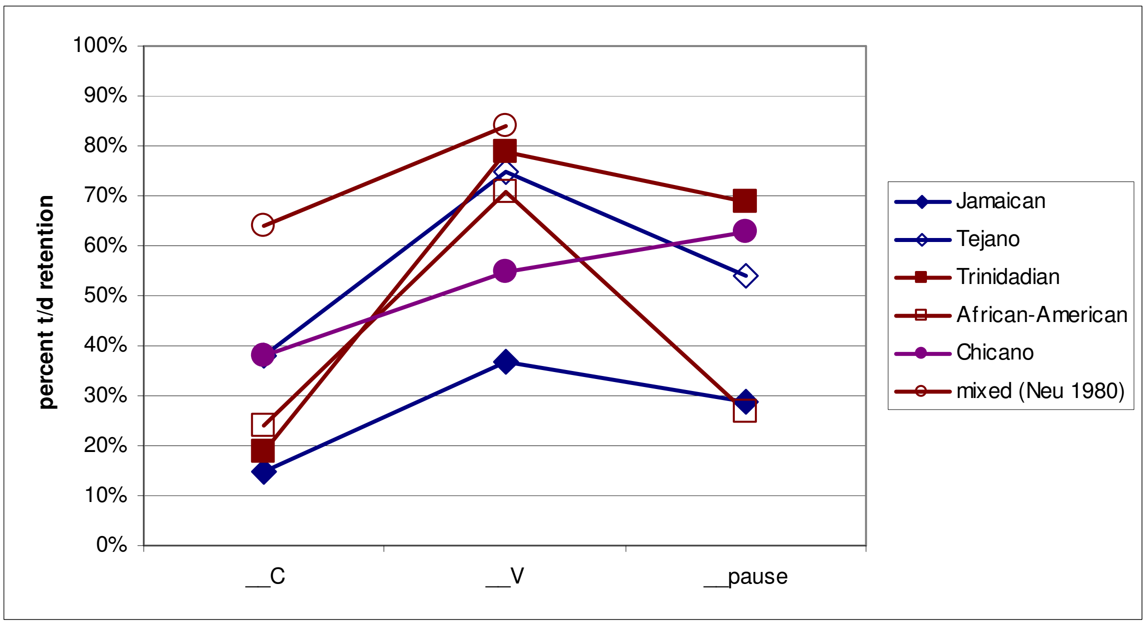
Class 17: Applying what we've seen to a famous case

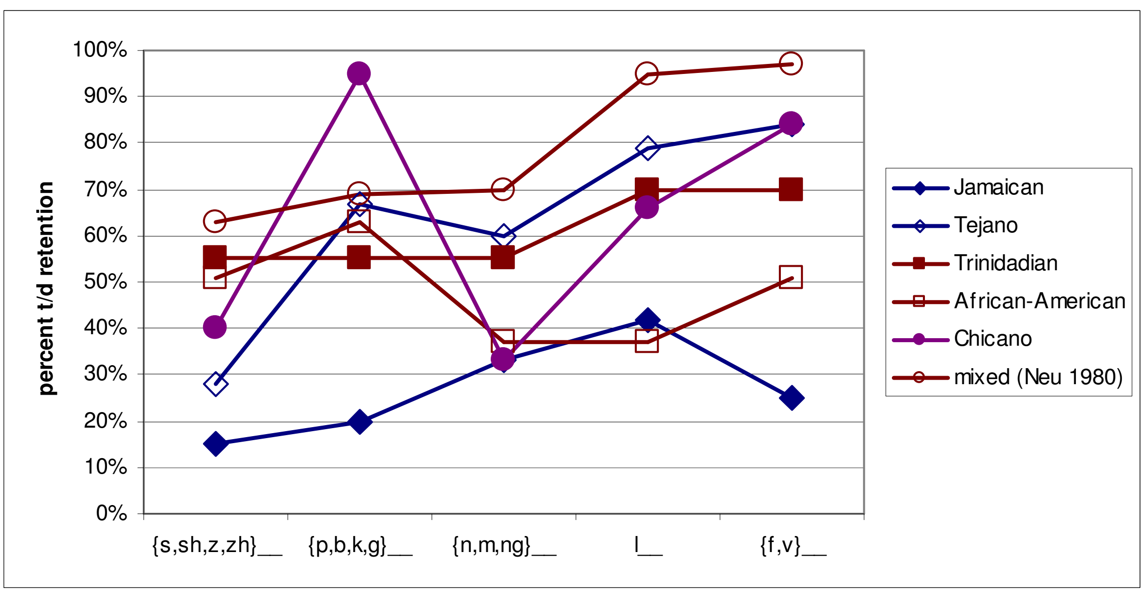
1 English t/d deletion (Labov, Cohen, et al. 1968 is an early paper)

- {t,d} tend to delete when C__#: *jus' fine, hol' these, san' castle*
- Has been studied in dialect after dialect
- Following sound matters: C vs. V vs. pause
- Preceding C matters
- Morphology matters: monomorpheme (*cost*) vs. irregular past (*lost*) vs. regular past (*tossed*)
  - Tagliamonte & Temple 2005 argue that in British English, there's no morphological effect (there is a non-significant trend in the right direction though)

Plotting each phono. effect across dialect, using counts compiled from literature by Coetzee 2004



(see p. 218)



(see p. 290)

2 Guy's exponentiation model

- Guy 1991a, Guy 1991b
- Assumes Lexical Phonology (e.g., Kiparsky 1982)
  - Seems to assume that within each level, phonological rules apply first, then morphology
  - Seems to assume just two levels within the lexical component
- If a variable rule is present at multiple levels, or at a level that words can cycle through multiple times, it has multiple chances to apply:

Lexicon

Level A phonology, including:
- **{t,d} → Ø / C__# (variable)**

↓

Level A Word Formation Rules, including:
- irregular inflection

↓

Level B phonology, including:
- **{t,d} → Ø / C__# (variable)**

↓

Level B WFRs, including:
- regular inflection

↓

Syntax

bracket erasure

↓

Postlexical phonology

Postlexical phonology, including:
- **{t,d} → Ø / C__# (variable)**

- Sample derivations

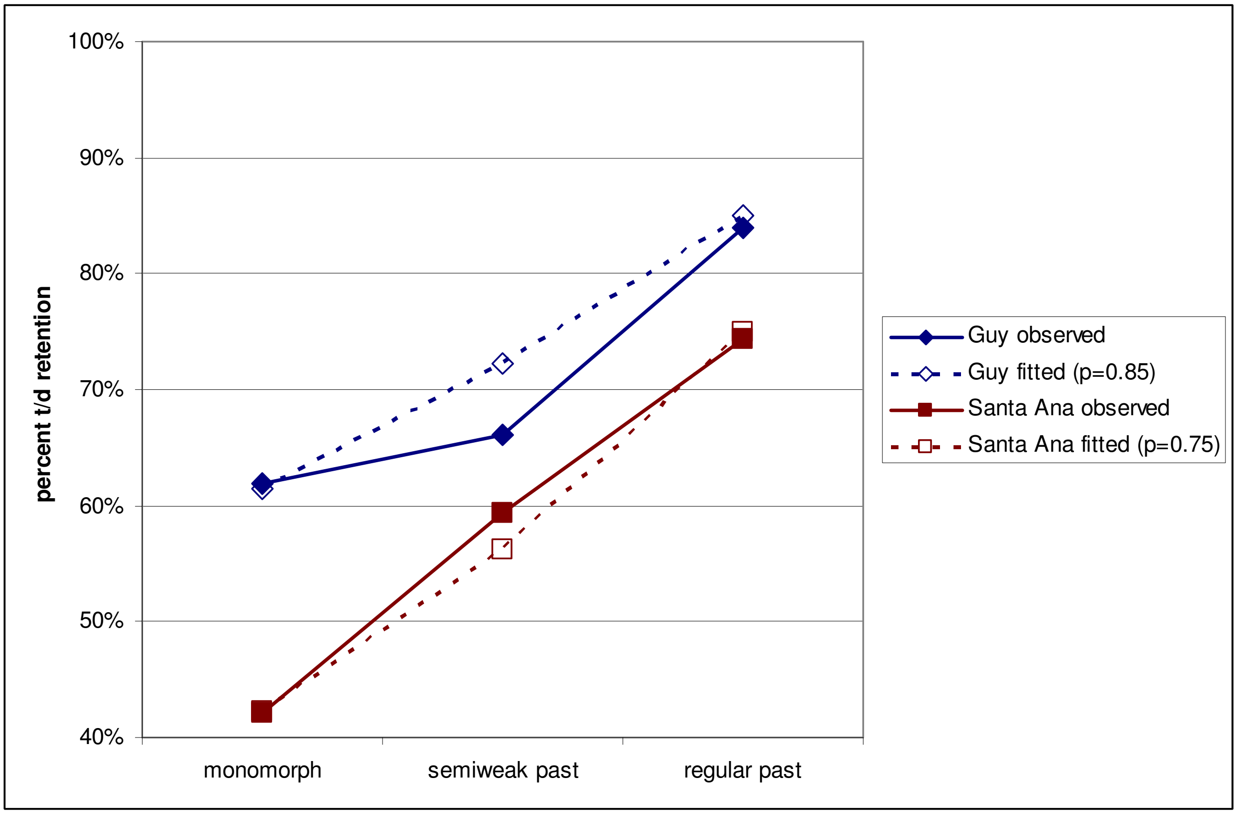
		/mɪst/	/liv/+[past]	/mɪs/+[past]
Level A	phon	*t* could delete	--	--
	morph	--	lɛft	--
Level B	phon	*t* could delete	*t* could delete	--
	morph	--	--	mɪst[1]
Postlex		*t* could delete	*t* could delete	*t* could delete
Result:		3 changes to delete	2 chances to delete	1 chance to delete

- Suppose that the rule always has the same probability of applying: How can we predict the probability that /t,d/ is retained?

[1] It doesn't really matter when voicing assimilates—could be part of the word-formation rule, could be part of the postlexical phonology

3 Data from two dialects (as reported in Guy 1991b)

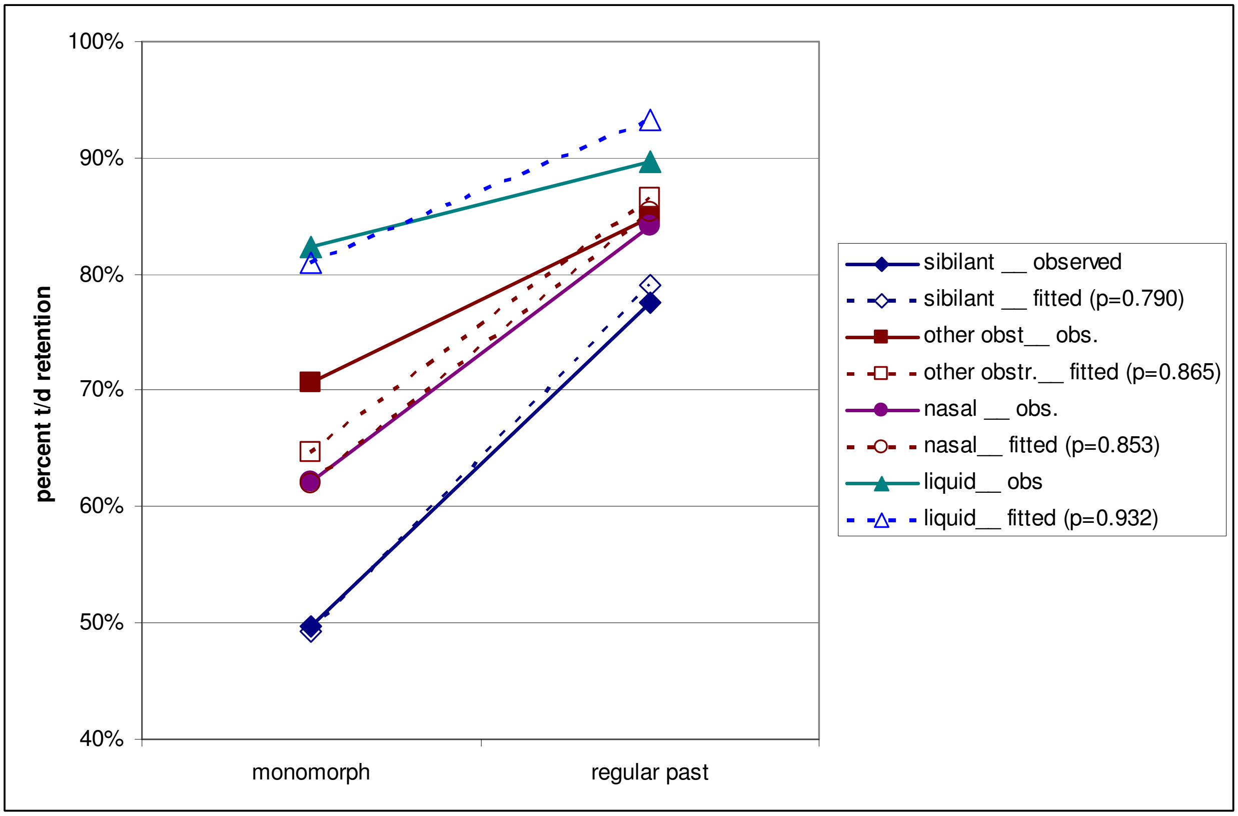
- "p" is the probability of the rule *not* applying (t/d retention)
  - regular past: prob. of retention is just p
  - semiweak past: prob. of retention is p^2
  - monomorpheme: past: prob. of retention is p^3



- ○ Model predicts that difference between dialects is be greatest for monomorphemes—why?

4 Factors internal to the word

- E.g., what the preceding sound is
- Relevant at every level, so essentially each internal context has its own probability of retention:

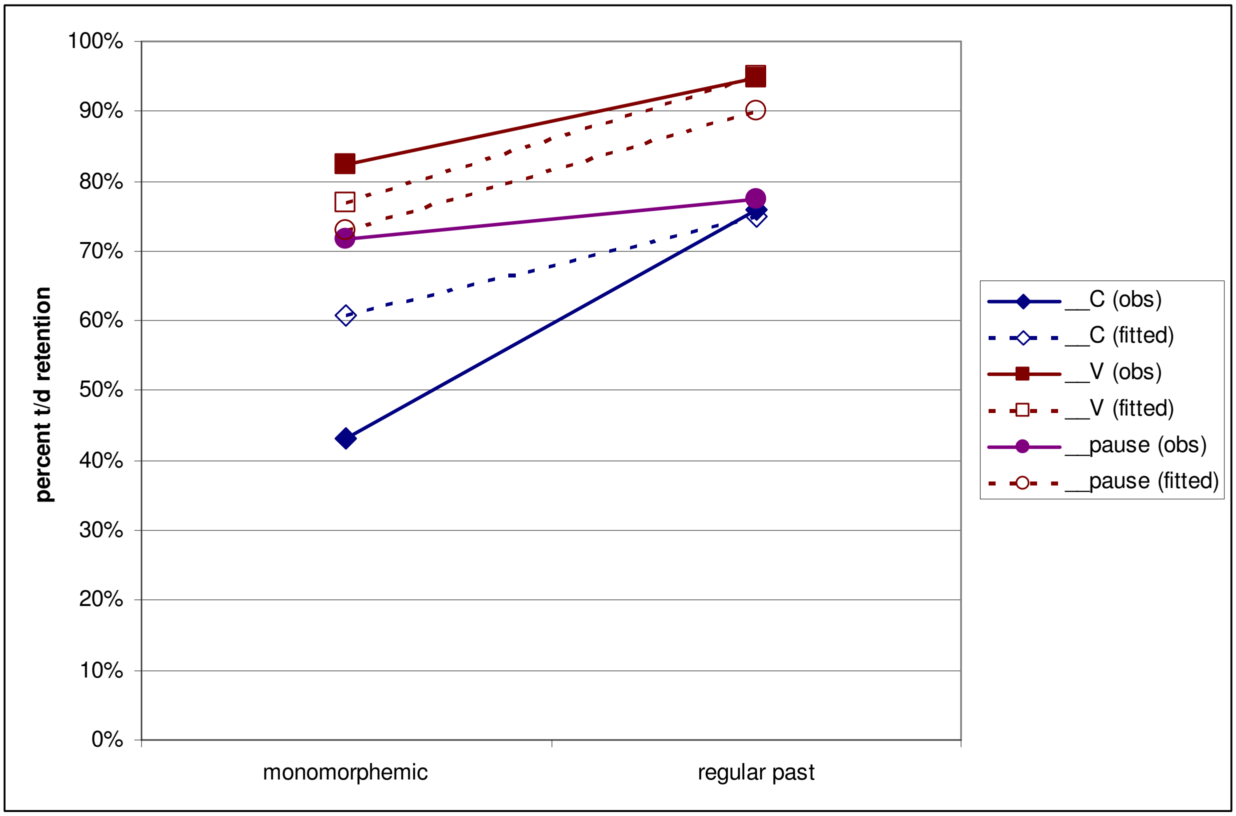


- As with the between-dialect difference, within a dialect the spread in probabilities is greater in the monomorphemes
  - difference between $(0.790)^3$ and $(0.932)^3$ is greater than the difference between 0.790 and 0.932

5 Factors external to the word

- E.g., does next word start with C or V?
- <u>Shouldn't matter till postlexical level</u>
  - monomorphemic rate of retention: $p * p * p_{\text{postlex}}$
  - regular past rate of retention: p_{postlex}
- Differences in external context should be similar regardless of morphology

- Guy gives these data from Wolfram, which (as Guy says) are a bit hard to fit in this model—here's the best I could do:

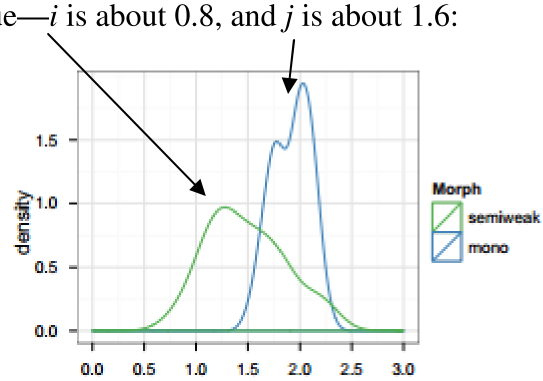


- In the fitted values, differences between external context no bigger in monomorphemes than in regular past
  - Is this also true in the observed values? Hard to say.

6 A refutation: Fruehwald 2012

- Data from Buckeye corpus
- Assume that $p_{retention}$ is the rate observed in regular pasts
- Assume that retention rate in irregular pasts is p_r^i, and in monomorphemes p_r^j,
- Use regression to estimate p_r, i, j
  - Exponential model says i should be 2, j should be 3
  - But this isn't true—i is about 0.8, and j is about 1.6:

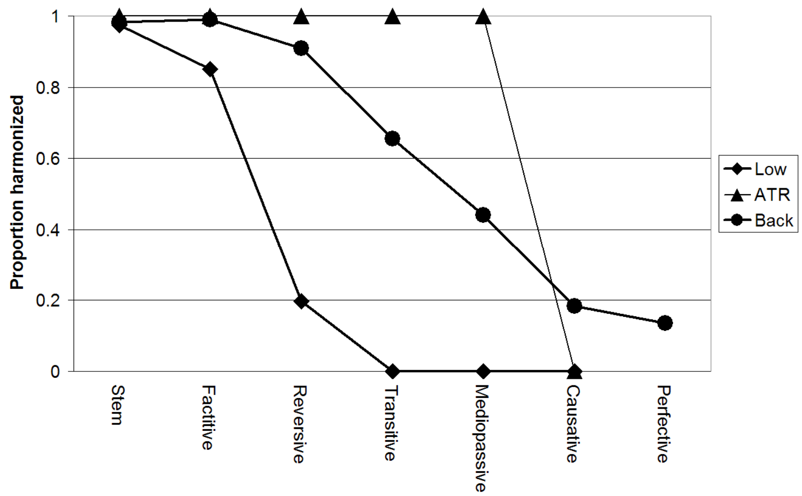
Plot of i and j for each speaker



(p. 80)

7 A refutation with different data: McPherson & Hayes's Tommo So data

- (You've seen this earlier this quarter). The Lexical Phonology/probability exponentiation model predicts the greater and greater differences among the 3 back-harmony rules as we move from outer to inner affixes.
- But instead, the three harmony types converge at both extremes:



(from Class 6 handout)

8 Kiparsky 1994: Partial-ordering model of Guy's data

- Four freely-ranked constraints (I've altered the names to reflect current constraint-naming practices)
  - *COMPLEX
    - Violated if there's a complex onset or complex coda[2]
  - ALIGN = ALIGN(Word,L; Syll,L), ANCHOR-R(Phrase)
    - Violated if word begins mid-syllable
  - ANCHOR = ANCHOR-R(Phrase)
    - Violated if phrase-final segment is deleted
  - MAX-C
- 24 rankings, all equally probable. How these translate into candidate probabilities:
  - __V

lost everything	*COMPLEX	ALIGN	ANCHOR	MAX-C	wins in...
lost.everything	*				8/24 rankings: 33%
los.teverything		*			8/24 rankings: 33%
los'.everything				*	**8/24 rankings: 33%**

  - __C

lost more	*COMPLEX	ALIGN	ANCHOR	MAX-C	wins in...
lost.more	*				12/24 rankings: 50%
los.tmore	*	*			0/24 rankings: 0%
los'.more				*	**12/24 rankings: 50%**

[2] Kiparsky uses a constraint that forbids *any* coda, but assumes that *lost* is a worse violation of this constraint than *los* (showing no violations for *los*, actually). This means we'd need a more-complicated definition of this constraint.

- __pause

lost	*COMPLEX	ALIGN	ANCHOR	MAX-C	wins in...
lost	*				16/24 rankings: 67%
los’			*	*	**8/24 rankings: 33%**

- Let’s discuss how this relates to the data on the previous page

9 Kiparsky on the morphological difference

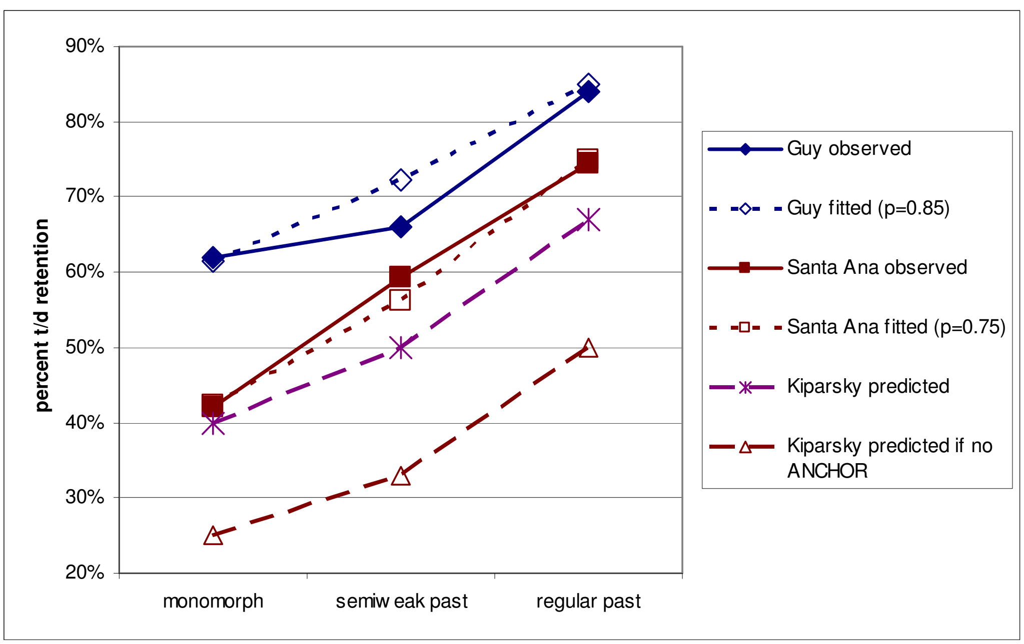
- In the 1994 paper, Kiparsky further splits *COMPLEX into root-level, stem-level, and word-level (in later work, OT and Lexical Phonology are combined)
  - This sort of does lexical phonology in a single tableau
  - We can ignore ALIGN (why?), so there are 5! = 120 rankings to consider.
    - I used OTSoft to check my work: “run” GLA with 0 learning trials (so all constraints have same ranking value), then look at output predictions

cost	*COMPL-root	*COMPL-word	*COMPL-phrase	ANCHOR	MAX-C	wins in...
cost	*	*	*			48/120 rankings: 40%
cos’				*	*	**72/120 rankings: 60%**

lost	*COMPL-root	*COMPL-word	*COMPL-phrase	ANCHOR	MAX-C	wins in...
lost		*	*			60/120 rankings: 50%
los’				*	*	**60/120 rankings: 50%**

tossed	*COMPL-root	*COMPL-word	*COMPL-phrase	ANCHOR	MAX-C	wins in...
tossed			*			80/120 rankings: 67%
toss’				*	*	**40/120 rankings: 33%**

- Let’s discuss how this relates to the data:



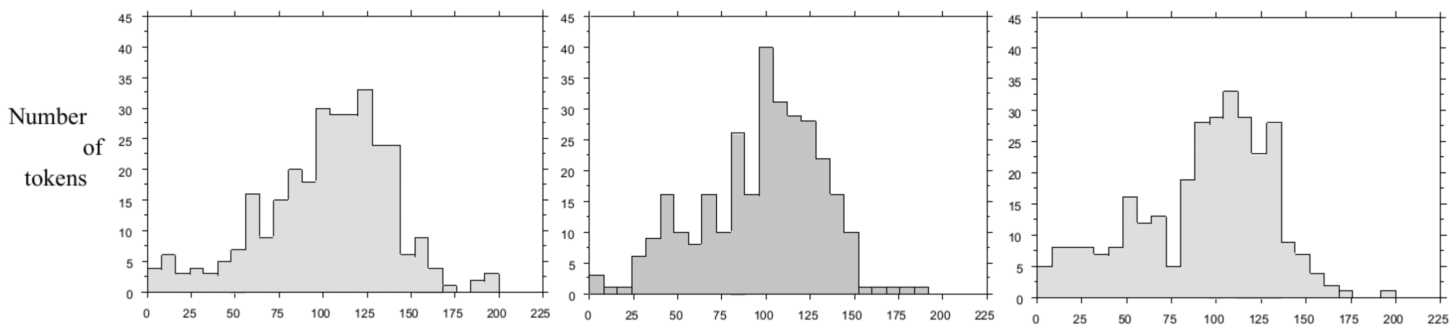
10 Myers 1995: a phonetic look at t/d deletion (see Bermúdez-Otero 2010 for related proposal)

- Points out a difference we might expect
    - *lexical* t/d deletion should be categorical—the phonological representation either does or doesn't have a consonant
    - *postlexical* t/d deletion might be gradient—more or less overlapping gestures, for instance

- Carries out a production study
    - Utterances embedded in a short paragraph
        - *...They tossed many cans into the trash...*
        - *...The toss many cans into the trash...*
        - *...They cost many people their lives...*
        - ...
    - measures duration from end of V to beginning of *many* ("V-m duration")

- Predictions
    - Words like *tossed* should show a diffuse, unimodal V-m-duration distribution, from gradient, postlexical deletion
    - Words like *toss* should show a tight distribution with a small mean: they have no /t,d/, so it's just the duration of the /s/ that's being measured.
    - Words like *cost* should look like a bimodal combination of the two:
        - a group of short-duration items (*t/d* was deleted at lexical level—these should look like *toss*)
        - a more-diffuse group that underwent postlexical deletion (should look like *tossed*)

- Compares mean, variances, normality, skewness of the distributions (p. 13):



regular past (*tossed*) "R" | no /t,d/ (*toss*) "no-t" | monomorphemic (*cost*) M

-
    - M has lower mean than R, as expected—true within each preceding C too (*p, t, k*)
    - No significant difference in variance
    - R is not significantly non-normal; M is
    - M is more skewed than R (significance not directly addressed)

- M doesn't look like R+no-t though—no-t's mean isn't even significantly different from the other two distributions' means! (Variance is significantly smaller though)

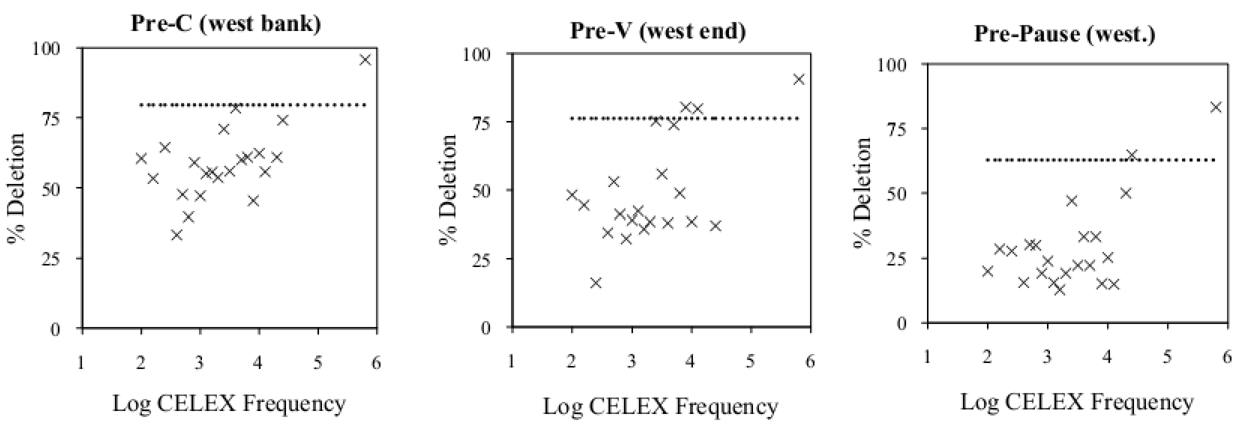
- Myers speculates that deletion at the lexical level, while still categorical, might not be complete
  - Suppose that the lexical-level process is not one of deletion but one of reduction to a category "short-t".

- My take is that we'd really want articulatory data so that we can measure the duration of the *t/d* gesture itself—and its overlap with the rest—instead of having to measure the preceding C also
  - What if the [s] in *toss* is longer than the [s] in *cost*, regardless of whether the /t/ in *cost* gets deleted?
  - o That is a little conceptually weird, though—at what level would [s]-duration be determined?

11 Coetzee 2004, Coetzee 2009, Coetzee & Kawahara 2013: frequency effects

- t/d deletion is more probable if the word is more frequent
  - *just* undergoes deletion more easily than *jest*
- Deletion rates in Buckeye Corpus, horizontal line at mean (over all tokens)

o Why is the mean higher than most of the points?
  - Most (?) literature doesn't use token frequency for this kind of thing, but instead uses token-weighted type frequency.



(p. 4)

- Lots of other processes are like this too

==> A theory that predicts free variation—same deletion rate for all words that have the same phonological properties—is inadequate.

- Not just frequency, but more generally predictability in context can matter.
  - ...at least in cases where one variant is clearly the "reduced" variant
  - o How about variable vowel harmony—which variant should we consider "reduced"?

(Though see Walker 2008: in a different corpus, frequency effects were driven entirely by a few very-frequent outliers: *went, first, different, want, most*)

12 Coetzee & Kawahara's weight scaling

- First, what the (very basic) Noisy HG grammar looks like before weight scaling:

(4) a. Faithful candidate optimal

	w	*nz*	*w*	*nz*	*w*	*nz*	
	5	-0.7	1.5	-0.4	1	0.2	
/lʊst/	DEP (4.3)		*COMPLEX (1.1)		MAX (1.2)		H
☞ lʊst			-1				-1.1
lʊs					-1		-1.2
lʊs.ti	-1						-4.3

"nz" means *noise*, and varies from instance to instance

b. Deletion candidate optimal

	w	*nz*	*w*	*nz*	*w*	*nz*	
	5	-0.7	1.5	0.1	1	-0.1	
/lʊst/	DEP (4.3)		*COMPLEX (1.6)		MAX (0.9)		H
lʊst			-1				-1.6
☞ lʊs					-1		-0.9
lʊs.ti	-1						-4.3

(p. 7)

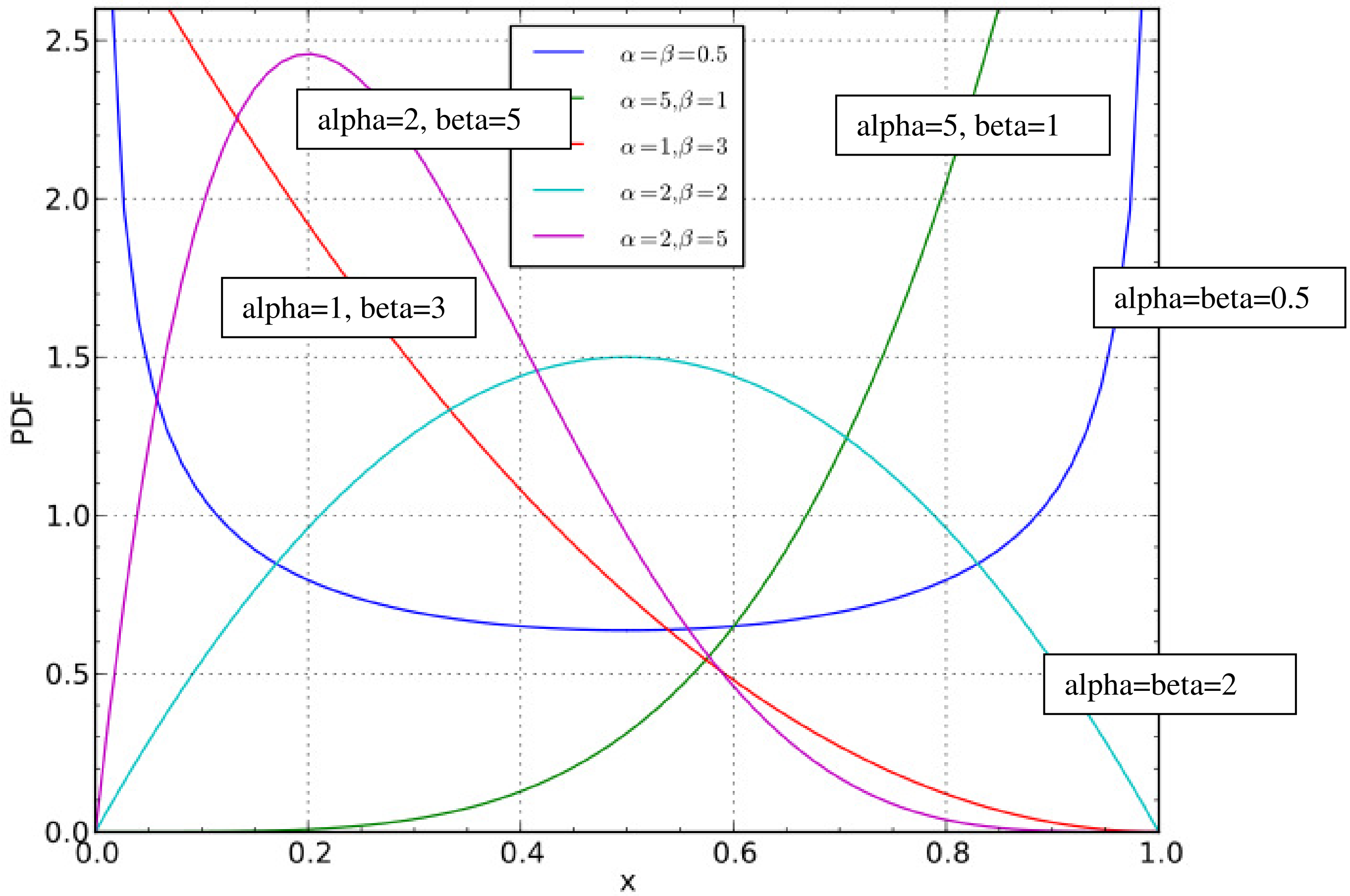
- Coetzee and Kawahara add one more element to Noisy Harmonic Grammar: *weight scaling*.
  - Recall Week 1's discussion of "knobs"
- Each word is associated with a *scaling factor*, which is added to the weight of each faithfulness constraint:

a. Evaluating frequent /lʊst/, with *sf* = -1

	w	*nz*	*sf*	*w*	*nz*	*w*	*nz*	*sf*	
	5	0.7	-1	1.5	0.1	1	0.2	-1	
/lʊst/	DEP (4.7)			*COMPLEX (1.6)		MAX (0.2)			H
lʊst				-1					-1.6
☞ lʊs						-1			-0.2
lʊs.ti	-1								-4.7

(p. 7)

- How to derive scaling factor from frequency: *beta distribution*
  - Has 2 parameters, alpha and beta
  - Difference between alpha and beta determines skewedness
    - When they're equal, probability density function is symmetrical



[3] $\alpha=\beta=0.5$

- Let the *reference point* be the log frequency of the median-frequency word in the set[4]
- Set alpha to that reference number
- Set beta, for each word, to the log frequency of that word
- (There's also a rho, which specifies the distribution's range—C & K fit it to each data set)
- The scaling factor is the mode (peak) of the corresponding distribution

o Suppose the median-frequency word's frequency is 100 (it's a small corpus) and we're interested in a word with frequency of 100,000. What will the beta distribution look like? What's the scaling factor (leaving rho out of it)?

- Why beta distribution? C & K say:
  - Because range is finite, there are upper and lower limits on the scaling factors

[3] http://commons.wikimedia.org/wiki/File:Beta_distribution_pdf.svg

[4] Roughly—see paper for details

13 Coetzee & Kawahara's grammar

- Weights learned by lumping all tokens together:

(14)

*CT	101.16
MAX	98.84
MAX-PRE-V	-1.51
MAX-PRE-PAUSE	0.96

(p. 18)

- Fit to all tokens lumped together is great:

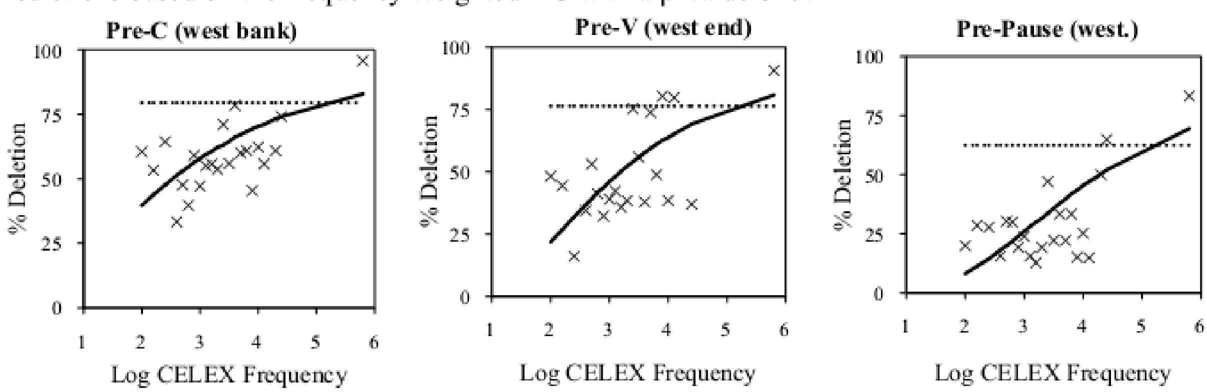
(15)

Context	Observed deletion	Expected deletion
Pre-C	80%	79.6%
Pre-V	76%	76.2%
Pre-Pause	63%	62.8%

(p. 19)

- What happens when scaling factor is included (for best rho):

Figure 4: Observed and predicted *t/d*-deletion rates in Columbus English. The broken line indicates the predictions based on the baseline, unscaled HG. The solid line shows the predictions based on the frequency weighted HG with a ρ-value of 5.



(p. 21)

  - Mean square error improves 79% compared to with no scaling factor

o Discuss: What does this mean for modeling variation? Are cases where any frequency differences are a nuisance rather than a point of interest, and if so what does this mean for those cases?

14 Hay 2003: frequency effects on deletion *within* a word, at stem-suffix boundary

- E.g., *sof'ly* vs. *softly*

(See Raymond, Dautricourt & Hume 2006 for a broad look at word-internal {t,d}-deletion in the Buckeye corpus.)

Uncontroversial assumptions about lexical access:

- Each lexical entry has a <u>resting activation</u>, largely determined by frequency (how many times you've encountered the word in your life)
  - Under the conception of resting activation as activation-before-lexical-access-starts, it's also affected by how recently you've encountered the word.
- During lexical access, various factors (depending on the model) increase or decrease item's activation
  - In production: e.g., activation of conceptual feature "furry" spreads to *cat*
  - In perception: e.g., activation of [k] spreads to *cat*
- The first item to cross some threshold "wins" and gets accessed.

Adds another, slightly controversial assumption:

- *daftly*, *daft* (and possibly *-ly*) can both exist as lexical entries
- If the goal is to produce *daftly*, *daft* and *daftly* compete for access.
  - *daftly* could win: whole-word access
  - *daft* could win, and then be composed with *-ly*: compositional access
- All else being equal, the outcome depends on which is more frequent, *daft* or *daftly*

Proposal

Phonological/phonetic reduction, in this case /t/ deletion, is more likely within an accessed item than across a boundary between two separately accessed items.

- Hay suggests this could actually be a listener-oriented effect
  - if perception includes a "Fast Phonological Preprocessor" (Pierrehumbert 2001) that posits a boundary in phonotactically illegal sequences like *ftl*...
  - ...then retaining the /t/ encourages the listener to attempt decomposed access, and in this case (because the /t/ belongs to the base) makes accessing the base easier too
  - this is good if decomposition is going to be the faster route

15 Hay's experiment

daftly-type words: word < base, word infrequent
softly-type words: word < base
swiftly-type words: word > base
briefly-type words: control; no /t/
(5 quadruples like this)

but *softly* and *swiftly* have similar frequency

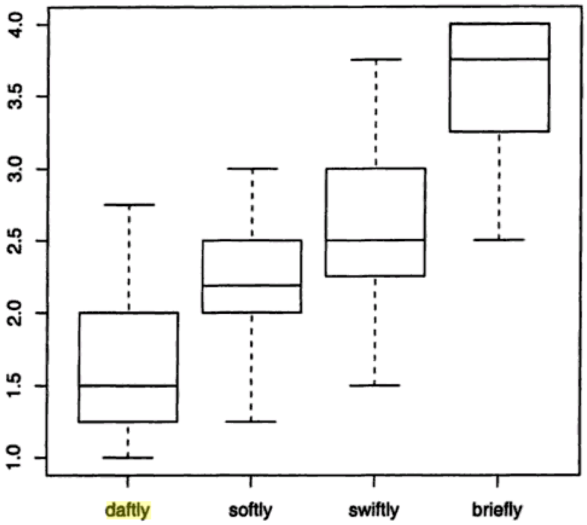
<u>Hypothesis</u>: more /t/-deletion in *swiftly* than in *daftly* and *softly*
<u>Alternate hypothesis</u>: high freq. -> reduction, so more /t/-deletion in *softly* and *swiftly* than *daftly*

- 6 participants read sentences like *Chris dropped by very briefly*
- For each repetition or each quadruple in each subject, transcriber ranked the 4 tokens for [t]-ness (/t/ duration if any, but geminated /f/ ranked above singleton).
  - Dependent variable is the average of these rankings.

Results

- Relative frequency matters: *softly* vs. *swiftly*
- Absolute frequency matters too: *daftly* vs. *softly*
  - Hay speculates that *soft*, being more frequent than *daft*, might itself have more [t]

less [t]



more [t]

Hay, p. 135 (Figure 6.5)

- Just for fun, I tried a regression model. 3 factors (word freq, base freq, which is more frequent), so tried 2^3 possibilities. Best model:

```
Linear mixed model fit by REML
Formula: rank ~ (1 | subject) + log(word_freq + 0.1) which_more_freq

                     Estimate Std. Error t value
(Intercept)           1.77670    0.07971  22.289
log(word_freq + 0.1)  0.06511    0.01813   3.592
which_more_freqword   0.37547    0.12437   3.019
```

  - (once you have a binary which-is-more-frequent factor, adding or substituting base frequency doesn't improve the model)

o Discuss: Let's speculate about how this might work under the Lexical Phonology and/or Noisy HG models we've just seen.

References

Bermúdez-Otero, Ricardo. 2010. Currently available data on English t/d-deletion fail to refute the classical modular feedforward architecture of phonology. *Manchester Phonology Meeting 18*. Manchester.

Coetzee, Andries W. 2004. What it means to be a loser: Non-optimal candidates in Optimality Theory. UMass Amherst Ph. D dissertation.

Coetzee, Andries W. 2009. An integrated grammatical/non-grammatical model of phonological variation. In Young-Se Kang, Jong-Yurl Yoon, Hyunkung Yoo, Sze-Wing Tang, Yong-Soon Kang, Youngjun Jang, Chul Kim, Kyoung-Ae Kim & Hye-Kyung Kang (eds.), *Current Issues in Linguistic Interfaces*, vol. 2, 267–294. Seoul: Hankookmunhwasa.

Coetzee, Andries W. & Shigeto Kawahara. 2013. Frequency biases in phonological variation. *Natural Language & Linguistic Theory* 31(1). 47–89. doi:10.1007/s11049-012-9179-z (27 May, 2013).

Fruehwald, Josef. 2012. Redevelopment of a Morphological Class. *University of Pennsylvania Working Papers in Linguistics* 18(1). http://repository.upenn.edu/pwpl/vol18/iss1/10.

Guy, Gregory R. 1991a. Explanation in Variable Phonology: An Exponential Model of Morphological Constraints. *Language Variation and Change* 3(01). 1–22. doi:10.1017/S0954394500000429.

Guy, Gregory R. 1991b. Contextual conditioning in variable lexical phonology. *Language Variation and Change* 3(2). 223–239.

Hay, Jennifer. 2003. *Causes and consequences of word structure*. Routledge.

Kiparsky, Paul. 1982. Lexical Morphology and Phonology. *Linguistics in the Morning Calm*. Hanshin Publishing Co.

Kiparsky, Paul. 1994. An OT perspective on phonological variation. Stanford University, Stanford, CA.

Labov, William, P Cohen, C Robins & J Lewis. 1968. A Study of the Non-Standard English of Negro and Puerto Rican Speakers in New York City. Volume I: Phonological and Grammatical Analysis. http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED028423 (28 May, 2013).

Myers, James. 1995. The categorical and gradient phonology of variable t-deletion in English. Paper presented at the International Workshop on Language Variation and Linguistic Theory, University of Nijmegen, ms.

Pierrehumbert, Janet. 2001. Why Phonological Constraints Are So Coarse-Grained. *Language and Cognitive Processes* 16(5). 691–98. (2 January, 2011).

Raymond, William D., Robin Dautricourt & Elizabeth Hume. 2006. Word-internal /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change* 18(01). 55–97. doi:10.1017/S0954394506060042.

Tagliamonte, Sali & Rosalind Temple. 2005. New perspectives on an ol' variable: (t,d) in British English. *Language Variation and Change* 17(03). 281–302. doi:10.1017/S0954394505050118.

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