Linguistics 251 Spring 2013
Variation in Phonology Hayes/Zuraw

Class 10, 5/2/13: More on Ratings vs. Probability

1. Assignments etc.

- Hand back previous exercise.
- New exercise on Logistic Regression, due Thurs. 5/9. Posted on web site.
  - Note: this will be the basis as well of our last exercise, on model comparison.
- Read: Lofstedt (2010) Phonetic Effects in Swedish Phonology: Allomorphy and Paradigms, UCLA dissertation. Read Chapter 4, "Vowel-vowel correspondence and *MAP". This extract is posted on line on the course website. Read for Tues. 5/7.

EXPLAINING THE CLASS EXERCISE

2. Background: the problem of "differential phonotactics"

- The general goal is to devise a grammar that distinguishes between two populations of words.
- Why would we ever want to do this? Several examples.

3. One case of differential phonotactics: product-oriented generalizations (Bybee)

- References for product-oriented generalizations:
  - Bybee, J. (2001). *Phonology and language use*. Cambridge University Press.
  - Bybee, J., & Moder, C. L. (1983). Morphological classes as natural categories. *Language*, *59*, 251–270.
  - Bybee, J., & Slobin, D. (1982). Rules and schemas in the development and use of the English past tense. *Language*, *58*, 265–289.
- What makes a word sound like a past tense?
  - Positive traits:
    - — ending in [ɔt] (*thought, caught, wrought, bought, brought, besought, sought, taught*)
    - — containing [ʌ] (*struck, snuck, dug, stuck, slunk, shrunk, stunk, flung, clung, slung, spun, wrung, sprung, strung, stung, won, swung, hung*)
    - — containing [oʊ] (*rode, strode, shone, smote, wrote, underwrote, dove, drove, strove, rose, arose, broke, woke, awoke, bore, forbore, tore, wore, swore, forswore, spoke, stole, wove, froze, chose, rode strode spoke*)
    - — ending in [−voice] + *t* or *d* (this makes you sound like a regular)
  - Negative traits:
    - — ending in a voiceless fricative [f, θ, s, ʃ] (an "island of reliability" for regularity in English past tenses; Albright and Hayes)

- This is a problem of differential phonotactics — how are past tenses different from words that are not past tenses?
- Differential phonotactics plausibly could be an important part of a past tense model, but it cannot be all of one.
    - [vɪd] is a conceivable past tense but would have to be irregular.
    - [krəmaɪd] is a conceivable past tense but would have to be regular.

4. Differential phonotactics for vocabulary strata

- For the phonology of many languages, it is useful to separate the vocabulary into strata.
- Japanese: Yamato, Sino-Japanese, Mimetic, Foreign (Ito/Mester)
    - Only Yamato undergoes rendaku (ori-kami → origami)
- English: Latinate, Native
    - Compare wug words: vennipation, vennistration, veniwation, venichation
- See Moreton and Amano (1999) for a nice psycholinguistic experiment on the psychological reality of strata in Japanese[1]
- Vocabulary strata are partly morphological, but partly phonotactic.

5. Differential phonotactics as a way of finding rule environments

- e.g., phonotactics of Hungarian stems that take [−nak], that take [−nɛk] would tell you the environment for vowel harmony
- This is the strategy pursued by Becker and Gouskova, described by Kie last time.
- This explains the opacity in Arto Anttila's famous Finnish example (1995): choice of genitive plural suffix depends on vowel height, but this is the vowel height of the *base* form, before coalescence processes alter it on the surface.
    - I am curious how widely this occurs — easily learnable opacity!

6. Differential phonotactics is a natural problem to handle in logistic regression

- There are two choices in parallel (the two systems of phonotactics), so we can do simple and easy binary logistic regression.
- All the benefits of maxent OT accrue.
- It's just fine if the weights go positive or negative, since we are setting up constraints in both directions in any event.

7. Differential phonotactics usually involves in lexical frequency

- Searching for the environment of an exceptionless phonological pattern (e.g. vowel harmony for most Turkish suffixes) doesn't need frequency, but exceptionful patterns do.

---

[1] http://www.unc.edu/~moreton/Papers/Eurospeech1999.pdf

8. Bruce's curiosity-problem: differential phonotactics for Dr. Seuss's coinages

- "Dr. Seuss" was Theodore Seuss Geisel (1904-1991), a noted American author of children's books.
- His books are mostly written in anapestic tetrameter, and include a great number of coined words (often naming imaginary beasts, people, or places).

And SPAZZ is a letter I use to spell Spazzim
A beast who belongs the Nazzim of Bazzim.
Handy for traveling. That's why he has 'im.

— From *On Beyond Zebra*

From a country called Frumm comes this drum-tummied Snumm
Who can drum any tune that you might care to hum.
(Doesn't hurt him a bit, cause his drum-tummy's numb.)

— From *If I Ran the Circus*

- Part of what distinguishes Seuss's coinages is simply phonotactic marginality; e.g. in *Snumm* [Socr.: what is it?] or more dramatically in *Nuh* [ˈnʌː].
- But there are also characteristic sequences (author-specific phonesthemes?) that are sharply overrepresented in Seuss's coinages.

9. The role of frequency

- This is a problem that can be treated, in part, with frequency; we suppose that there are specifically Seussian phonesthemes that will be identifiable by having much higher frequency than in ordinary English.
- … and thus that we can use logistic regression as a tool for more confidently identifying the author-specific phonesthemes.

10. Data corpus

- From a pile of Seuss books left over at home from my son's childhood, I gathered 179 nonce words and transcribed them using Carnegie-Mellon dictionary transcription:

Obsk	AA1 B S K	Gitz	G IH1 T S	Nerd	N ER1 D	Walloo	W AO2 L UW1
Um	AH1 M	Gluppity-Glupp	G L AH2 P AH0 T IY0 G L AH1 P	Nerkle	N ER1 K AH0 L	Winkibus	W IH1 NG K AH0 B AH0 S
Umbus	AH1 M B AH0 S	glurk	G L ER1 K	Nipswich	N IH1 P S W IH2 CH	Winna-Bango	W IH2 N AH0 B AE1 NG G OW0
Offt	AO1 F T	Glikk	G L IH1 K	noozer	N UW1 Z ER0	Yop	Y AA1 P
Olf	AO1 L F	Glikker	G L IH1 K ER0	o'Grunth	OW2 G R AH1 N TH	Yupster	Y AH1 P S T ER0
Balber	B AA1 L B ER0	gleap	G L IY1 P	Palooski	P AH0 L UW1 S K IY0	Yuzz	Y AH1 Z
Bopps	B AA1 P S	Gractus	G R AE1 K T AH0 S	Pelf	P EH1 L F	Yuzz-a-ma-Tuzz	Y AH1 Z AH0 M AH0 T AH2

							Z
Bar-ba-loot	B AA2 R B AH0 L UW1 T	gruvvulous	G R AH1 V Y AH0 L AH0 S	Preep	P R IY1 P	Yekk	Y EH1 K
Bazzim	B AE1 Z AH0 M	Grickle-grass	G R IH1 K AH0 L	Proo	P R UW1	Yekko	Y EH1 K OW0
Brigger-ba-Root	B AH0 R UW1 T	Grinch	G R IH1 N CH	Redd-Zoff	R EH1 D Z AO2 F	Yerka	Y ER1 K AH0
Bustard	B AH1 S T ER0 D	Gootch	G UW1 CH	rippulous	R IH1 P Y AH0 L AH0 S	Yertle	Y ER1 T AH0 L
Ben-Deezing	B EH2 N D IY1 Z IH0 NG	Gwark	G W AA1 R K	Sala-ma-goox	S AE2 L AH0 M AH0 G UH1 K S	Ying	Y IH1 NG
Biffer-Baum	B IH1 F ER0	Huffle	HH AH1 F AH0 L	Sala-ma-Sond	S AE2 L AH0 M AH0 S AA1 N D	Yink	Y IH1 NG K
Biggel-Ball	B IH1 G AH0 L	Humpf	HH AH1 M P F	Skeegle-mobile	S K IY1 G AH0 L	Zomba-ma-tant	Z AA2 M B AH0 M AH0 T AE1 N T
Bingle-bug	B IH1 NG G AH0 L	Hiffer	HH IH1 F ER0	Squitsch	S K W IH1 CH	Zans	Z AE1 N Z
Bip	B IH1 P	Hinkle-Horn	HH IH1 NG K AH0 L	smogulous	S M AO1 G Y AH0 L AH0 S	zang	Z AE1 NG
Beers	B IH1 R Z	Itch-a-pod	IH1 CH AH0 P AA2 D	snop	S N AA1 P	Zatz	Z AE1 T S
Beezlenut	B IY1 Z AH0 L	Ish	IH1 SH	Snarp	S N AA1 R P	Zatz-it	Z AE1 T S IH0 T
bloop	B L UW1 P	It-Kutch	IH1 T K AH2 CH	Snumm	S N AH1 M	Zuff	Z AH1 F
bloozer	B L UW1 Z ER0	Jawks	JH AO1 K S	Snuvv	S N AH1 V	Zuk	Z AH1 K
Chugg	CH AH1 G	Jeers	JH IH1 R Z	Sneth	S N EH1 TH	Zumm	Z AH1 M
Dungus	D AH1 NG G AH0 S	Jorn	JH AO1 R N	Snee	S N IY1	Zummzian	Z AH1 M Z IY0 AH0 N
Dutter	D AH1 T ER0	Jounce	JH AW1 N S	Sneedle	S N IY1 D AH0 L	Zorn	Z AO1 R N
Dawf	D AO1 F	Jedd	JH EH1 D	Sneeden	S N IY1 D AH0 N	Zed	Z EH1 D
Dofft	D AO1 F T	Jill-ikka-Jast	JH IH2 L AH0 K AH0 JH AE1 S T	Sneelock	S N IY1 L AA2 K	Ziff	Z IH1 F
Dake	D EY1 K	Joat	JH OW1 T	Snookers	S N UH1 K ER0	Ziffer-Zoof	Z IH1 F ER0 Z UW2 F
Didd	D IH1 D	Katta-ma-side	K AE2 T AH0 M AH0 S AY1 D	Spazz	S P AE1 Z	Zinn-a-Zu	Z IH1 N AH0 Z UW2
Joggoon	JH AA2 G UW1 N	Katroo	K AH0 T R UW1	Spritz	S P R IH1 T S	Zind	Z IH1 N D
Fotichee	F AA1 T AH0 CH IY0	Keck	K EH1 K	Strookoo	S T R UW1 K UW2	Zinzibar-Zanzibar	Z IH1 N Z AH0 B AA2 R
Fa-Zoal	F AH0 Z OW1 L	clop	K L AA1 P	Soobrian	S UW1 B R IY0 AH0 N	Zeep	Z IY1 P
Fuddle	F AH1 D AH0 L	Klopfer	K L AA1 P F ER0	Swomee-swans	S W OW1 M IY0		
Fibbel	F IH1 B AH0 L	Krox	K R AA1 K S	Schloppity-Schlopp	SH L AA2 P AH0 T IY0 SH L AA1 P		
Fizza-ma-Wizza-ma-Dill	F IH2 Z AH0 M AH0 W IH2 Z AH0 M AH0 D IH1 L	cruffulous	K R AH1 F Y AH0 L AH0 S	Tobsk	T AA1 B S K		
Flupp	F L AH0 P	Quan	K W AA1 N	Tudd	T AH1 D		
Flunn	F L AH1 N	Kwong	K W AO1 NG	Tidder	T IH1 D ER0		
Flunnel	F L AH1 N AH0 L	Kwigger	K W IH1 G ER0	Tinkibus	T IH1 NG K AH0 B AH0 S		
Floob	F L UW1 B	Kweet	K W IY1 T	Truffula	T R AH1 F Y AH0 L AH0		
floop	F L UW1 P	Lass-a-lack	L AE1 S AH0 L AE2 K	Thidwick	TH IH1 D W IH2 K		
Frumm	F R AH1 M	Lorax	L AO1 R AE2 K S	Thnad	TH N AE1 D		
Frink	F R IH1 NG K	Lerkim	L ER1 K IH0 M	Thnadner	TH N AE1 D N ER0		
Far Foodle	F UW1 D AH0 L	Malber	M AA1 L B ER0	Thneed	TH N IY1 D		
Foon	F UW1 N	Motta-fa-Potta-fa-	M AA2 T AH0 F AH0 P AA2 T AH0 F AH0 P EH1 L	Thwerll	TH W ER1 L		

		Pell			
Foona-Lagoona	F UW1 N AH0	Mupp	M AH1 P	Va-Vode	V AH0 V OW1 D
Gox	G AA1 K S	Natch	N AE1 CH	Van Vleck	V L EH1 K
Gack	G AE1 K	Nadd	N AE1 D	Vroom	V R UW1 M
Gump	G AH1 M P	Nazzim	N AE1 Z AH0 M	Voom	V UW1 M
Gekko	G EH1 K OW0	Nuh	N AH1	Wum	W AH1 M
G-r-r-zopp	G ER0 Z AA1 P	Nubb	N AH1 B	Wumbus	W AH1 M B AH0 S
G-r-r-zapp	G ER0 Z AE1 P	Nutch	N AH1 CH	Wump	W AH1 M P
G-r-r-zibb	G ER0 Z IH1 B	Nungus	N AH1 NG G AH0 S	Wog	W AO1 G

11. Camparison population of non-Seuss words

- I used my groomed version of the CMU dictionary (used e.g. in Daland et al. (2009), Hayes and White (2013)), omitting suffixed and compound forms.

12. Constraints

- All target constraints are of the form, "Be a Seuss word if you have property X".
- I can't think of any salient properties of non-Seuss words.
- I did a more than this but I'm giving just three constraints for pedagogical purposes.
- Part of SeussViolationsFile.txt, selected to show violations:

Word	Transcription	IsSeuss	InitialZ	InitialTHConsonant	TH
Zomba-ma-tant	[Z AA2 M B AH0 M AH0 T AE1 N T]	1	1	0	0
Zans	[Z AE1 N Z]	1	1	0	0
Zind	[Z IH1 N D]	1	1	0	0
Zinzibar-Zanzibar	[Z IH1 N Z AH0 B AA2 R]	1	1	0	0
Zeep	[Z IY1 P]	1	1	0	0
Thnad	[TH N AE1 D]	1	0	1	1
Thnadner	[TH N AE1 D N ER0]	1	0	1	1
Thneed	[TH N IY1 D]	1	0	1	1
Thwerll	[TH W ER1 L]	1	0	1	1
Thidwick	[TH IH1 D W IH2 K]	1	0	0	1
Obsk	[AA1 B S K]	1	0	0	0
Um	[AH1 M]	1	0	0	0
Umbus	[AH1 M B AH0 S]	1	0	0	0
Offt	[AO1 F T]	1	0	0	0

- However, the bulk of the file consists of thousands of real words, with IsSeuss = 0 and constraint violations duly assessed.

13. Assessing constraint violations in Excel

- Use the string functions.

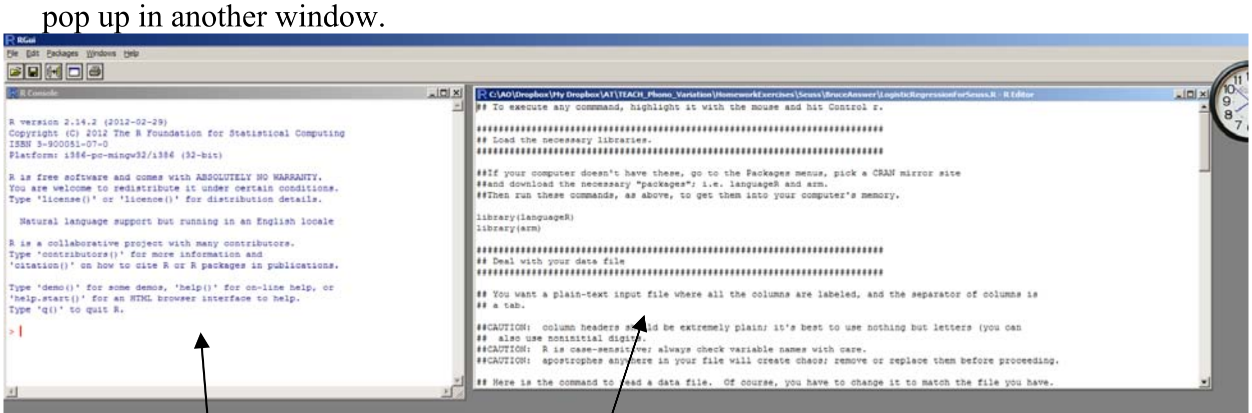
- Locating words with initial [z]:
  - =if (left(A1, 1) = “Z”, 1, 0)
- Others: right(cell, length), mid(cell, start, maxlength[2])
- A weird one: =if(iserror(find(“SearchString”, cell)), 0, 1)
  - This produces 1 if the cell contains the searchstring anywhere, else 0.
- Paste the formula all the way up and down the column to get what you want.
- Disjunction (e.g. θr, θl, θw): make multiple columns and add them up.

14. A way to locate more complicated constraint violations

- My little Phonology Search utility (http://www.linguistics.ucla.edu/people/hayes/EnglishPhonologySearch/) lets you search on natural classes, using syllables; likewise stress patterns/vowel patterns. Use the input file SeussPlusEnglish.txt, part of the package.
- Use Excel to paste the results, gathered in “ViolationsFile.txt,” into SeussViolationsFile.txt.

15. Starting up with R

- Download R from http://www.r-project.org/ (all platforms)
- Download from the course web site the zipped bundle of files. It will unzip as a working folder with all relevant files.
- Start up R.
- On the **File** menu, select **Change dir** and navigate to the relevant folder.
- On the **File** menu, select **Open script** and choose LogisticRegressionForSeuss.R. It will pop up in another window.



- Results are here Script is here
- To run any line of the script, highlight it with the mouse and hit Control R.
- The script has lots of comment lines and tells you what to do.

---

[2] You have to put in a maximum length; I use 100, which catches all cases and does no harm.

16. The R Script

```
##########################################################################
## Load the necessary libraries.
##########################################################################
```

##If your computer doesn't have these libraries, go to the **Packages** menu, pick a CRAN mirror site, then (again from **Packages** menu) pick **Install Packages**, find languageR and arm.
##Once the packages are downloaded, run these commands to get them into your computer's memory.

```
library(languageR)
library(arm)
```

17. Deal with your data file

You want a plain-text input file where all the columns are labeled, and the separator of columns is a tab.

##CAUTION: column headers should be extremely plain; it's best to use nothing but letters (you can also use noninitial digits.
##CAUTION: R is case-sensitive; always check variable names with care.
##CAUTION: apostrophes anywhere in your file will create chaos; remove or replace them before proceeding.

```
## Here is the command to read a data file.
## sep="t" is needed so that it will assume that tab is the column separator.

MyData=read.table("SeussViolationsFile.txt", header=T, sep="\t")

   ##You can look at the column names with this command:
      colnames(MyData)
```

18. Logistic regression

```
## For linguistics, the best r function for logistic regression is probably
bayesglm().
## This is because there are often exceptionless principles--
##   you don't want the weights to go sky high without good justification.
##   bayesglm() employs a prior to enforce this principle
## The reference source for bayesglm() is
http://www.stat.columbia.edu/~gelman/research/unpublished/priors7.pdf.
## If you want, you can leave the word "bayes" in this command and get
classical glm instead.

MyModel = bayesglm(IsSeuss ~ +
InitialZ +
TH +
InitialTHConsonant, data = MyData, family = "binomial")

## This command merely reports the weights that were found:
MyModel

## This one is nicer, because it also gives you a significance test for each
weight:
summary(MyModel)
```

```
## Print out the model's predictions.
## This next line uses the actual formula for logistic regression to create
probabilities,
## and put the computed probabilities into a new column in MyData.
MyData$Prediction <- exp(predict(MyModel)) / (1 + exp(predict(MyModel)))
## Print the result out as a tab-delimited file.
write.table(MyData, sep="\t", file = "ModelPredictions.txt")

## Make a spreadsheet of the grammar.
idx <- coef(summary(MyModel))
idx
MyConstraints = round(idx, digits=3)
write.table(MyConstraints, sep="\t", file = "ConstraintsAndWeights.txt")
```

19. Examining output files from R

- Excel works well.

20. Constraints and Weights

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-4.699	0.08	-58.979	0
InitialZ	3.631	0.294	12.354	0
TH	-0.446	0.731	-0.61	0.542
InitialTHConsonant	3.241	0.894	3.626	0

Socrates:

- What does the big negative weight on Intercept mean?
- What's going on with the two constraints involving θ?

21. Performance of the grammar

- Looking at and interpreting ModelPredictions.txt
- You can use Excel to produce simple assessments of the grammar; we'll return to this later on.
- If you sort descending on IsSeuss, Prediction, you can get the most "Seussian" Seuss words according to the grammar:

	Word	IsSeuss	Frequency	InitialZ	TH	InitialTHConsonant	Prediction
164	Zomba-ma-tant	1	0	1	0	0	0.255834
165	Zans	1	0	1	0	0	0.255834
166	zang	1	0	1	0	0	0.255834
167	Zatz	1	0	1	0	0	0.255834
168	Zatz-it	1	0	1	0	0	0.255834
169	Zuff	1	0	1	0	0	0.255834
170	Zuk	1	0	1	0	0	0.255834
171	Zumm	1	0	1	0	0	0.255834

172	Zummzian	1	0	1	0	0	0.255834
173	Zorn	1	0	1	0	0	0.255834
174	Zed	1	0	1	0	0	0.255834
175	Ziff	1	0	1	0	0	0.255834
176	Ziffer-Zoof	1	0	1	0	0	0.255834
177	Zinn-a-Zu	1	0	1	0	0	0.255834
178	Zind	1	0	1	0	0	0.255834
179	Zinzibar-Zanzibar	1	0	1	0	0	0.255834
180	Zeep	1	0	1	0	0	0.255834
139	Thnad	1	0	0	1	1	0.129686
140	Thnadner	1	0	0	1	1	0.129686
141	Thneed	1	0	0	1	1	0.129686
142	Thwerll	1	0	0	1	1	0.129686
1	Obsk	1	0	0	0	0	0.009024

- And, for that matter the most Seussian real words:

4265	czar	0	24	1	0	0	0.255834
16954	tsar	0	26	1	0	0	0.255834
17822	xenophobia	0	18	1	0	0	0.255834
17823	xenophobic	0	5	1	0	0	0.255834
17824	xerox	0	18	1	0	0	0.255834
17825	xylophone	0	7	1	0	0	0.255834

These get more interesting if you add more constraints; in my current best grammar the most Seussian Seuss words are *Zomba-ma-tant*, *Zatz*, and *Zummzian*; the most Seussian real words are *xerox*, *snuff*, *snuggle*, *snug*, *snub*, *zoom*, *zoo*, *flux*, *flummox.*[3]

22. If time

Take a look at the data ((10)) and conjecture a few constraints that might work well.

MORE ON RATINGS DATA

23. What are we trying to do?

- There is pretty clearly a connection between frequency and intuitive well-formedness; e.g.
  - status of [dw] onsets in English vs. (say) some Bantu language where [dw] is very ordinary.

[3] *Flummox* is actually used by Seuss as the name of a animal.

- ditto for [ts] in English (*tsetse*, *tsunami*, *Tsongas*) vs. Japanese
- badness of [ʃɛd] as past tense of wug form [ʃɛi] (just only real example to support it) vs. [splʌŋ] as past test of [splɪŋ] (*fling*, *cling*, *string*, *ring*, sting, shrink, slink)

- This connection is obviously non-trivial; cf. all discussion so far on the Law of Frequency Matching and the various types of bias that make it imperfect.

24. Why should people have well-formedness judgments at all?

- One view:
  - To speech-perceive well, you need vast amounts of information about the probability of what you're likely to be hearing (this comes from *all* areas of linguistic knowledge, and some extra-linguistic ones as well)
  - People can, to varying degrees, consciously detect what their inner probability-assigning mechanisms are saying and translate the result into a a judgment.
  - The lower ends of the scale: **,*, ?? correspond to items that the grammar assigns a low probability.

25. The research that has to be done

- Acquisition model: mimic how humans can take in a childhood's worth of language data and produce a grammar that assigns probabilities to everything.
- Judgment model: understand how the grammatical probabilities are used in performing the various tasks that psycholinguistic subjects are asked to do.
- This, in turn, gets us into the question of variation between experimental tasks …

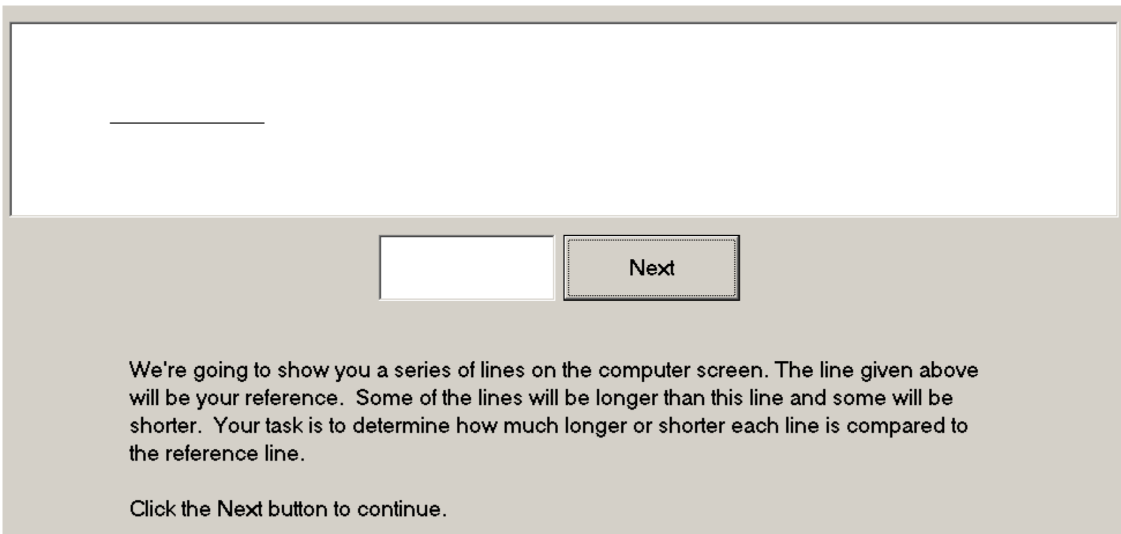
KINDS OF RATINGS DATA

26. A listing of types of ratings data

- Make a choice
  - Hungarian vowel harmony: moleːb-nɔk or moleːb-nɛk?
  - English past tenses: spling ~ splung/splang/splinged
- Give a yes/no verdict:
  - "Is *splung* appropriate as the past tense of *spling*?
  - "Could *blick* be a word of English?", "Does *blick* sound like a word of English?" etc.
- Rate on a Likert scale
  - Please check one of the options below (1-7) for how good *splung* sounds as the past tense of *spling*.
  - Now do the same for *splinged*.
- Binary comparisons

- Example from Daland et al. (2009): "[subjects were asked to] choose the non-word that seemed more like a typical English word. The practice items were *stallop* vs. *thmeffle*, *lbobbib* vs. *priffin*, *thrishal* vs. *ftemmick*, *skeppick* vs. *mzibbus*, *shmernal* vs. *dwiffert* and *shthokkith* vs. *thpellop*.

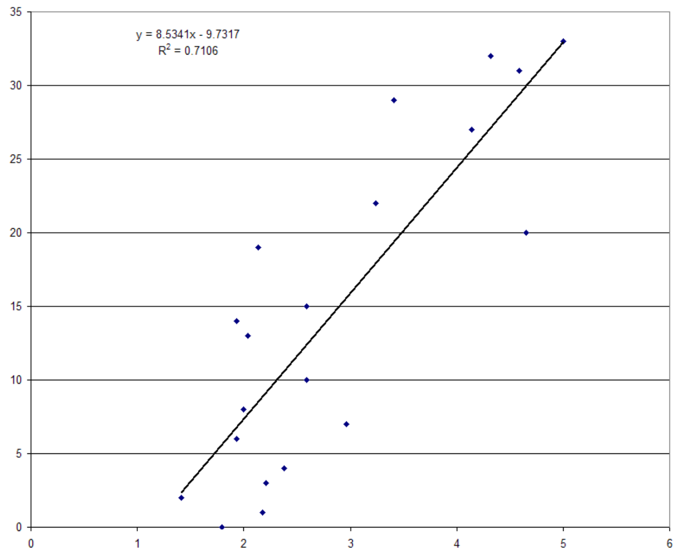
* Magnitude estimation
  - Please draw a line with the mouse that matches the goodness of check one of the options below (1-7) for how good *splung* sounds as the past tense of *spling*.
  - Software screen for Hayes and White (2013):



  - The other task in magnitude estimation is simply to type in a number.

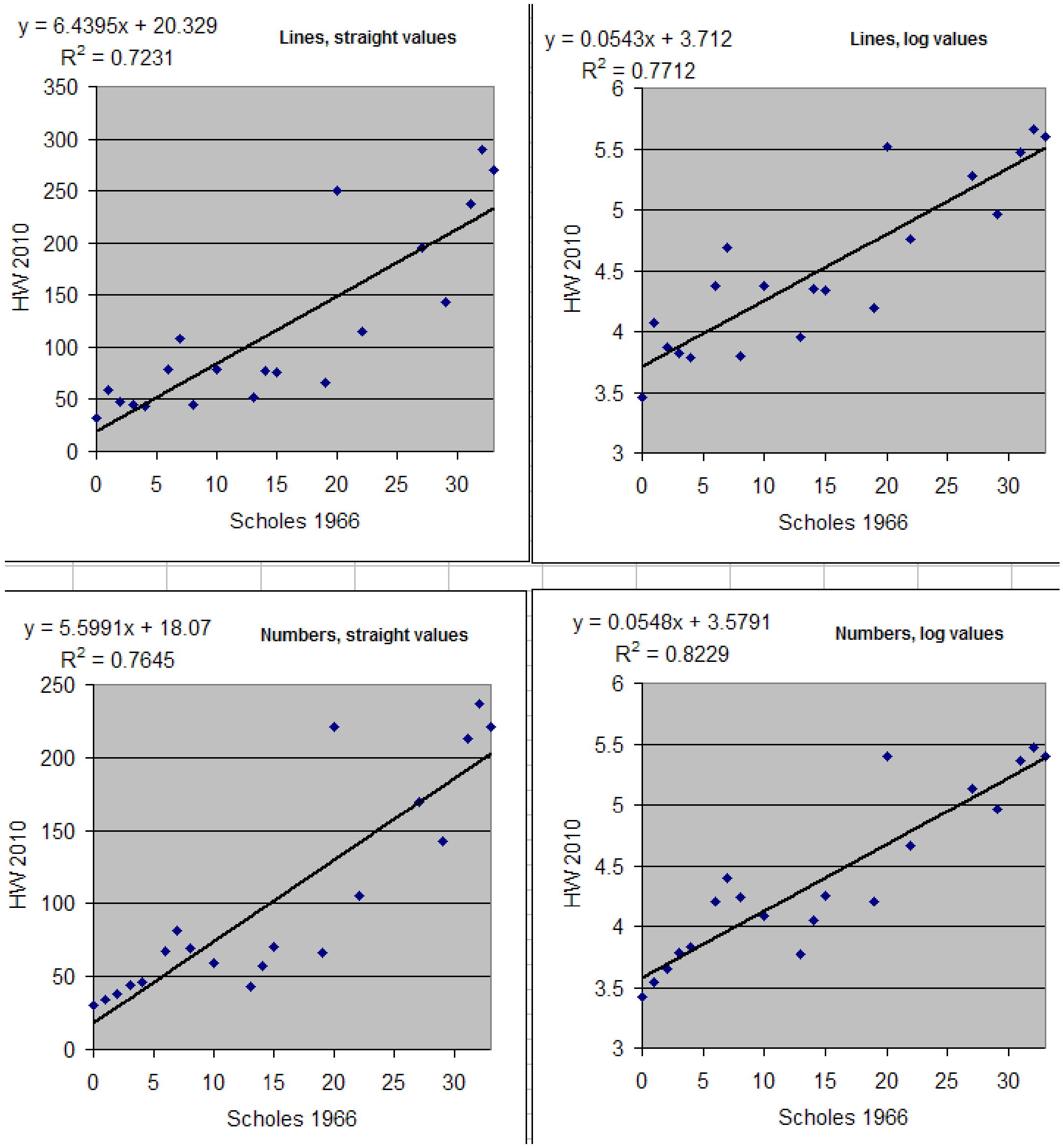
27. Binary choice vs. ratings

* I believe that these typically give similar results.
* One case I can remember: Hayes and White (2013) used filler items that matched those of Scholes (*Phonotactic Grammaticality* 1965).
  - Forms: *blung, fnet, frun, glung, shlurk, shmat, shnet, shtin, skeep, smat, srun, stin, vkeep, vlurk, vnet, vrun, zhmat, zlurk, znet, zrun*
  - Scholes: a class of seventh graders making up-down decisions
  - Hayes/White I (not published): a bunch of Mechanical Turkers rating on a 1-7 Likert scale
  - Correl. = .843
  - Scatterplot:



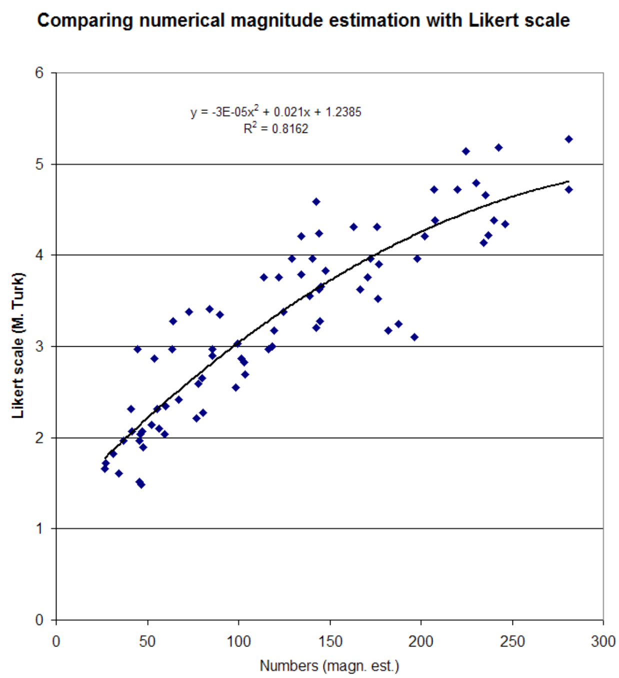
28. Binary choice vs. magnitude estimation

- Again Hayes and White, expt. 2 (published)
- The fit is not bad and works a bit better if you take the log of the magnitude estimation values.



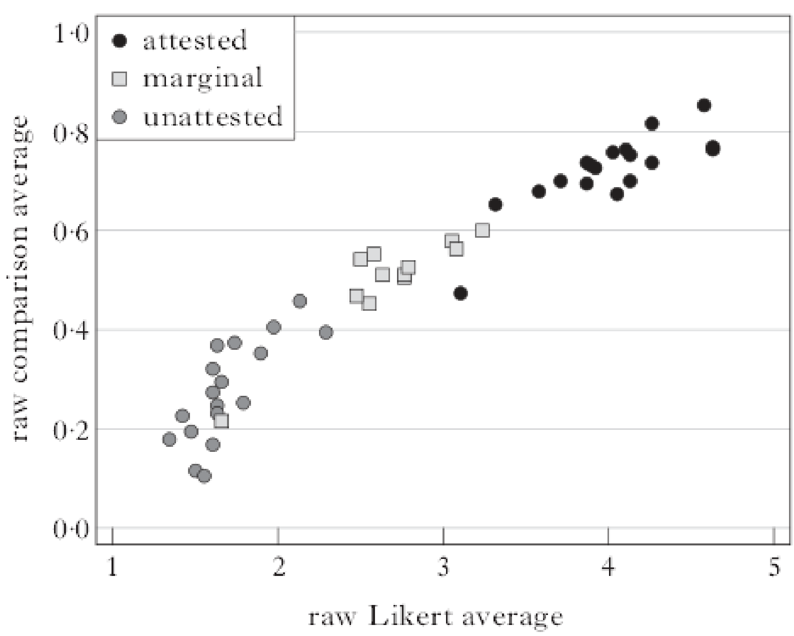
29. Likert scale vs. magnitude estimation

- Here are all the Hayes/White forms, the two experiments compared.
- It seems a bit nonlinear.



30. Ratings vs. comparisons

- Daland et al. got hard-to-interpret results with Likert-scale ratings of their stimuli.
- They then redid the experiment with comparison — every stimulus compared with every other, and got results that seemed more meaningful.
- Here is a scatterplot of the same stimuli across experiments:



➢ The curving at the left side means: distinctions made in direct comparison were largely not made in Likert-scale rating, suggesting comparison is more sensitive.

- The paper says, "This fact suggests the following methodological point: in non-word acceptability studies, head-to-head comparison is preferable to Likert rating whenever the stimuli of interest are concentrated at one end of the well-formedness scale, owing to ceiling/floor effects in Likert ratings. Similar conclusions have been reached by [various researchers]; we mention this methodological point here in the hope of averting unnecessary replication of effort in the future."

31. The controversy over magnitude estimation

- In principle, magnitude estimation is nice:
  - scale is refined as much as the subject would like
  - scale can be instantly extended, e.g. if you hear a new words that is unprecedently awful or wonder
  - In uncontroversial cases "how long is this line?", people behave reliably and consistently.
- For discussion of the method, see the following:
  - Pro: Bard, Ellen Gurman, Dan Robertson and Antonella Sorace. 1996. Magnitude estimation of linguistic acceptability. *Language* 72:32-68.
  - Pro: Lodge, Milton. 1981. *Magnitude scaling: Quantitative measurement of opinions*. Beverly Hills/London: Sage.
  - Con: Sprouse, John (2011) A Test of the Cognitive Assumptions of Magnitude Estimation: Commutativity does not Hold for Acceptability Judgments. *Language* 87.2

32. Upshot

- Reassuringly, different methods do seem to yield different results.
- Actually determining what methods are most reliable is something where I would want to rely on expert opinion (i.e., based on extensive comparative work by experienced experimental psychologists).

A FREQUENCY-BASED MODEL THAT DOESN'T USE PROBABILITY:
ALBRIGHT AND HAYES (2003)

33. Minimal generalization

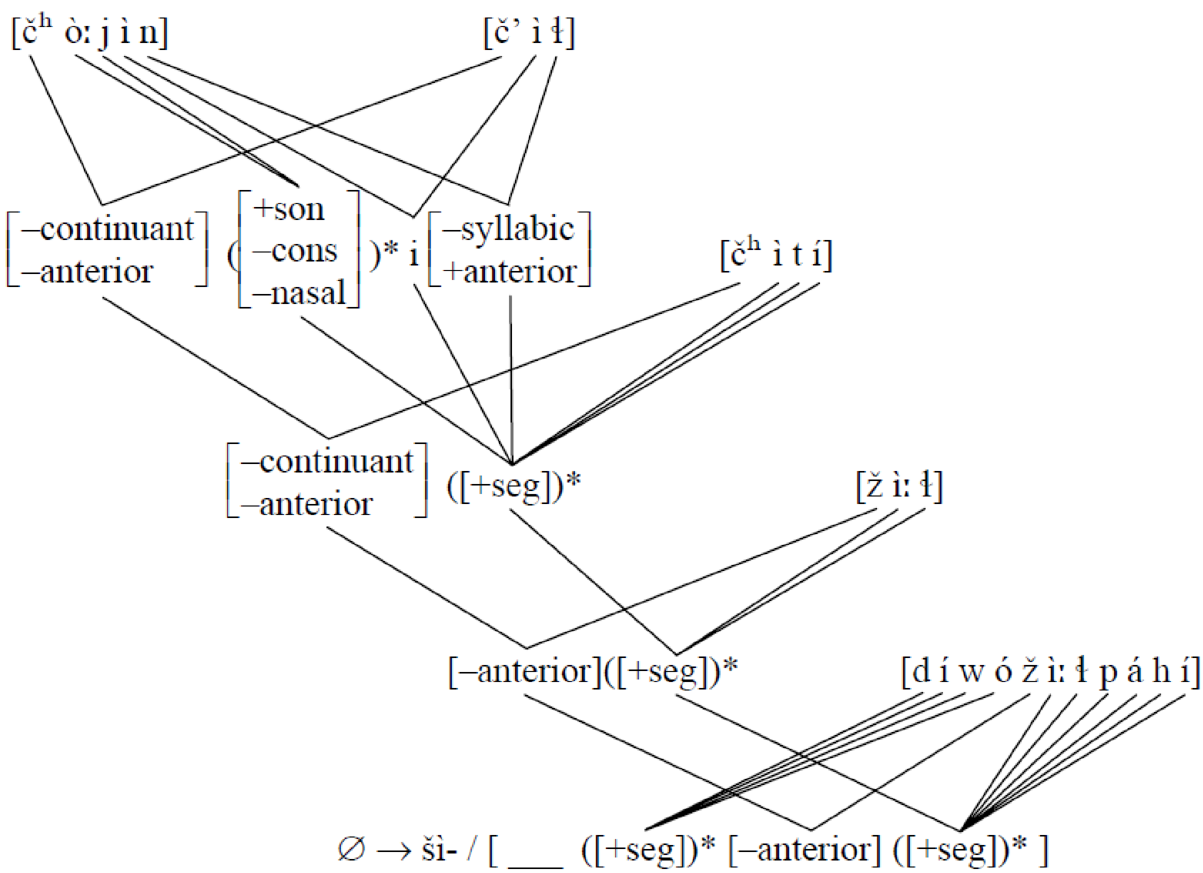
If you were trying to find environments where Navajo prefixes [ʃi-] instead of [si-] (this turns out to be sibilant harmony), you could take each case as a "microrule" and start generalizing:

a. $\varnothing \rightarrow$ šì / [___ tàš]
$\varnothing \rightarrow$ šì / [___ t͡ɬé:ž]

b.

$$\begin{array}{llllllllll} & \varnothing & \rightarrow & \text{šì} & / & [& \_\_\_ & \text{t} & \text{à} & \text{š} \quad] \\ + & \varnothing & \rightarrow & \text{šì} & / & [& \_\_\_ & \text{t͡ɬ} & \text{é:} & \text{ž} \quad] \\ \hline = & \varnothing & \rightarrow & \text{šì} & / & [& \_\_\_ & \begin{bmatrix} -\text{sonorant} \\ -\text{continuant} \\ -\text{spread gl.} \\ +\text{anterior} \end{bmatrix} & \begin{bmatrix} +\text{syllabic} \\ -\text{high} \\ -\text{round} \end{bmatrix} & \begin{bmatrix} -\text{sonorant} \\ +\text{continuant} \\ -\text{anterior} \\ +\text{strident} \end{bmatrix} \quad] \end{array}$$

If you keep going, you'll get the final version:



34. Reliability

- The "correct" rule for English past tenses is "Add -t/d/əd".
- But you can do an unorthodox special rule: "Add -t after a voiceless fricative."
- Unorthodox, but perfect! All 352 voiceless fricatives in Albright/Hayes's corpus are regular.

- And indeed, in a wug test people really like past tenses like *bliffed* or *daced*.

35. The accuracy/scope tradeoff

- People like generalizations that are really accurate.
- Cf. voiceless fricative "island of reliability", above.
- ɪ → ʌ / C liquid ___ ŋ is perfect, but there are only 4 examples.
- So Albright/Hayes use a statistical adjustment that reflects both principles; i.e. 95% lower confidence limit on the "batting average" of the rule.
- This is an evaluation score for each rule.

36. The "use the best rule" principle

- To wug-test a form, for each applicable past tense type, find the applicable rule with the best evaluation score (ranging from zero to one).
- That is the score assigned to the past tense candidate.

37. The resulting model is not a probability model

- Outputs are evaluated individually, not in competition.
- Indeed, Albright/Hayes produce four categories of wug verbs:
  - regular predicted good, irregular predicted good
    ***dize** [daɪz] (doze [doz]); **fro** [fro] (frew [fru]); **rife** [raɪf] (rofe [rof], riff [rɪf])*
  - regular predicted not so good, irregular predicted good
    fleep *[flip] (*flept *[flɛpt]);* **gleed** *[glid] (*gled *[glɛd],* gleed*);* **spling** *[splɪŋ] (*splung *[splʌŋ],* splang *[splæŋ])*
  - regular predicted good, irregular predicted bad
    [brɛdʒ] (*broge* [brodʒ]); ***gezz*** [gɛz] (*gozz* [gaz]); ***nace*** [nes] (*noce* [nos])
  - regular predicted not so good, irregular predicted bad
    gude [gud] (*gude*); ***nung*** [nʌŋ] (*nang* [næŋ]); ***preak*** [prik] (*preck* [prɛk], *proke* [prok])

38. Models can be "probabilized"

- Take their scores and treat them like Harmony; then do maxent.
- This doesn't help with Albright/Hayes; in the aggregate correlations go down.

regulars: .714 → .500
irregulars: .485 → .510