Introduction to constraint-based phonology

ACTL Summer School 2018

Adam J. Chong

a.chong@qmul.ac.uk

DAY 4

Within-language variation using constraints I

Where we've been:

- ~~M: Phonotactics and alternations; Intro to OT~~
- ~~T: More on OT (finish up basic mechanics)~~
- ~~W: Constraint interaction: TETU + Contrasts + Conspiracies~~
- Th: Constraint approaches to within-lg. variation I
- F: Constraint approaches to within-lg. variation II; Wrap-up

Where we've been

- A short historiography of why phonology when in the constraint direction (M)
- Mechanics of OT (M, T)
- Some consequences of constraint interaction (W)
- CROSS-linguistic variation with constraints (W) – software that can help! OTSoft!

- What about WITHIN-language variation? (Th, F)
  - Augmenting constraints to account for variable patterns within a language
  - (See also Yuni Kim's class here at ACTL!)

What is variation?

- Phonological variation: “a situation in which a single morpheme can be realized in more than one phonetic form in a single environment.” (Coetzee & Pater, 2011)

- Phonological variation a central part of current research (see also Yuni’s class here at ACTL for specific kinds of non-phonological influences on phonology: morphological influences).

Phonological variation

- The claim here is that variation is **grammatical** – i.e. the degree of variability (so, say, relative proportion of forms) should be modeled in the grammar

- It can't just be boiled down to 'performance' (vs. 'competence') – you can't just put this into the bin of 'performance'

Phonological variation in phonological theory

- Studies of linguistic variation have achieved a high level of quantitative precision in describing the systematic patterns of “orderly heterogeneity" (Weinreich et al. 1968, 100) that permeate human use of language. [:::] With [variable rule] techniques and an impressive body of empirical studies completed, the field can be said to have achieved a certain level of descriptive adequacy. This descriptive precision is not generally matched, however, by explanatory precision. That is, the analyst usually cannot say why the quantitative values obtained should have the precise values they do. [:::] Theories that predict particular quantitative values for linguistic variables are in very short supply in linguistics. The development of models that have explanatory value in this sense - models from which one can derive precise quantitative predictions - is one of the fundamental challenges facing our discipline (Guy 1991a, 1-2).

Phonological variation in phonological theory

- "This study is an attempt to reconcile variation with generative phonology. It arose out of the frustrating experiences I had establishing certain basic facts of Finnish phonology: rules that were absolutely central to the system sometimes applied with full force, sometimes they were optional. Solving the problem by labelling these rules "optional" did not even begin to work. It was hard to figure out the environments where optionality was supposedly invoked and, even worse, I soon began to realize that I was faced with optionality of the more-or-less sort: in many cases several variants were possible, but one of them was clearly preferred. The crucial problem was how to capture these gradient, yet robust facts in a generative framework. Given theoretical linguists' goal of moving away from rich descriptive devices towards restrictive theories, it was all the more puzzling to be faced with a phenomenon that current phonological theories seemed unable even to describe." (Anttila, 1997:1)

Variation

- Within-language phonology – a single morpheme in the same context can have more than one phonological form.

- Historically – realm of sociolinguists!

A classic case of variation: social/stylistic variation

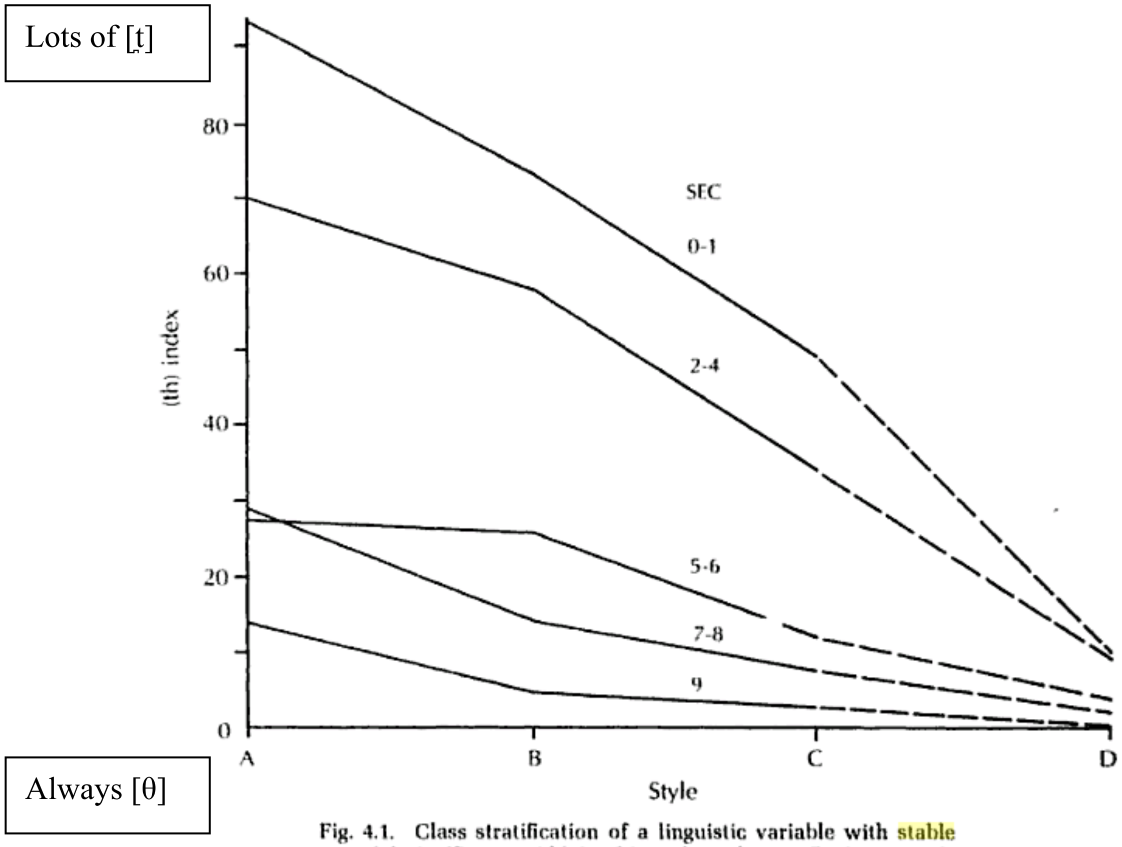


Fig. 4.1. Class stratification of a linguistic variable with stable social significance: (th) in *thing, through*, etc. Socioeconomic class scale: 0–1, lower class; 2–4, working class; 5–6, 7–8, lower middle class; 9, upper middle class. A, casual speech; B, careful speech; C, reading style; D, word lists.

- New York City English, variation in words like *three*
- Within each social group, more [θ] as style becomes more formal
- Each utterance of /θ/ scored as 0 for [θ], 100 for [t̪]
- Each speaker gets an overall average score.

Labov 1972 p. 113

Some common types of variation

- **Token** variation ('free' variation)
  - The same speaker can produce both variants of a given word
  - No differences in meaning (although maybe different stylistic factors, etc.)
  - One variant may be more frequent, but the rate is the same for all target morphemes/triggers

A classic case of variation: social/stylistic variation

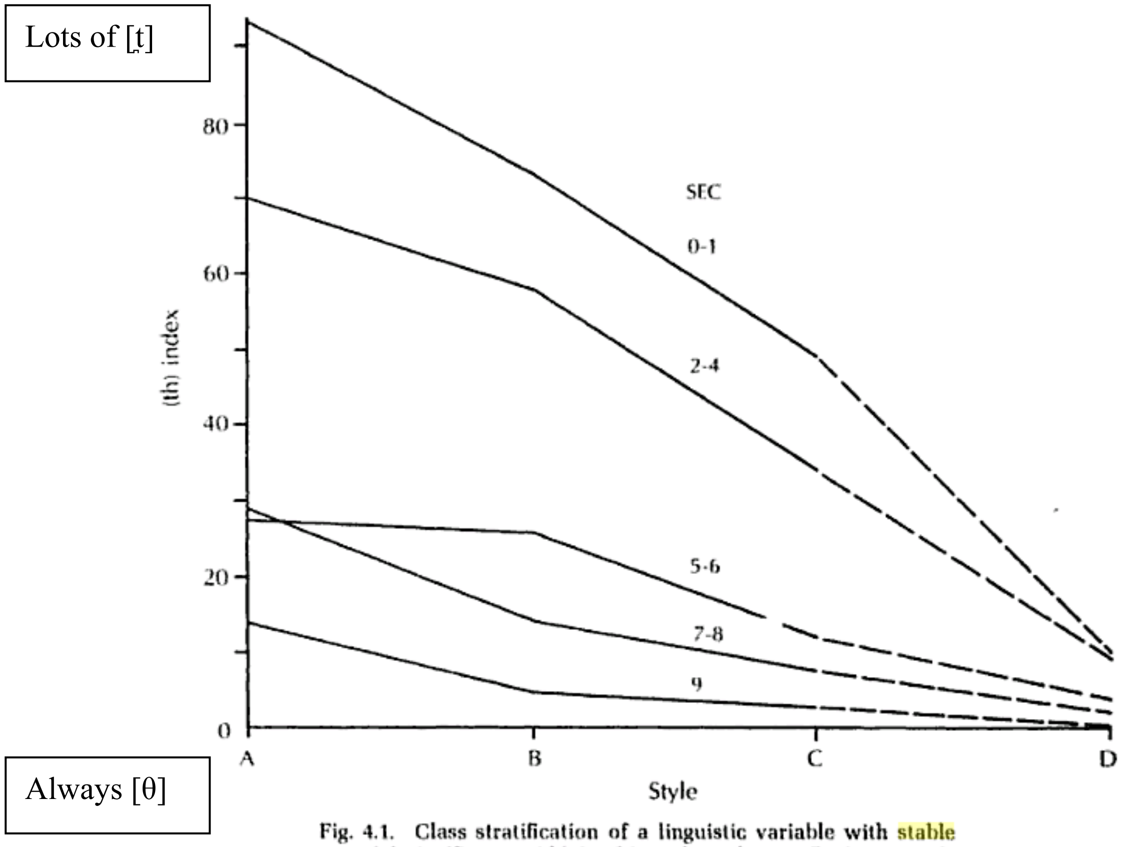


Fig. 4.1. Class stratification of a linguistic variable with stable social significance: (th) in *thing, through*, etc. Socioeconomic class scale: 0-1, lower class; 2-4, working class; 5-6, 7-8, lower middle class; 9, upper middle class. A, casual speech; B, careful speech; C, reading style; D, word lists.

- Each word with /θ/ can vary
  - *think*: [θɪŋk] ~ [t̪ɪŋk]
  - *Cathy*: [kæθi] ~ [kæt̪i]
  - etc.
- The variation can be conditioned by various factors
  - style (what Labov looks at here)
  - word frequency
  - part of speech
  - location of stress in word
  - preceding or following sound
  - etc.

Labov 1972 p. 113

Some common types of variation

- **Type** variation (lexical variation)
  - A particular lexical item has a fixed behaviour
  - So variation is **across**, not between items

- E.g. in English v/f plurals – can we think of the pattern?

E.g. Tagalog (Zuraw 2009)

Tagalog: Austronesian language from the Philippines with ~17 million native speakers (Ethnologue 2005, data from Zuraw 2009's corpus ; see also Schachter & Otanes 1972)

- d → ɾ / V__V :

dunoŋ	'knowledge'	ma-**ɾ**unoŋ	'intelligent'
dinig	'heard'	ma-**ɾ**inig	'to hear'
dupok		ma-**ɾ**upok	'fragile'

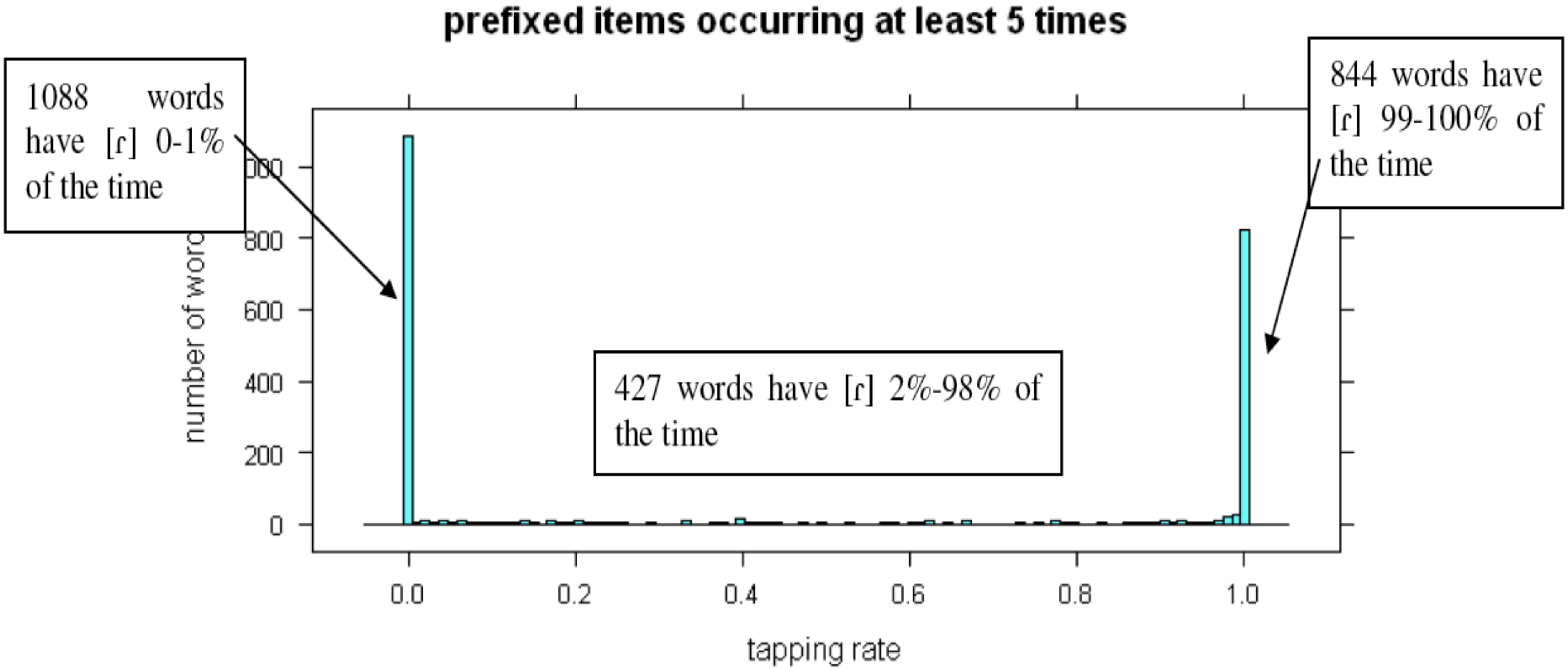
- But, there are also words like this

daʔig	'beaten'	ma-**d**aʔig	'beaten'
dulas	'slipperiness'?	ma-**d**ulas	'slippery'
daʔan	'road'	ma-**d**aʔan-an	'passable'

- and like this

duŋis	'dirt on face'	ma-**ɾ**uŋis ~ ma-**d**uŋis	'dirty (face)'
dumi	'dirt'	ma-**ɾ**umi ~ ma-**d**umi	'dirty'

Data from Zuraw (2009)



Accounting for variation using rules

- Traditionally rules that were variable were just marked with a diacritic to say they were *optional*

  /θ/ → [–continuant] , **optional rule**

Labov makes a model of the whole speech community

- but let's be more conservative and suppose that each group can have a different grammar.

Each group's grammar has its own **numbers** *a* and *b* such that:

- (th)-index = *a* + *b**Style
- where Style A=0, B=1, C=2, D=3

Some things to note

- Phonological factors and stylistic/social ones are modeled in the same model
- They have the same status – just another independent predictor

How is variation accounted for in constraint-based grammars?

- Brief look at four different models:
  - Partial-ranking (Today)
  - Stochastic OT (Today)
  - Noisy Harmonic Grammar (Friday)
  - MaxEnt Harmonic Grammar (Friday)

Constraint rankings

- Recall that cross-linguistic variation is captured by different rankings of constraints.

- By the same token, within language variation should fall out from this very same property!

Partial-ranking in OT (Anttila, 1997)

- Key idea: Some constraints do have fixed rankings, but some don't and are crucially unranked with respect to one another

- Each time a speaker wants to produce a form, the unranked constraints need to be put into a ranking.

Illustration of partial-ranking

	/θɪk/	Max-C	Ident(place)	*θ	Ident(cont)	*Dental
☞ *a*	[θɪk]			*		*
☞ *b*	[t̪ɪk]				*	*
c	[ɪk]	*!				
d	[sɪk]		*!			

Linearization

- In order to generate a form, the constraints have to be put into a linear order
- Each linear order consistent with the grammar's partial order is equally probable

grammar

Max-C

Id(place)

*θ   Id(cont)

*Dental

linearization 1 (50%)	**linearization 2 (50%)**
Max-C	Max-C
Ident(place)	Ident(place)
*θ	Ident(cont)
Ident(cont)	*θ
*Dental	*Dental
☞ [t̪ɪk]	☞ [θɪk]

Some properties of this theory

- Unfortunately: No straightforward learning algorithm

- Makes strong predictions about variation *numbers*:
  - If there are 2 constraints that are partially-ranked, what are the possible Anttilan grammars?

Finnish example (Anttila 1997)

- The genitive suffix has two forms
  - “strong”: -iden/-iten (with additional changes)
  - “weak”: -(j)en (data from p. 3)

Strong form: heavy penult (CVV, CVVC), final syllable onset /t, d/
Weak form: light penult (CV), final syllable onset /j/ or absent

a.	/puu/ ‘tree’	**pui.d**en
	/potilas/ ‘patient’	po.ti.**lai.d**en
b.	/kala/ ‘fish’	ka.**lo.j**en
	/margariini/ ‘margarine’	mar.ga.rii.**ni**.en
c.	/naapuri/ ‘neighbor’	naa.pu.**rei.d**en ~ naa.pu.**ri**.en
	/Reagani/ ‘Reagan’	Rea.ga.**nei.d**en ~ Rea.ga.**ni**.en
	/moskeija/ ‘mosque’	mos.kei.**joi.d**en ~ mos.kei.**jo.j**en
	/ministeri/ ‘minister’	mi.nis.te.**rei.d**en ~ mi.nis.te.**ri**.en

Factors affecting variation

Stem length	Stem-final syllable	Variant(s)
1 syllable	(irrelevant)	strong
2 syllables	heavy (CVV, CVC)	strong
	light (CV)	weak
3 syllables	heavy (CVV, CVC)	strong
	light (CV)	strong~weak
4 syllables	heavy (CVV, CVC)	strong
	light (CV)	strong~weak
5 syllables	heavy (CVV, CVC)	strong
	light (CV)	strong~weak

Factors affecting variation

Strong form: heavy penult (CVV, CVVC), final syllable onset /t, d/
Weak form: light penult (CV), final syllable onset /j/ or absent

a.	/puu/ 'tree'	**pui.d**en
	/potilas/ 'patient'	po.ti.**lai.d**en
b.	/kala/ 'fish'	ka.**lo.j**en
	/margariini/ 'margarine'	mar.ga.rii.**ni.**en
c.	/naapuri/ 'neighbor'	naa.pu.**rei.d**en ~ naa.pu.**ri.**en
	/Reagani/ 'Reagan'	Rea.ga.**nei.d**en ~ Rea.ga.**ni.**en
	/moskeija/ 'mosque'	mos.kei.**joi.d**en ~ mos.kei.**jo.j**en
	/ministeri/ 'minister'	mi.nis.te.**rei.d**en ~ mi.nis.te.**ri.**en

- Anttila shows that choice is governed by...
  - avoiding sequence of heavy or light syllables (*HH, *LL)
  - avoiding high vowels in heavy syllables (*H/I) or low vowels in light syllables (*L/A)

Anttila's grammar (p. 21)

We won't go through the entire analysis

SET 1	SET 2	SET 3	SET 4	SET 5
*X́.X́	*Ĺ	*H/I	*H/O	*H/A
	*H	*Í	*Ó	*Á
		*L.L	*L/A	*L/O ≫ *L/I
			*H.H	*A ≫ *O ≫ *I
			*H́	*L
			*X.X	

Sample of the results (p. 23)

(53) Predictions and observed frequencies: 3-syllabic stems

	Motif	Pred. %	Obs. %	Examples
1a	L.TÁA	100	99.4 (720)	ká.me.ròi.den
1b	L.TA	0	0.6 (4)	ká.me.ro.jen
2a	L.TÓO	100	99.5 (389)	hé.te.ròi.den
2b	L.TO	0	0.5 (2)	hé.te.ro.jen
3a	L.TÍI	33	36.9 (215)	náa.pu.rèi.den
3b	L.TI	67	63.1 (368)	náa.pu.ri.en
4a	H.TÁA	50	50.5 (46)	máa.il.mòi.den
4b	H.TA	50	49.5 (45)	máa.il.mo.jen
5a	H.TÓO	20	17.8 (76)	kór.jaa.mòi.den
5b	H.TO	80	82.2 (350)	kór.jaa.mo.jen
6a	H.TÍI	0	1.6 (13)	pó.lii.sèi.den
6b	H.TI	100	98.4 (806)	pó.lii.si.en

Local summary: Partial-ranking

- Re-ranking of constraints – gives us different CROSS-linguistic possibilities

- But the same principle can allow us to account for WITHIN-language grammatical variation!

Stochastic OT

- A less-restrictive model of variation than partial-ranking
- ‘Stochastic’ – loosely ‘probabilistic’
- But first – a quick word on probability

What is a probability distribution

- It’s a function from possible outcomes (of some random variable) to probabilities.
- A simple example: flipping a fair double-sided coin

which side lands up	probabiliy
heads	0.5
tails	0.5

There are only two possible outcomes! So you expect each case to be equally probable (all else being equal)

Rolling 2 dice

sum of 2 dice	probability
2 (1+1)	1/36
3 (1+2, 2+1)	2/36
4 (1+3, 2+2, 3+1)	3/36
5 (1+4, 2+3, 3+2, 4+1)	4/36
6 (1+5, 2+4, 3+3, 4+2, 5+1)	5/36
7 (1+6, 2+5, 3+4, 4+3, 5+2, 6+1)	6/36
8 (2+6, 3+5, 4+4, 5+3, 6+2)	5/36
9 (3+6, 4+5, 5+4, 6+3)	4/36
10 (4+6, 5+5, 6+4)	3/36
11 (5+6, 6+5)	2/36
12 (6+6)	1/36

Probability distributions over grammars

- One way to think about within-speaker variation is that, at each moment, the speaker has multiple grammars to choose between.
  - So at each time, there's a probability of picking one possible grammar
  - But – generally in models that account for variation, the probability distributions are constrained somehow
    - (Not completely unconstrained variation)
- So it is similar to Anttila's model – but it differs in the possible outcomes – it is less restrictive than Anttila's model

Anttilan partial ranking as a probability distribution over Classic OT grammars

Id(place)

*θ Id(cont)

means

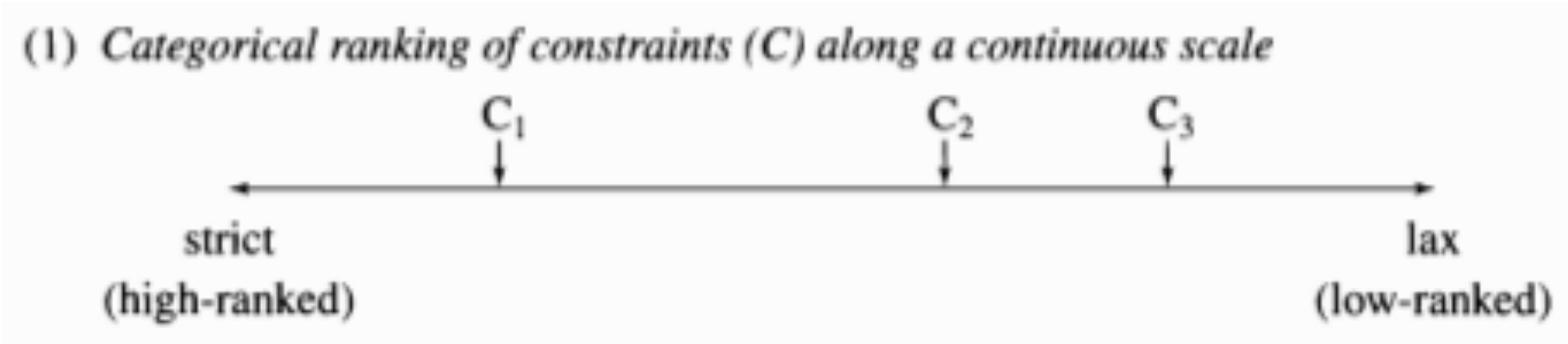
- Id(place) >> *θ >> Id(cont): 50%
- Id(place) >> Id(cont) >> *θ: 50%
- *θ>> Id(place) >> Id(cont): 0%
- *θ>> Id(cont) >> Id(place): 0%
- Id(cont) >> *θ>> Id(place): 0%
- Id(cont) >> Id(place) >> *θ: 0%

You could also imagine a scenario where each of these constraint rankings gets some %. But no one has actually proposed a model like that!

Stochastic OT (Boersma & Hayes, 2001)

- Still within the realm of 'strict'-ranking

- Instead of discrete rankings, constraints are on a **linear** scale of strictness

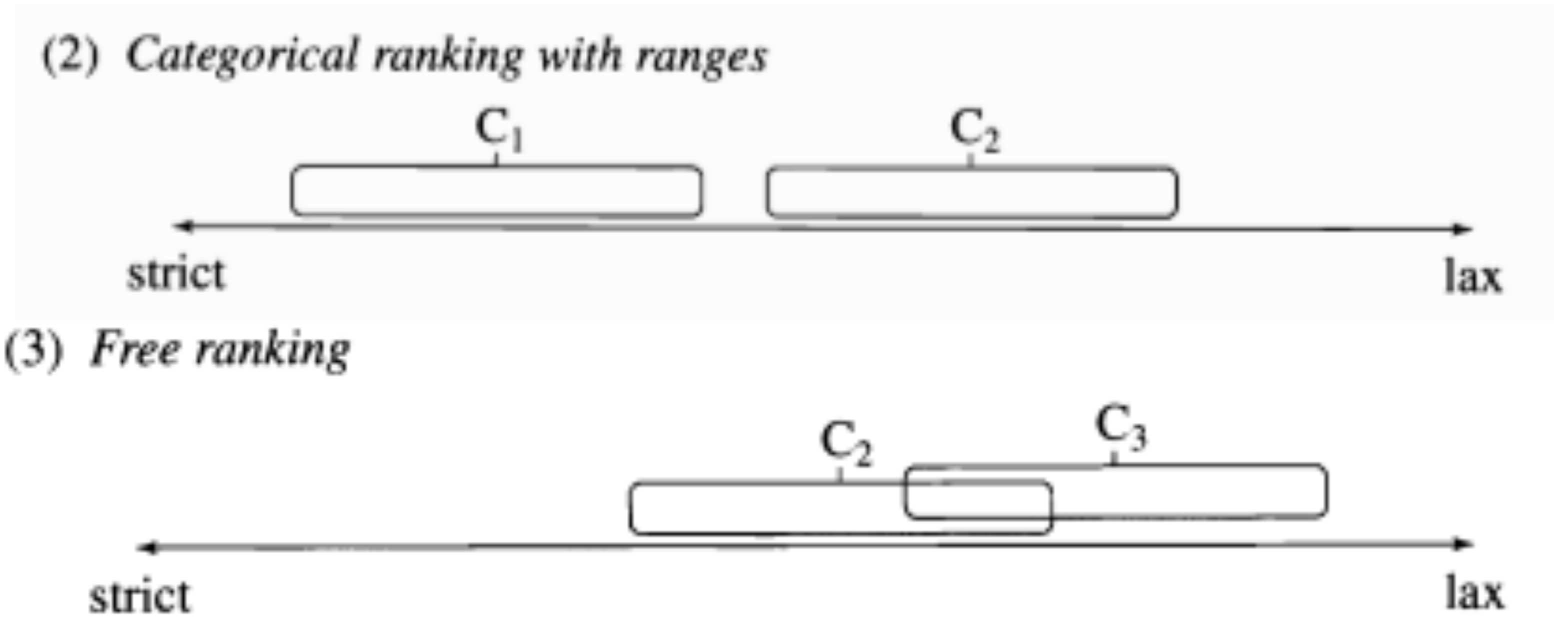


Stochastic OT (Boersma & Hayes, 2001)

- Constraints are associated with a range of values instead of a fixed value
- When a speaker has to produce a form (i.e. at *evaluation time*), the position of constraints are perturbed by a random positive/negative value of noise (within the range of possible values)
- Choose one point from each constraint's range, then use a total ranking according to those points.

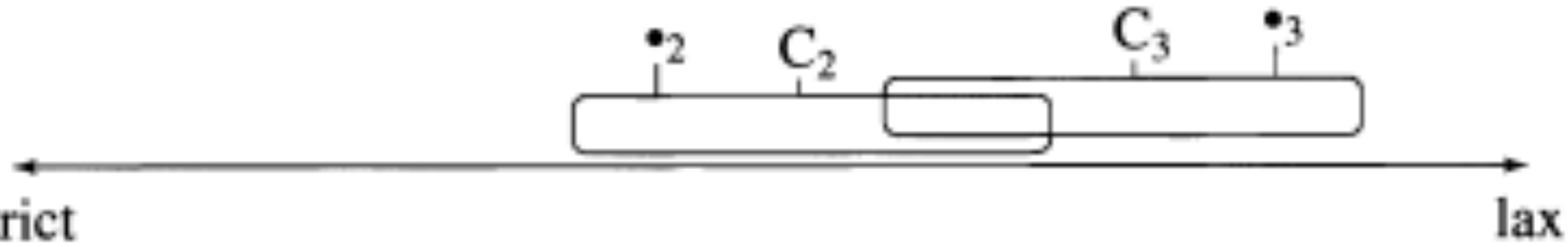
Stochastic OT (Boersma & Hayes, 2001)

- When a speaker has to produce a form (i.e. at *evaluation time*), the position of constraints are perturbed by a random positive/negative value of noise.
- Constraints are associated with a range of values instead of a fixed value

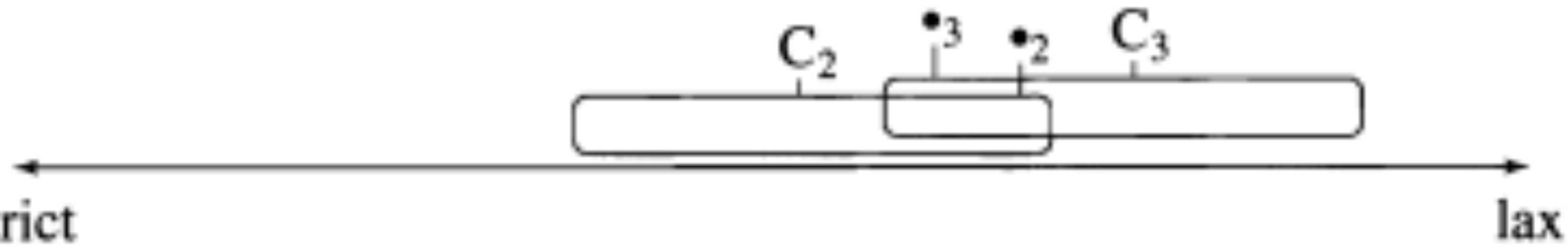


Stochastic OT (Boersma & Hayes, 2001)

(4) a. *Common result:* $C_2 \gg C_3$



b. *Rare result:* $C_3 \gg C_2$



- [demo in Excel]

- (from Kie Zuraw)

Example: Hungarian Vowel Harmony
(Hayes & Londe, 2006)

- Dative suffix [nɛk]~[nɔk]
- Basic pattern: allomorph selection is based on final vowel in stem.

(2)	BB	[ɔblɔk-nɔk]	*ablaknak*	'window-DAT'
	NB	[biːroː-nɔk]	*bírónak*	'judge-DAT'
	FB	[glykoːz-nɔk]	*glükóznak*	'glucose-DAT'

(3)	F	[yʃt-nɛk]	*üstnek*	'cauldron-DA
	BF	[ʃoføːr-nɛk]	*sofőrnek*	'chauffeur-D

Variability with vowel harmony

- Hayes & Londe focus on particular stem-type: cases where a stem-final neutral vowel is preceded by back-vowels
- These can either take the back or front suffixes, some also vacillate between both options.

(7)	BN	[pɔlleːr-nɔk]	*pallérnak*	'foreman-DAT'
	BN	[ɔrzeːn-nɔk, ɔrzeːn-nɛk]	*arzénnak, arzénnek*	'arsenic-DAT'
	BBN	[mutɔgeːn-nɛk]	*mutagénnek*	'mutagen-DAT'

Statistical generalizations

- **Height** effect: the lower the vowel, the more likely the back suffix is selected. E.g. %back is low with [ɛ] but highest with [i]

- **Count** effect: BN stems take back suffixes more than BNN stems (i.e. the further away the trigger for harmony is (B), the less likely you get harmony occurring

Statistical generalizations

tem ype	back	vacillator	front	total stems	backness index
	6251	39	0	6290	0·999
N	603	78	83	764	0·831
Bi	458	17	0	475	0·989
Bi:	52	0	1	53	0·980
Be:	93	18	9	120	0·845
Bε	0	43	73	116	0·104
NN	6	21	44	71	0·206
BNi	1	12	17	30	0·223
BNi:	1	7	0	8	0·358
BNe:	4	2	6	12	0·421
BNε	0	0	21	21	0
	14	23	259	296	0·078
N	0	4	939	933	0·002
	0	0	698	698	0

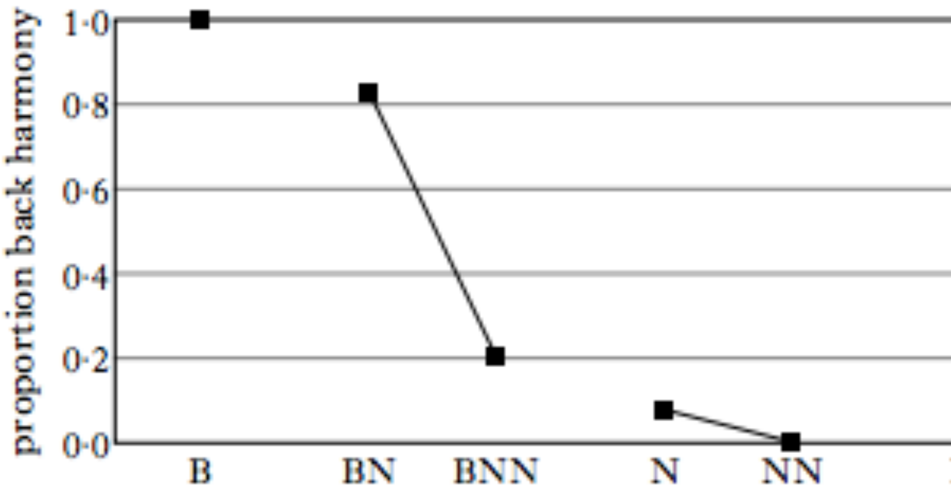


Figure 1
Google data: basic stem types.

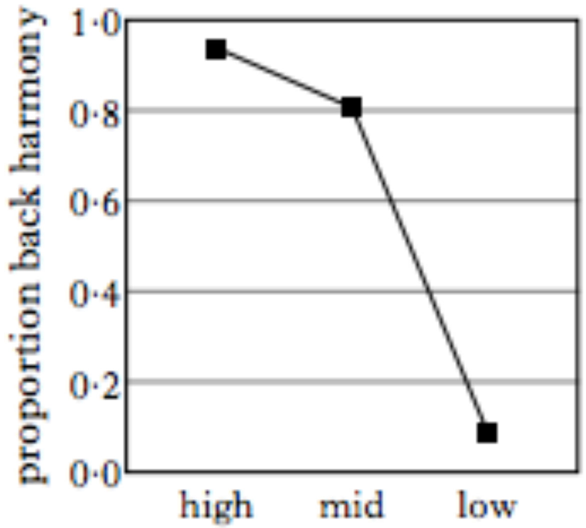


Figure 2
Google data: height effect.

Native speaker confirmation

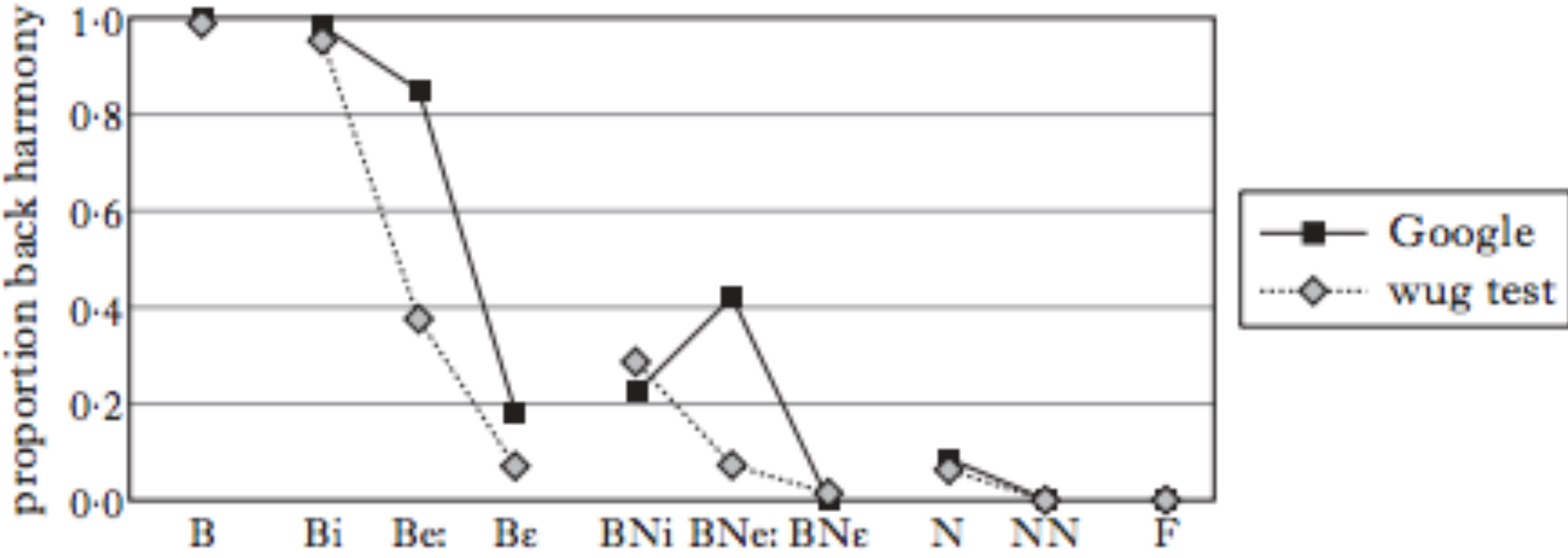


Figure 5

Wug-test data compared with Google data.

Final grammar

(without going through the full analysis)

- Using Stochastic OT: able to generate a grammar that generates outputs in proportions similar to those produced by native speakers!

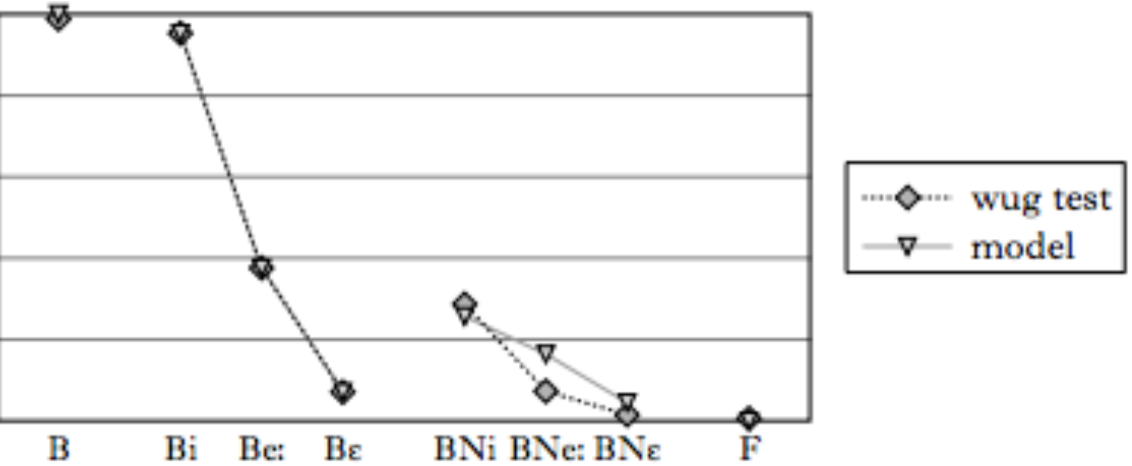
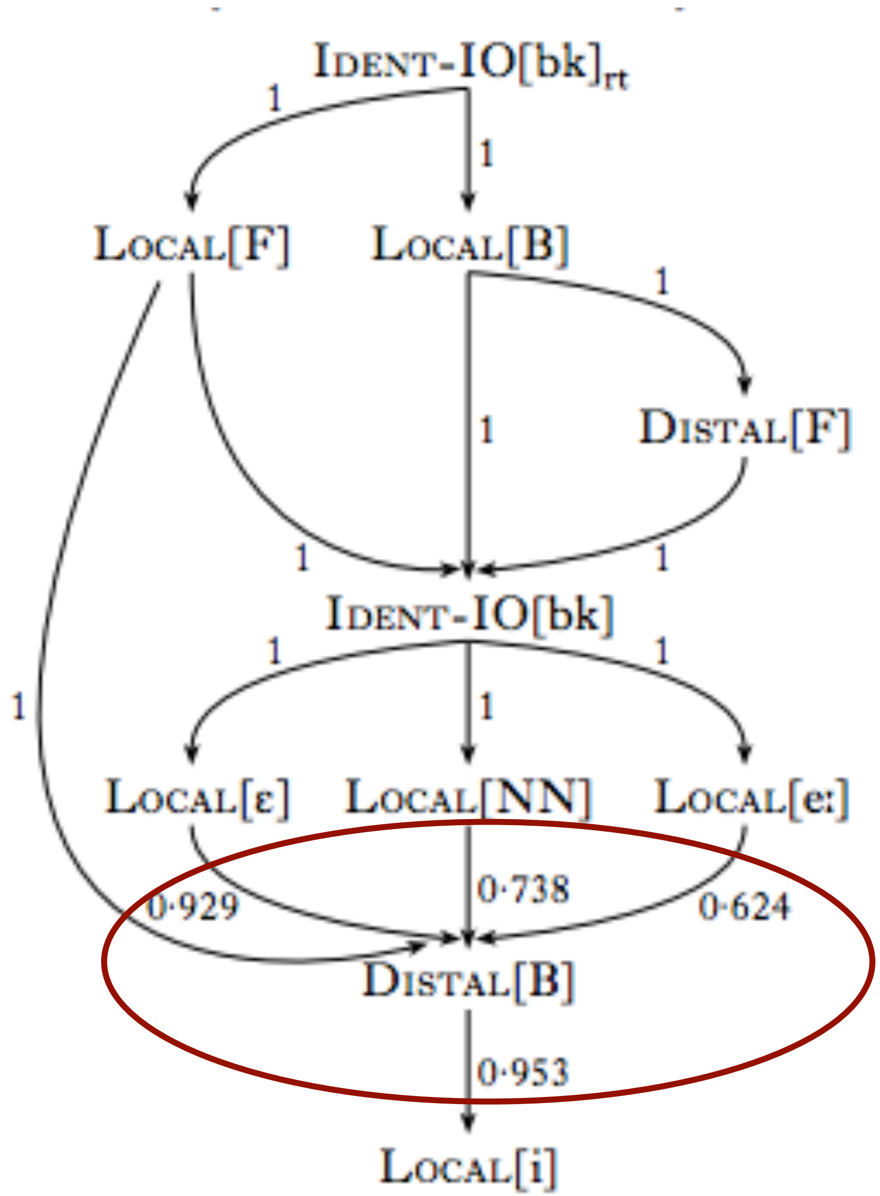


Figure 6
Match-up of model to wug-test data.

(34)



A few things to note

- Good example of triangulation: Hayes & Londe use data from:
  - Google search – lexical search
  - Confirmed with Wug test (test of productivity)
  - Tested model against data from Wug test
- Here, the model more-or-less replicates the behaviour of people!
- There's a learning algorithm for this – Gradual Learning Algorithm (Boersma, 1997) (though doesn't reliably converge – see Pater 2008)

How is variation accounted for in constraint-based grammars?

- Brief look at four different models:
  - Partial-ranking
  - Stochastic OT
  - Noisy Harmonic Grammar
  - MaxEnt Harmonic Grammar

} Use strict-ranking of constraints

Summary

- Phonological variation as something the grammar needs to account for
- Constraint re-ranking can also give us within-language variation
- Two ways:
  - Partial-ranking
  - Probabilistic ranking (Stochastic OT)

- Tomorrow: Weighted (vs. ranked) constraints
  - Noisy Harmonic Grammar
  - MaxEnt Harmonic Grammar

References

Anttila, A. (1997). Deriving variation from grammar. In Hinskens, F., van Hout, R., Wetzels, L. (eds.), *Variation, change, phonological theory.* Amsterdam: John Benjamins, p. 33-68

Boersma, P. (1997). How we learn variation, optionality, and probability. *IFA Proceedings* 21, 43-58.

Boersma, P. & Hayes, B. (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry,* 32, 45-86.

Coetzee, A. & Pater, J. (2011). The place of variation in phonological theory. In *Handbook of Phonological Theory.*

Hayes, B. & Londe, Z. (2006). Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology, 23,* 59-104.

Labov, W. (1972). The reflection of social processes in linguistic structures. *Sociolinguistic Patterns*, 110–121. Philadelphia: University of Pennsylvania Press.

Zuraw, K. (2009). Frequency influences on rule application within and across words. *Proceedings of CLS 43.*