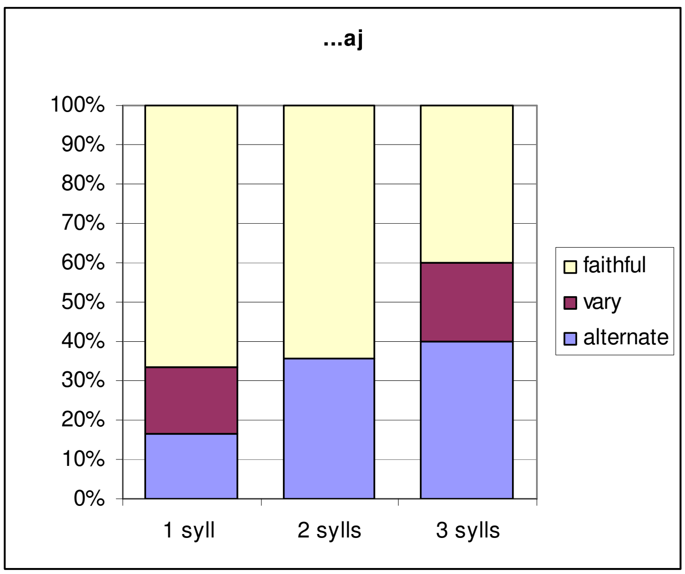
(Some of the above have a different plural with a different meaning.)

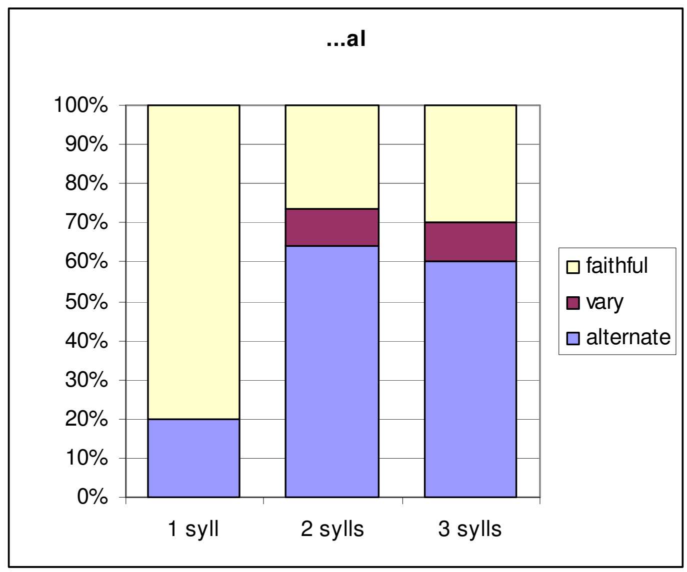
- Becker & al. extracted the 118 relevant nouns and adjectives, and had a speaker mark their plurals.
- Syllable count matters in the real lexicon: monosyllables tend to be faithful:



(p. 5)

- No real difference between 2 and 3 syllables (graph made from their posted data):



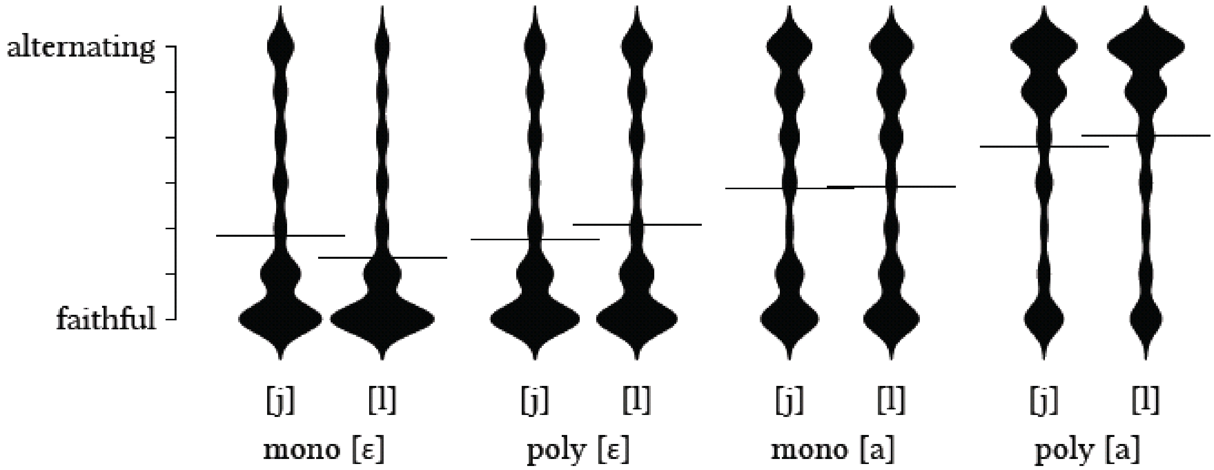


- Becker & al.'s interpretation: initial syllables are protected by a special faithfulness constraint (see Becker, Ketrez & Nevins 2011, Becker, Nevins & Levine 2012)

2 Becker & al.'s experiment

- Experiment with nonce words.
  - Singular presented auditorily in frame sentence: e.g., [dal]
  - 2 options presented auditorily for plural: e.g., candidate plurals [dal] and [do]
  - Participants (185!) responded on a scale from 1 (strongly prefer option on left) to 7 (strongly prefer option on right)

- "Beanplot" of ratings (horizontal line is mean):



(p. 11)

sg. dɛj	dɛl	fanɛj	fanɛl	daj	dal	dɛʁnaj	dɛʁnal
pl. dɛj/dø	dɛl/dø	fanɛj/...ø	fanel/...ø	daj/do	dal/do	dɛʁnaj/o	dɛʁnal/o

- Subjects hated alternation for [ɛj, ɛl]
- Within [aj, al], alternation was a lot better in polysyllables than in monosyllables.
- Unlike in lexicon, no real [j]/[l] difference
  - As you can see in the plots, there's a small difference within the polysyllables
- Becker & al.'s interpretation: even though the alternating pattern is not productive, people are sensitive to the syllable-count effect, because it's phonologically natural

[There's a similar lexical and experimental study of Brazilian Portuguese plurals in the paper—it's a little more complex.]

3 The "less is more" part of Becker & al.: factors that don't seem to play a role

- We might expect that monosyllabic/polysyllabic is a just a crude measure of length
- Reasons why longer words would be more willing to alternate
  - The shorter the word is, the less unchanging material there is between singular and alternated plural
    - E.g., from [mo] it's pretty hard to recover the singular [mal], since they share only one segment
    - This actually isn't one of the reasons they consider
  - The longer the word is, the shorter each segment is.
    - Changing the [al] in [ɛ̃tɛʁnasjɔnal] is not so bad, because the material changed is short (see Giavazzi 2010)
- So does actual duration improve the regression model for predicting participants' responses?
  - No.
- Same story for bigram probability (monosyllables have higher bigram probability, since there are fewer bigrams whose probability needs to be multiplied)—no improvement

- Neighborhood density?
  - Monosyllables have more neighbors
    - Why? (not a logical necessity, but tends to be true in languages)
    - Why would having more neighbors make a word resist the irregular plural? (See Ussishkin & Wedel 2002)
  - Log of neighborhood density does make a significant contribution in French, but only if density is measured based on a sample of well over 30,000 words
    - Suggests it would matter only if the participant has a quite large vocabulary
    - Should predict an age effect
  - In Portuguese, density doesn't improve the model
- Their interpretation: learners prefer to keep it simple.

- What do you think?

4 Loucherbem

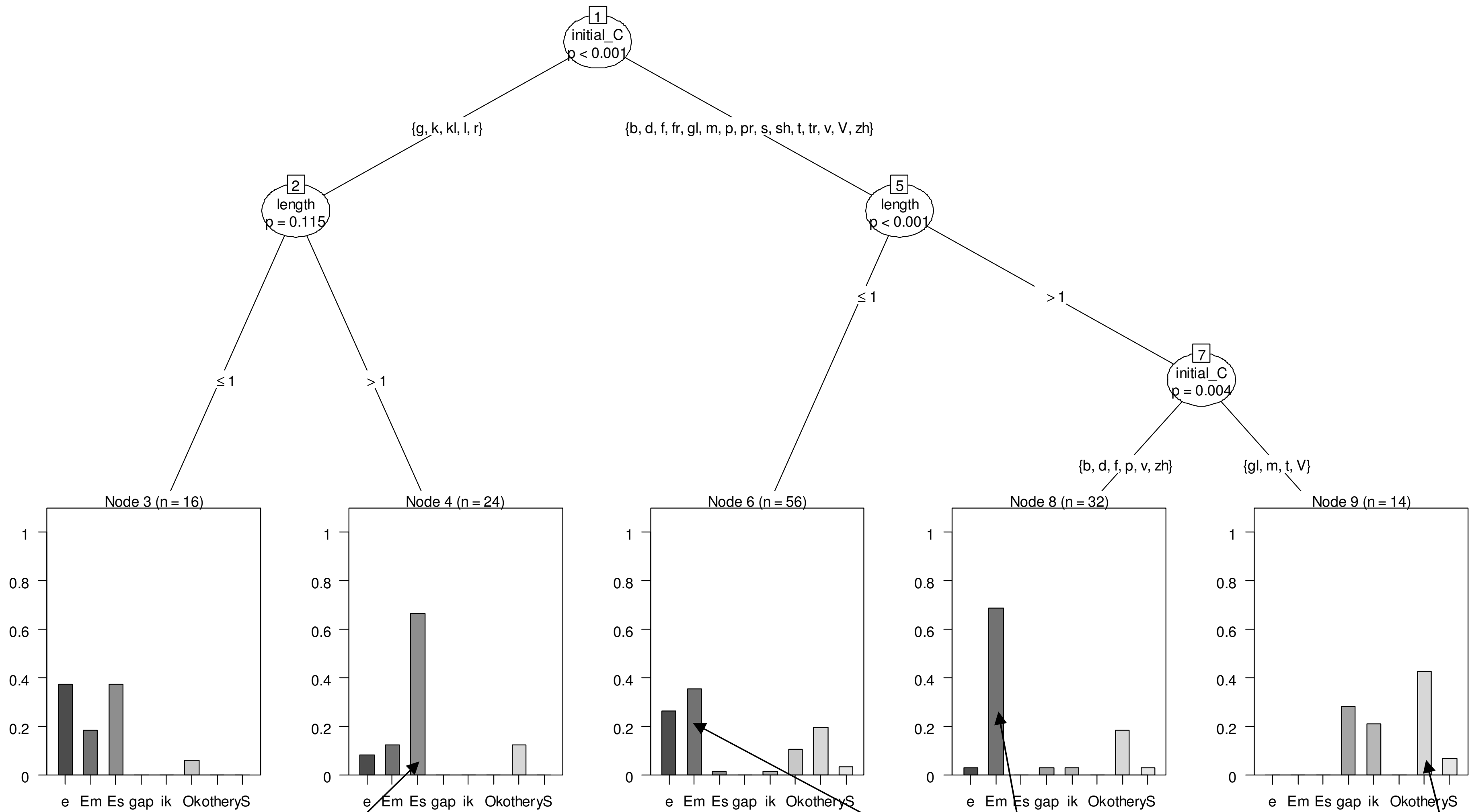
- I thought it'd be fun to think about a case where recoverability is not a priority: speech disguise, though as we'll see there were no conclusive results
- Loucherbem (aka Louchébem) is the argot of Parisian butchers (l' Argenton 1991)
  - Used for fun and solidarity
  - But also for privacy—l'Argenton gives the example of a customer asking for some special cut that they're out of
    - Junior butcher can ask the boss which *lorsomik de liandvem* (morceau de viande, 'piece of meat') to give the customer
    - Boss can say to give the customer some *loubem de lanchrem loujrem* (bout de tranche rouge, 'red end of a slice') that resembles it (p. 121)
  - Other examples: complaining about a disliked customer, giving a special price to one customer without letting the rest know

- I'll let you have the fun of figuring out how it works from the above.

- Among l'Argneton's examples, most are pretty regular (except that the choice of suffix isn't fully predictable), but the 3-syllable words all have something special going on.

Beaujolais	paradigm gap: they use *rouge* 'red' instead
tire-bouchon	paradigm gap: they use *couteau* 'knife' instead
pardessus	lardeuss
portefeuille	larfeuille
avarié	larisok (??)

  - Presumably, **leaujolaisbem* is just too recoverable for listeners
- So will longer words show more irregularity? Results not conclusive (see over).



- The 1-syllable words have most of the –*e*, no gaps, all the 0s, no gaps, low rate of "other".
- *g,k,kl* tend to take -*εs* the rest (except *m*) tend to take -*εm*

V tends to take *other*

5 Variation in underlying form: Pater et al. 2012

- This has come up here and there this quarter
  - Should the high *t/d*-deletability of certain high-frequency words be attributed to multiple underlying forms: {/mowst/, /mows/}? (Walker 2008)
- This idea also has a history outside of variation
  - E.g., Tranel 1996: in OT you can have multiple allomorphs in the lexicon, and leave it up to the grammar to pick the most unmarked one in each environment.
  - To take the Korean nominative suffix as an example...
    - The 2 constraints at the bottom are hypothetical—either your theory doesn't have them at all, or if it does, in this case they're ranked low

/pi/+{/i/, /ka/}	*COMPLEX	ONSET	NOCODA	NOM=...i#	NOM=...ka#
pi.i		*!			*
☞ pi.ka				*	

/nun/+{/i/, /ka/}	*COMPLEX	ONSET	NOCODA	NOM=...i#	NOM=...ka#
☞ nu.ni					*
nun.ka			*!	*	

/salm/+{/i/, /ka/}	*COMPLEX	ONSET	NOCODA	NOM=...i#	NOM=...ka#
☞ sal.mi			*		*
salm.ka	*!		*	*	

- The puzzle: how does a learner know when to resort to this?
  - Should a Russian speaker learn 'bread' as {/xlep/, /xleb/}, using /xleb/ only when suffixed with a vowel?
    - If they do that for the whole lexicon, won't they fail to devoice in a wug test? (Kenstowicz & Kisseberth 1979)
- Pater & al.'s solution: learn a MaxEnt grammar that gives you a probability distribution over underlying forms
  - Candidates are input-output pairs, and some constraints prefer/penalize the input part (à la Zuraw 2000)

- Russian-type language (final devoicing)

	3.65	0	6.05	401.41	1.94	
'bread' *nom. sg.*	BREAD→/xleb/	BREAD→/xlep/	IDENT(voice)	NOVOICEDCODA	*VTV	*p*
/xleb/→[xleb]		*		*		0.00%
/xleb/→[**xlep**]		*	*			8.32%
/xlep/→[xleb]	*		*	*		0.00%
/xlep/→[**xlep**]	*					91.68%

'bread' *gen. sg.*	BREAD→/xleb/	BREAD→/xlep/	IDENT(voice)	NOVOICEDCODA	*VTV	*p*
/xleb/+a→[**xleba**]		*				99.59%
/xleb/+a→[xlepa]		*	*		*	0.03%
/xlep/+a→[xleba]	*		*			0.01%
/xlep/+a→[xlepa]	*				*	0.37%

- Where do constraints like BREAD→/xleb/ come from?
  - Every observed surface form for 'bread' spawns such a constraint
    - This means no abstract underlying forms, no composite underlying forms
    - Also means that in, e.g., a language with no final devoicing, we don't need to worry about the wrong underlying form:

- English-type language (no final devoicing)

	0	0	43.62	39.83	39.83	
'dog'	DOG→/dag/	PUP→/pʌp/	IDENT(voice)	NOVOICEDCODA	*VTV	*p*
/dag/→[**dag**]				*		97.79%
/dag/→[dak]			*			2.21%

	0	0	43.62	39.83	39.83	
'dog' *diminutive*	DOG→/dag/	PUP→/pʌp/	IDENT(voice)	NOVOICEDCODA	*VTV	*p*
/dag/+i→[**dagi**]						100.00%
/dag/+i→[daki]			*		*	0.00%

	0	0	43.62	39.83	39.83	
'pup'	DOG→/dag/	PUP→/pʌp/	IDENT(voice)	NOVOICEDCODA	*VTV	*p*
/pʌp/→[pʌb]			*	*		0.00%
/pʌp/→[**pʌp**]						100.00%

	0	0	43.62	39.83	39.83	
'pup' *diminutive*	DOG→/dag/	PUP→/pʌp/	IDENT(voice)	NOVOICEDCODA	*VTV	*p*
/pʌp/+i→[pʌbi]			*			2.21%
/pʌp/+i→[**pʌpi**]					*	97.79%

6 Learning

- What if we just try to fit a MaxEnt grammar to the schematic Russian tableaux above, in the usual way?
- What's the important information not in the tableaux above that we'd have to fill in for the MaxEnt Grammar Tool?

- In simple MaxEnt, you ask the computer to maximize the probability of the data (minus a prior that penalized non-default weights).
- Pater et al. instead *minimize* the difference between the observed distribution of surface forms (p^*)—not candidates—and the distribution predicted by the current set of weights (p_w)
  - More specifically, minimize the Kullback-Leibler divergence (KL divergence) of the two distributions:

$$\sum_i \sum_j p^*\left(y_{ij} \mid x_i\right) \log \frac{p^*\left(y_{ij} \mid x_i\right)}{p_w\left(y_{ij} \mid x_i\right)}$$

observed probability of surface form given intended meaning

predicted " "

sum over each surface form (j) for each tableau (i)

- Gaussian prior, sigma squared = 10,000
- Objective function also contains a third term, to maximize (a constant times) the difference between all the faithfulness weights and all the markedness weights.

7 Pater & al.'s application to a case of variation

- We've talked about French schwa (really a front rounded vowel) deletion before
- Different words seem to have idiosyncratically different rates of deletion:

underlying full V (non-schwa)
underlying schwa, deletes a little

underlying schwa, deletes a lot

no underlying V
*CCC prevents deletion

	Word	UR	SR	*p*
a.	femelle	/fømɛl/	[fømɛl]	1
b.	semestre	/sVmɛstʁ/	[sømɛstʁ]	0.8
			[smɛstʁ]	0.2
c.	semelle	/sVmɛl/	[sømɛl]	0.5
			[smɛl]	0.5
d.	Fnac	/fnak/	[fnak]	1
e.	breton	/bʁøtɔ̃/	[bʁøtɔ̃]	1

(p. 69)

- Weights the system ends up learning:

Constraint	Weight
*CCC	467.26
MAX	4.93
'SEMESTRE'→/sømɛstʁ/	4.23
'SEMELLE'→/sømɛl/	2.71
*[ø]	2.58
'SEMELLE'→/smɛl/	0.10
'SEMESTRE'→/smɛstʁ/	0.03
DEP	0.00

(p. 69)

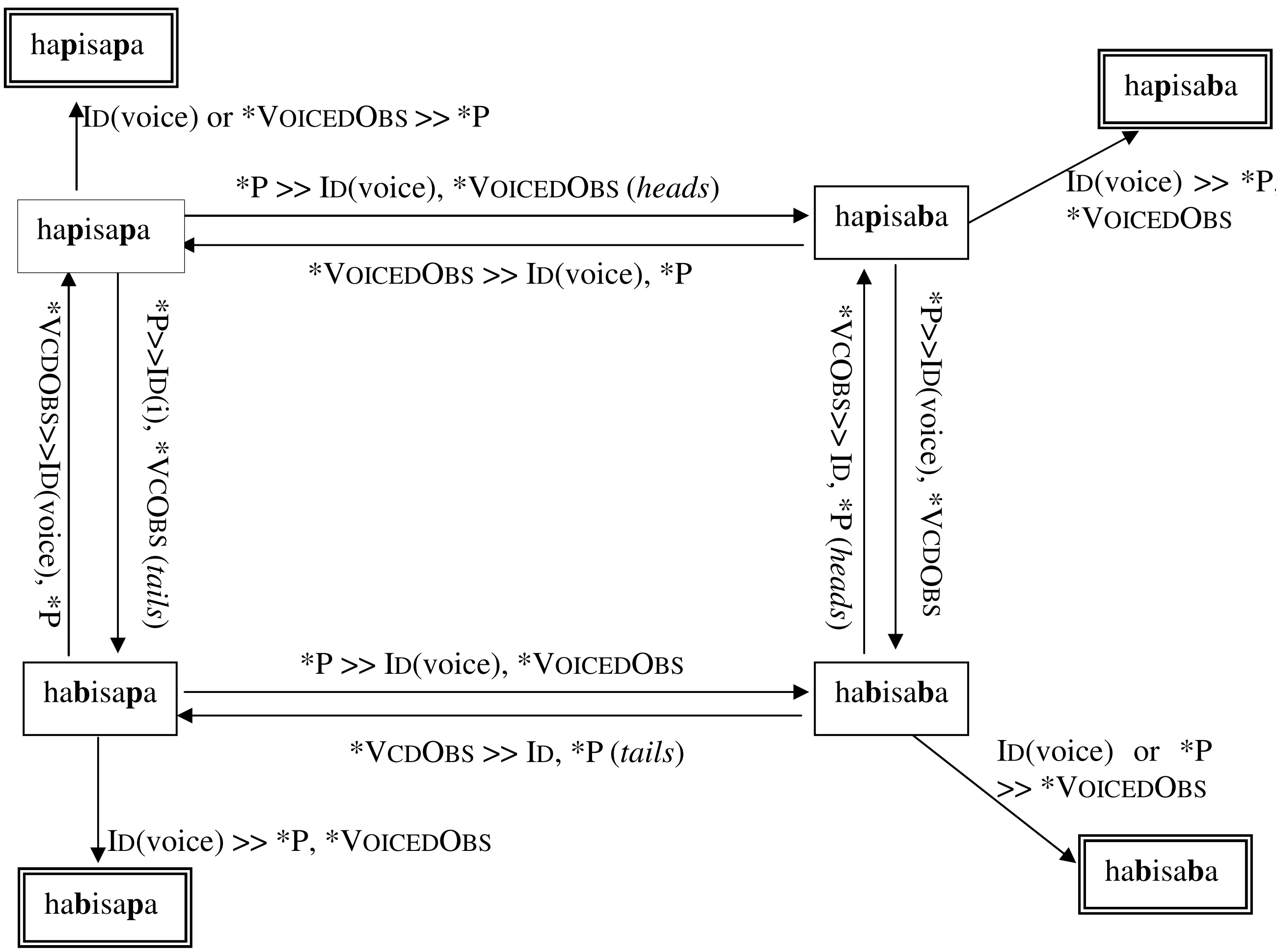
- Probability each *candidate* ends up getting (remember, the system was trained not on candidate probabilities but on surface-form probabilities)

candidate	*p*	*total deletion probability*
femelle→f'melle	0.09	0.09
femelle→femelle	0.91	
semestre→s'mestre	0.08	0.23
s'mestre→s'mestre	0.15	
semestre→semestre	0.77	
s'mestre→semestre	0.01	
semelle→s'melle	0.04	0.49
s'melle→s'melle	0.45	
semelle→semelle	0.47	
s'melle→semelle	0.03	
Fnac→Fnac	0.93	0.93
Fnac→Fenac	0.07	
breton→br'ton	0.00	0.00
breton→breton	1.00	

- What drives the difference between *semestre* and *semelle*?

8 Learning variation in Harmonic Serialism: Staubs & Pater to appear

- Recall from Kimper 2011 the potential for derivational loops (adapting Staubs & Pater's figure for a difference phenomenon)
  - Assume variable constraint ranking (for illustration—weights will work similarly)
    - Ties resolved by coin flip
  - This is Harmonic Serialism, so we can make only one change at a time (no diagonal arrows)



- Any of these candidates' probabilities of winning is the sum of the probabilities of the infinite number of paths that could lead to it.
- We can attach a probability to each arrow above, according to whatever our underlying model of variation is (MaxEnt, NoisyHG, Stochastic OT, etc.).
- Using matrix algebra (see the paper) we could figure out the probability of getting from any beginning state (input) to any final state (output).
- Now the approach from Pater et al. above kicks in
  - Learn weights that minimize the KL divergence between the observed probability distribution (over surface forms) and the predicted one.

- Staubs & Pater run a case of French schwa deletion here too, but with the simplifying assumption that the underlying forms all have the schwa:

(13) Distributions in learning data (Obs.), and produced by the grammar (Fit.)

	Obs.	Fit.			Obs.	Fit.
a. Je me prépare	0.13	0.125		Je te répondrais	0.6	0.605
b. J' me prépare	0.07	0.075		J' te répondrais	0.38	0.374
c. Je m' prépare	0.8	0.8		Je t' répondrais	0.02	0.02
d. J' m' prépare	0	0		J' t' répondrais	0	0

(p. 11)

9 Some interesting properties of the resulting grammar

- Although *Je te...*'s probability of winning on the first round is only 46.3%, it has a higher probability of winning overall
  - If (b) or (c) wins on the first round, (a) still might win on a subsequent round.

(14) First step from /ʒœtœ/

ʒœtœ	*CCC 4.7	*t' 3.6	*SCHWA 1.98	*j' 1.84	
a. 0.425 ʒœtœ			–2		–3.97
b. 0.49 ʒtœ			–1	–1	–3.82
c. 0.08 ʒœt		–1	–1		–5.58

(p. 12)

- Why is the dual deletion so improbable—candidates (d), *J'm'* and *J't'* :
  - Going from (b) or (c) to (d) is improbable because of *CCC
    - 0.1%, 0.2%
  - And even if you do get to (d), going from (d) to (d) (convergence) is improbable too, again because of *CCC

(17) A third step from /ʒœte/

ʒt	*CCC 4.7	*t' 3.6	*SCHWA 1.98	*j' 1.84	
a. 0.851 ʒtœ			–1	–1	–3.82
b. 0.147 ʒœt		–1	–1		–5.58
c. 0.002 ʒt	–1	–1		–1	–10.13

(p. 13)

10 Exemplar models

- Pater et al. use (indirectly) a probability distribution over underlying forms, each identical to some observed surface form.
- Exemplar approaches assume that a "lexical entry" is actually a huge set of remembered surface forms.

Background

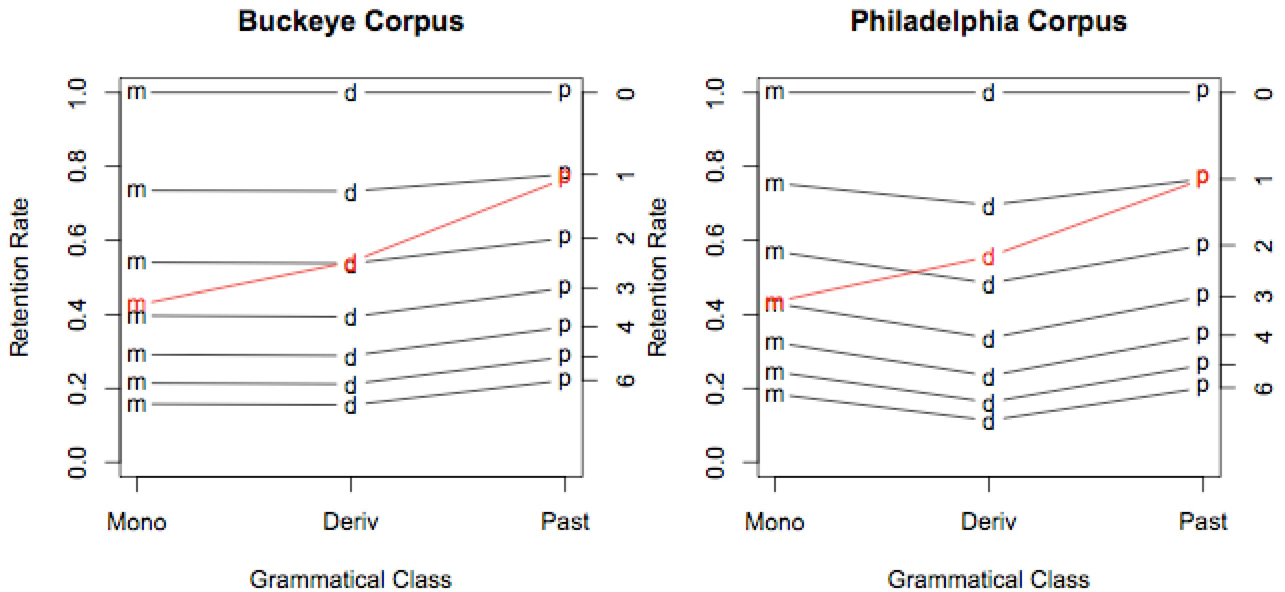
- Medin & Schaffer 1978, a picture-categorization study, is an important early reference
  - Johnson 1997 is an important ref in phonetics
  - Bybee (2001) has been a big advocate
  - Kirchner 1998 applies exemplar ideas to the problem of phonologization
- <u>Categories</u> are represented mentally as clouds of <u>remembered tokens</u> (in a multidimensional perceptual space).
  - That space might include dimensions of talker identity (e.g., Blair/Serena, male/female—see Johnson 2006), social situation
- Highly similar tokens may be grouped into a single <u>exemplar</u> (i.e., the similarity map is granular)
  - The exemplar's strength is augmented when tokens are added to that "bin"
  - Each exemplar decays over time, so that more-recent tokens have more say
- Low-frequency words, having fewer items already in their cloud, are more susceptible to influence from recent tokens (see Goldinger 1998).
- An incoming <u>stimulus</u> is categorized according to the number of exemplars from each category that are similar to it
  - with a weighting in favor of stronger exemplars.
- A given token may contribute to categories at multiple levels of structure (feature, phoneme, word—see Nielsen 2011.

Production

- To produce an instance of a category, choose an exemplar at random
  - But stronger exemplars have a higher probability of being chosen.
- Add some noise, so that the actual production may differ slightly from the exemplar chosen.
  - See Pierrehumbert 2001 for some considerations of averaging and drift

11 Fruehwald 2008: applying an exemplar model to *t/d*-deletion

- Assumes a word is represented as a cloud of tokens, with more and less *t/d*
- In production, after an exemplar is drawn from the cloud, it's subject to further *t/d*-deletion according to its phonological context (__V, __C, __*pause*)
- The resulting production is added back into the cloud
- Result: words that tend to occur in more-deleting context (__C) should have a less *t/d*-ful representation
  - They should get produced with less *t/d* even when they're in a different context.
- Runs a simulation starting with 100% *t/d*-ful representations
  - Tests the idea that distribution alone can explain the differences among monomorphemes (*just*), semi-weak pasts (*left*), and regular pasts (*jumped*).



(slide 30)

- Result: from iteration 0 (top) to 6 (bottom), amount of retention goes down
  - but the morphological difference never gets as big as the observed one (slanted line)
- Concludes that distribution isn't enough to explain the morphological differences
  - And that the exponential model is doing pretty well.

Integrating exemplar ideas into constraint grammars?

- The proposals I've seen are not implemented models, but rather more schematic ideas.
- Let's take a tour...

12 Ettlinger 2007: How variation could make opacity less bad

- Take a typical counterfeeding example (schematic Sea Dayak, from Ettlinger):

Nasality spreads rightward	/mata/	mãta	'eye'
Post-nasal stop deletion fails to feed nasal spreading			
	/ramboʔ/	ramoʔ *ramõʔ	'plant sp.'

- But actually, there's variation—could this help?

  /ramboʔ/ rambo? ~ ramoʔ

(The paper also discusses the Shimakonde variable vowel reduction case we saw earlier, Canadian French laxing, and Finnish assibilation.)

- Sea Dayak analysis
  - You've heard tokens of [mãta]—those form your underlying representation
  - For the sake of argument (Richness of the Base), suppose that you stripped the redundant nasalization out of some or all tokens
  - No matter, the ranking still ensures nasal spreading:

/m{a,ã}ta/	*NV_{ORAL}	*Ṽ	IDENT(NASAL)/V
mata	*!		?
✦ mãta		*	?

    - This wasn't in the paper, so I'm guessing a little—sorry to Ettlinger if I got it wrong

  - You've heard tokens of both [ramboʔ] and [ramoʔ]
  - Again for the sake of Richness of the Base, suppose you also hallucinated nasalization
  - No matter, the ranking still ensures nasal spreading:

/ram{b,Ø}{o,õ}ʔ/	*NV_{ORAL}	*Ṽ	IDENT(NASAL)/V
ram{b,Ø}õʔ		*!	?
✦ ram{b,Ø}oʔ			?

    - I'm taking yet more liberties here—Ettlinger doesn't discuss the rich-base possibility
  - Ettlinger's idea here is that the decision as to whether to produce the [b] is made in a later phonetic component, taking into account factors like formality.

13 A prediction Ettlinger makes

- You couldn't have a language where instead stop *insertion* counter*bleeds* nasal spreading
- Ettlinger refers to this impossible pattern as Sky Dayak:

/ramoʔ/ ramõʔ ~ rambõʔ

- Why would Sky Dayak be impossible in Ettlinger's approach?

14 An earlier theory this is based on: Kawahara 2002's variant-variant faithfulness

- How I think Kawahara would handle Sea Dayak
  - Careful speech—no deletion

/ramboʔ/	MAX-C	*NC	*NV_{ORAL}	*Ṽ	IDENT(nasal)/V
✦ ramboʔ		*			
rambõʔ		*		*	*
ramoʔ	*!		*		
ramõʔ	*!			*	*

  - Casual speech—promote *NC / demote MAX-C

/ramboʔ/ cf. variant [ramboʔ]	**IDENT-OutputVariant(nas)**	*NC	MAX-C	*NV_{ORAL}	*Ṽ	IDENT(nas)/V
ramboʔ		*!				
rambõʔ	***!**	*			*	*
✦ ramoʔ			*	*		
ramõʔ	***!**		*		*	*

- It's kind of like Sympathy (McCarthy 1999), except playing the role of the sympathy candidate is an existing output
- This requires one form to be canonical ("Output") and the other non-canonical ("Variant")

o Ettlinger points out that this approach *could* get "Sky Dayak"—how?

15 van de Weijer 2012: *output* candidates as exemplars

- Input is the communicative intent
  - Sloos 2013 explores the idea that the input is the word's prototype
    - Prototype: an exemplar (or composite) that's selected as most typical/best example of its category
- Candidates are exemplars, with their frequency/activation
- Roughly, Harmonic Grammar
  - Satisfying a constraint gives you a bonus point
  - Violations/satisfactions are multiplied by frequency (as well as by weight)
- Schematic example, for 'incomplete'

(24)	AGREE PLACE 4	*V_{NASAL} 3	NOCODA 2	IDENT-PFX 1	result
7 [ɪŋk...]	28	21	-14	-7	28
3 [ɪnk...]	-12	9	-6	3	-6
2 [ĩk...]	8	-6	4	-2	4

(p. 58)

- Open questions
  - Productivity: what if you haven't yet heard the exemplar you need?
  - Does it make sense that more-frequency exemplars should violate both markedness and faithfulness more strongly?
  - How should faithfulness constraints be evaluated?

16 One last performance-oriented account of variation: Bíró 2005

- Assumes totally standard OT—BUT, the algorithm for generating an output isn't guaranteed to find the right answer
- **Simulated annealing**: like the hill-climbing techniques we discussed for finding the best MaxEnt weights...
  - ...except sometimes you can go downhill[1] instead
  - A parameter called **temperature** determines how willing you are to go downhill.
    - Temperature decreases over time to force convergence

[1] In MaxEnt, we were maximizing our objective function, so we wanted to go uphill. More often you're trying to *minimize* something (e.g., error), so you actually want to go *downhill*. When reading discussions of hill-climbing and the like, first figure out whether up is good or down is good.

- This is all a metaphor from metallurgy: as long as you're applying heat, the material can use that heat energy to go from a "better" (lower-energy) to a "worse" state
  - but as the heat decreases, there's nowhere to "steal" energy from, and the material has to settle for the local optimum.
- This can be a good idea when your search space isn't convex
  - It can rattle you out of a local optimum
- Instead of moving around in a continuous space, in the OT case, move from one OT candidate to an "adjacent" candidate
  - Bíró defines a topology on the candidates—which ones are **connected**?
  - Assumes something similar to harmonic serialism: two candidates are connected if they differ minimally.
  - Example for stress/foot assignment. *s* = stressed, *u* = unstressed, [ ] = foot

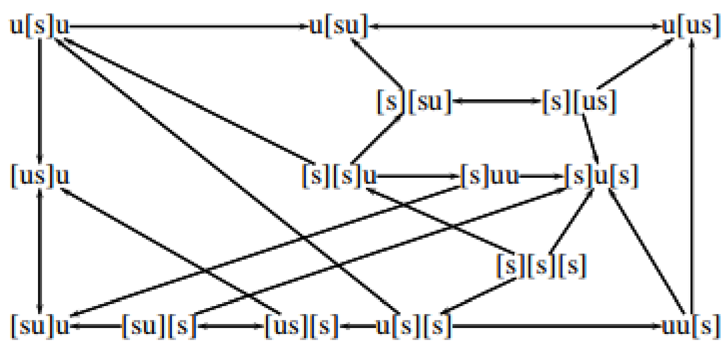


Figure 4: Search space (candidate set, neighbourhood structure) for a three-syllable word. (p. 12)

- How do you know whether you're going uphill or downhill?
  - Find the first constraint that distinguishes between candidates (here, C_k)
  - Find the difference in violations on that candidate (here, -1)

	C_n	C_{n-1}	...	C_{k+1}	C_k	C_{k-1}	C_{k-2}	...
$E(w')$	2	0		1	2	3	0	
$E(w)$	2	0		1	3	1	2	
$E(w') - E(w)$	0	0		0	-1	2	-2	

(p. 6 of corrected version)

  - Difference between *E(w')* and *E(w)* is ⟨*k*, -1⟩ --that is, *E(w)* is worse, at constraint *k*—the **fatal constraint**—by 1 violation.
- How do you decide whether to make the move?
  - Temperature also has two components: ⟨*K, t*⟩
  - *K* is the highest that *k* can be
    - The higher the temperature, the higher-ranked can be the fatal constraint
    - If the fatal constraint is higher than *K*, no move
    - If it's lower than *K*, yes move
  - If the fatal constraint is *K*, *t* determines how probable the move is (probability: $e^{-d/t}$)

17 Bíró's model of stress shift in Dutch rapid speech

- Cites a production study (Schreuder & Gilbers 2004) of Dutch stress
  - Participants were forced to talk fast (answer multiple-choice question as quickly as possible) and normally (reading words in frame sentences)
  - There were tendencies for certain stress shifts in rapid speed (S&G p. 184):

"andante" (slow) O-O faithful	"allegro" (fast) seems to ignore morphology	reason for shift	
stú.die+**tòe**.la.ge	stú.die.toe.**là**.ge	*penultimate default?*	'study grant'
wég.werp+**àan**.ste.ker	wég.werp.aan.**stè.k**er		'disposable lighter'
ká.mer+**vòor**.zit.ker	ká.mer.voor.**zìt**.ler		'chairman of Parliament'
per.**fèc**.tio.n+íst	**pèr**.fec.tio.níst	*initial dactyl preference?*	'perfectionist'
a.**mè**.ri.k+áan	**à**.me.ri.káan		'American'
pi.**rà**.te+ríj	**pì**.ra.te.ríj		'piracy'
zùid+**à**.fri.káans	zùid**a**frikáans	*eliminate clash*	'South African'

- Schreuder and Gilbers think of this in terms of constraint re-ranking (Bíró's tableaux):

(7) *Slow (andante) speech:*

fototoestel	OO-CORR.	*ΣΣ	PARSE-σ
☞ (fóto)(tòestel)		*	
(fóto)toe(stèl)	*!		*

(8) *Fast (allegro) speech:*

fototoestel	*ΣΣ	OO-CORR.	PARSE-σ
(fóto)(tòestel)	*!		
☞ (fóto)toe(stèl)		*	*

(p. 10)

- But Bíró proposes that speakers are using the same ranking, but are under more time pressure in rapid speech
  - Temperature (the *t* part, in ⟨*K*, *t*⟩) decreases more rapidly, increasing probability of "wrong" winner.

Results

fo.to.toe.stel 'camera'	*uit.ge.ve.rij* 'publisher'	*stu.die.toe.la.ge* 'study grant'	*per.fec.tio.nist* 'perfectionist'
OOC to: susu	ssus	susuu	usus
fó.to.tòe.stel fast: *0.82* slow: *1.00*	*úit.gè.ve.rìj* fast: *0.65* / **0.67** slow: *0.97* / **0.96**	*stú.die.tòe.la.ge* fast: *0.55* / **0.38** slow: *0.96* / **0.81**	*per.féc.tio.nìst* fast: *0.49* / **0.13** slow: *0.91* / **0.20**
fó.to.toe.stèl fast: *0.18* slow: *0.00*	*úit.ge.ve.rìj* fast: *0.35* / **0.33** slow: *0.03* / **0.04**	*stú.die.toe.là.ge* fast: *0.45* / **0.62** slow: *0.04* / **0.19**	*pér.fec.tio.nìst* fast: *0.39* / **0.87** slow: *0.07* / **0.80**

Table 1: Simulated (in italics) and observed (in bold; Schreuder, 2005) frequencies. The simulation used $T_{step} = 3$ for fast speech and $T_{step} = 0.1$ for slow speech.

(p. 14)

- Does a great job of not picking a different, truly ungrammatical candidate.
- Does a decent job of match the frequencies for the two top candidates in the two speech styles
  - Except for predicting far too much slow-speech shift in *perfectionist*.

- See Bíró 2006 for full dissertation, with additional cases (Dutch voice assimilation, Hungarian article alternation,

References

Argenton, Françoise Robert l'. 1991. Larlepem largomuche du Louchébem: Parler l'argot du boucher. *Langue Française*(90). 113–125. doi:10.2307/41558516 (2 June, 2013).

Becker, Michael, Lauren Eby Clemens & Andrew Nevins. 2011. A richer model is not always more accurate: the case of French and Portuguese plurals. Manuscript. Indiana University, Harvard University, and University College London, ms.

Becker, Michael, Nihan Ketrez & Andrew Nevins. 2011. The surfeit of the stimulus: analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87(1). 84–125. (15 March, 2012).

Becker, Michael, Andrew Nevins & Jonathan Levine. 2012. Asymmetries in generalizing alternations to and from initial syllables. *Language* 88(2). 231–268. doi:10.1353/lan.2012.0049 (1 June, 2013).

Bíró, Tamás. 2005. When the hothead speaks: simulated annealing Optimality Theory for Dutch fast speech. In Ton van der Wouden, Michaela Poss, Hilke Reckman & Crit Cremers (eds.), *Computational Linguistics in the Netherlands 2004: Selected papers from the fifteenth CLIN meeting*, 13–28. (LOT Occasional Series). Utrecht.

Bíró, Tamás. 2006. *Finding the right words: implementing Optimality Theory with simulated annealing*. Groningen: GroDiL.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Ettlinger, Marc. 2007. Variation as a Window into Opacity. *Proceedings from the Annual Meeting of the Chicago Linguistic Society* 43(1). 61–75. (4 June, 2013).

Fruehwald, Josef. 2008. Simulation and evaluation of exemplar theoretic -t/-d deletion. *NWAVE 37*. University of Pennsylvania.

Giavazzi, Maria. 2010. The phonetics of metrical prominence and its consequences for segmental phonology. MIT PhD dissertation.

Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105(2). 251–279. doi:10.1037/0033-295X.105.2.251.

Johnson, Keith. 1997. Speech perception without speaker normalization. In K Johnson & J Mullenix (eds.), *Talker variability in speech processing*, 145–165. San Diego: Academic Press.

Johnson, Keith. 2006. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics* 34(4). 485–499. doi:10.1016/j.wocn.2005.08.004 (3 June, 2013).

Kawahara, Shigeto. 2002. Faithfulness among variants. *On-in-kenyuu [Phonological Studies]* 5. 47–54.

Kenstowicz, Michael & Charles Kisseberth. 1979. *Generative Phonology: Description and Theory*. New York: Academic Press.

Kimper, Wendell A. 2011. Locality and globality in phonological variation. *Natural Language & Linguistic Theory*.

Kirchner, Robert. 1998. *Preliminary thoughts on "phonologization" within an exemplar-based speech-processing system*. Edmonton, AB.

McCarthy, John J. 1999. Sympathy and Phonological Opacity. *Phonology* 16(3). 331–399.

Medin, Douglas L. & Marguerite M. Schaffer. 1978. Context theory of classification learning. *Psychological Review* 85(3). 207–238. doi:10.1037/0033-295X.85.3.207.

Nielsen, Kuniko. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39(2). 132–142. doi:10.1016/j.wocn.2010.12.007 (3 June, 2013).

Pater, Joe, Robert Staubs, Karen Jesney & Brian Smith. 2012. Learning probabilities over underlying representations. *Proceedings of the Twelfth Meeting of the ACL-SIGMORPHON: Computational research in phonetics, phonology, and morphology*. 62–71.

Pierrehumbert, Janet. 2001. Exemplar dynamics: word frequency, lenition, and contrast. In Joan Bybee & Paul Hopper (eds.), *Frequency effects and the emergence of linguistic structure*, 137–157. Amsterdam: John Benjamins.

Schreuder, Maartje & Dicky Gilbers. 2004. The influence of speech rate on rhythm patterns. In Dicky Gilbers, Maartje Schreuder & N Knevel (eds.), *On the boundaries of phonology and phonetics*. University of Groningen.

Sloos, Marjoleine. 2013. *Phonological Grammar and Frequency: an Integrated Approach: Evidence from German, Indonesian and Japanese*.

Staubs, Robert & Joe Pater. to appear. Learning serial constraint-based grammars. In John J McCarthy & Joe Pater (eds.), *Harmonic Grammar and Harmonic Serialism*. London: Equinox Press.

Tranel, Bernard. 1996. Exceptionality in Optimality Theory and final consonants in French. In Karen Zagona & Karen Zagona (eds.), *Grammatical Theory and Romance Languages*, 275–291. Amsterdam: John Benjamins.

Ussishkin, Adam & Andrew Wedel. 2002. Neighborhood density and the root-affix distinction. In Mako Hirotani & Mako Hirotani (eds.), *Proceedings of NELS 32*, 539–549. Amherst, MA: GLSA.

Walker, James. 2008. Form, function, and frequency in Phonology: (t/d)-Deletion in Toronto. *New Ways of Analyzing Variation 37*. Rice University, Houston.

Weijer, Jeroen van de. 2012. *Grammar as selection: combining Optimality Theory and exemplar theory*. Nagoya: Kougaku.

Zuraw, Kie. 2000. Patterned exceptions in phonology. University of California, Los Angeles PhD dissertation.