Voice-specific lexicons: acoustic variation and semantic association

Ed King and Meghan Sumner
Stanford Linguistics

Overview

* spoken words are highly variable, and are never produced the same way twice

* types of variability
  * linguistic (allophonic)
    * larely predictable from phonology and phonetics
  * speaker-specific (acoustically cued)
    * due to speaker age, sex, etc.

* how do listeners understand speaker-specific variation?

Overview

* current theories (e.g., exemplar theory) claim that phonetic variation is important, and stored in the lexicon
  * Goldinger, 1996
  * Johnson, 2006

* semantic relationships in the lexicon are also important in understanding spoken words
  * Neely 1977
  * Lucas, 2000 for review

* so far, little linkage between the two

Speaker-specific lexicons

speaker	word	referent?
older British woman	“princess”	Diana, Kate Middleton
young girl	“princess”	Jasmine, Cinderella
nerdy man	“princess”	Leia, Peach

our question:
do phonetic variation and semantic association interact in spoken word recognition?

Speaker-specific variation

* listeners have better memory for words with phonetic characteristics consistent with prior exposure (over the course of an experiment)
    * speaker sex (Schacter & Church, 1992)
    * individual voices (Goldinger 1996)
    * speaking rate (Bradlow et al, 1999)

Speaker-specific variation

* listeners show perceptual shifting to tokens that match expected speaker characteristics
  * speaker dialect (Niedzielski, 1999; Hay, Warren, & Drager, 2006)
  * speaker sex (Johnson et al, 1999; Strand 1999)

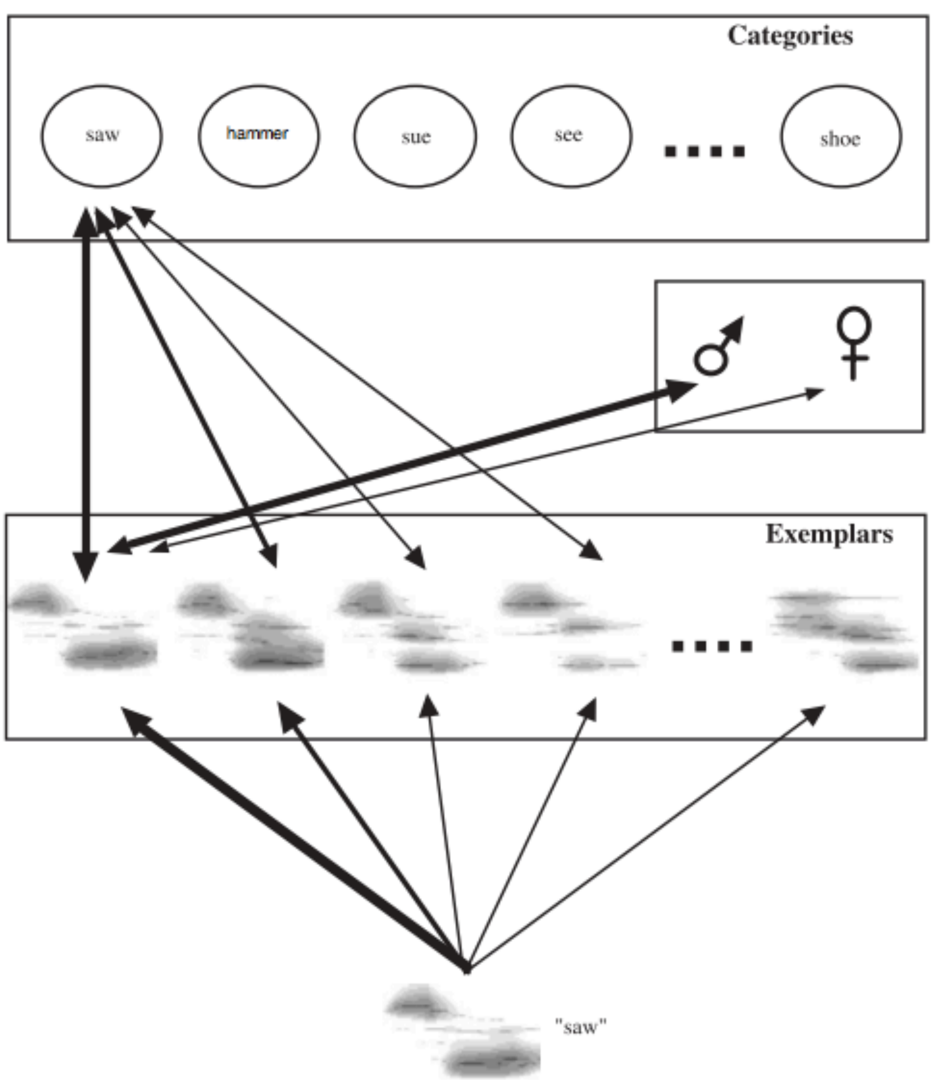
Speaker-specific variation

* listeners show perceptual benefits for **words** that match speaker characteristics
  * emotional prosody (e.g., “dye” v. “die”, Nygaard & Lunders, 2002)
  * speaker age (Walker & Hay, 2011)
  * speaker sex (Hay & Walker, 2013)

Modeling speaker-specific variation

* exemplar models store detailed phonetic representations in the lexicon (Goldinger, 1996; Johnson, 2006)
    * incoming speech signals are directly compared to these phonetic representations
    * signals matched to the lexicon based on phonetic similarity

Modeling speaker-specific variation

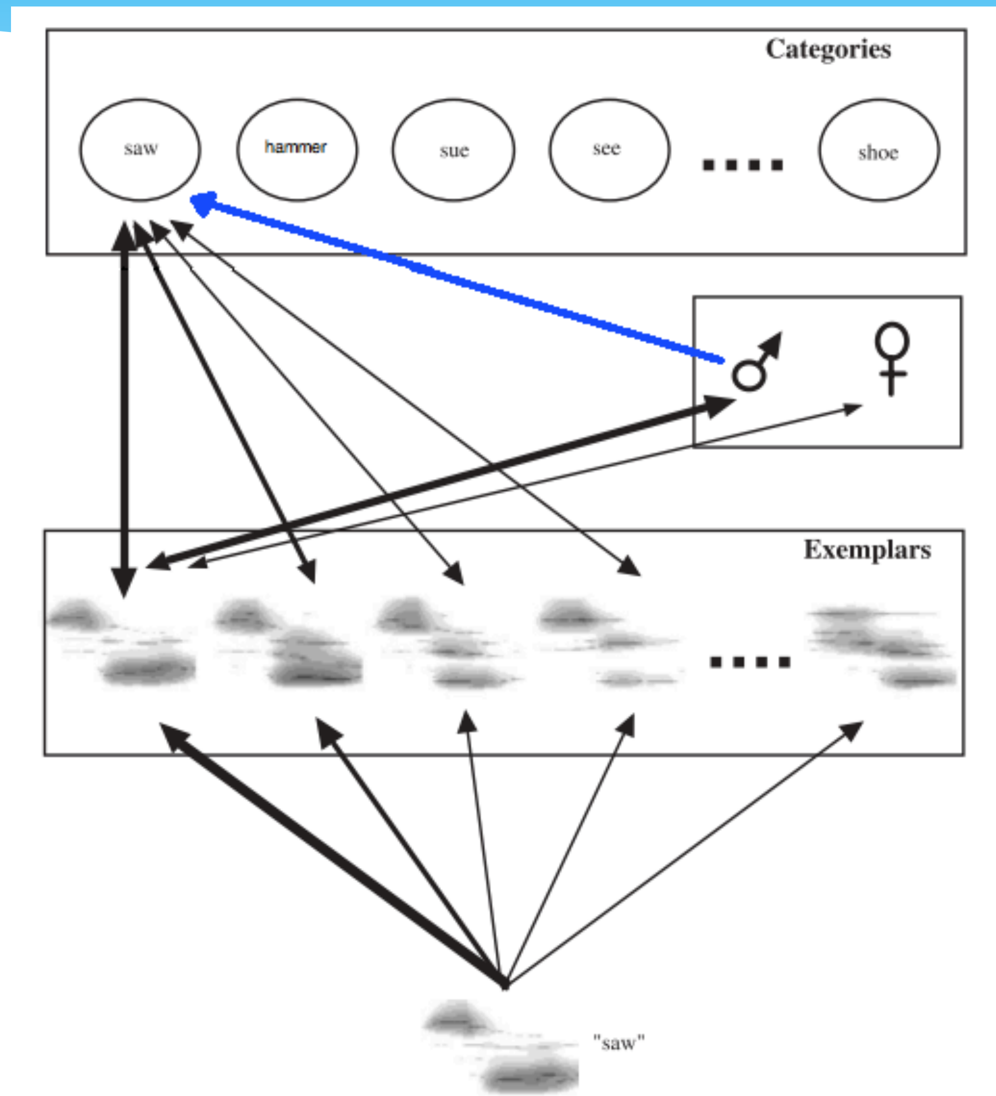


Johnson (2006)

Modeling speaker-specific variation

* exemplar models can capture data on emotional prosody, age, sex
  * lexical activation is sensitive to the phonetic distribution of *experienced* tokens
  * some words heard more often (globally) with particular speaker-specific variation

Modeling speaker-specific variation

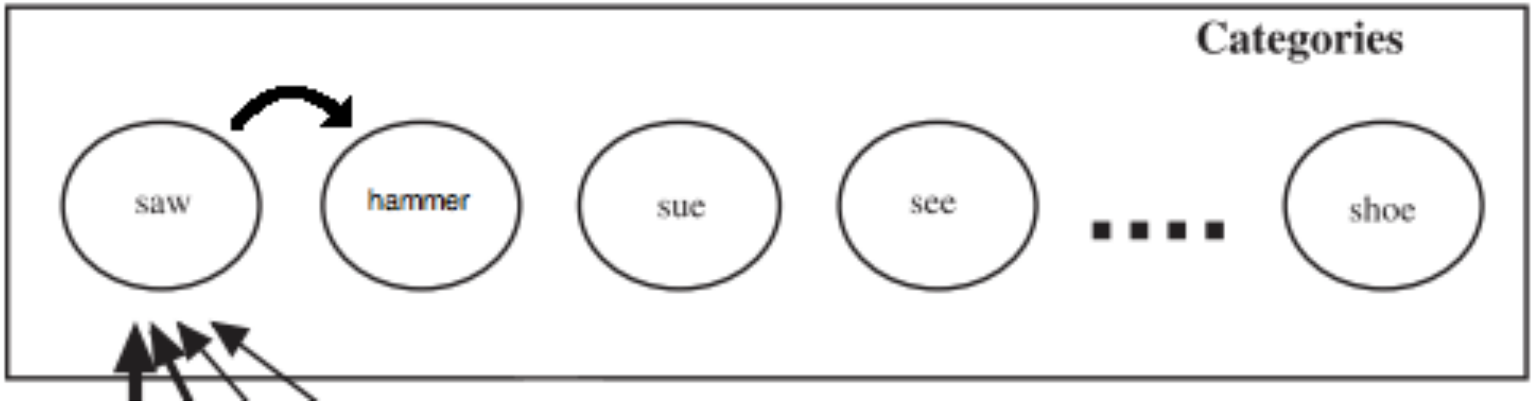


Johnson (2006)

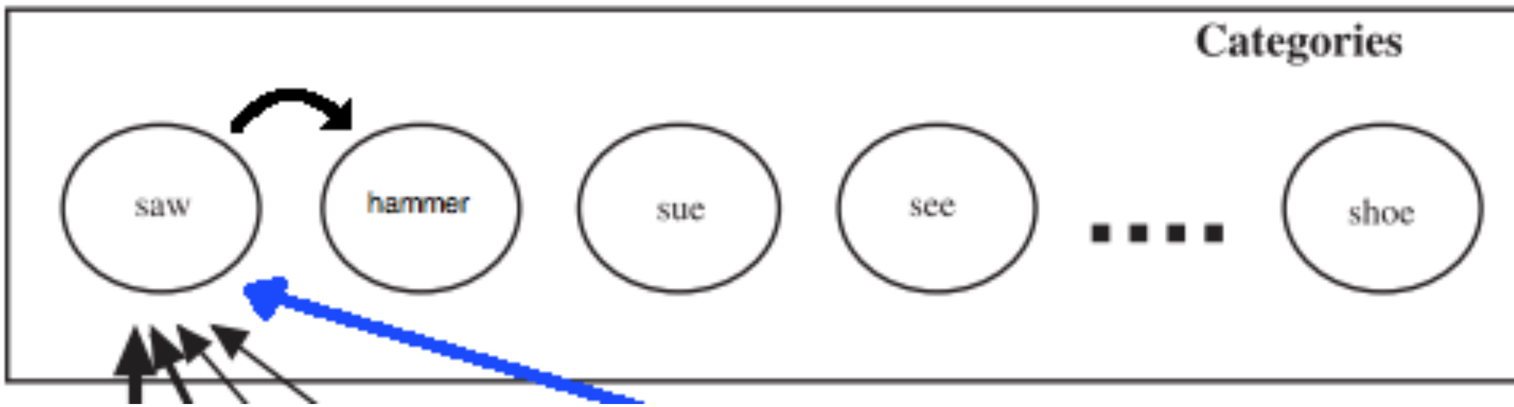
Modeling speaker-specific variation

* but these are primarily studies of *representation*
  * what effects do speaker features have on further lexical activation (e.g., semantically-related words)?

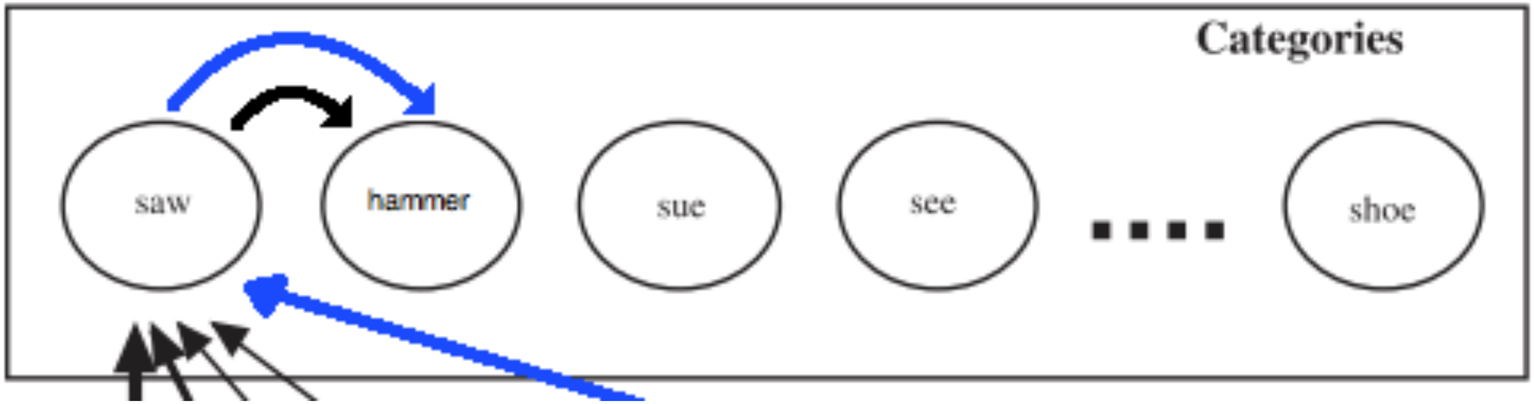
Modeling lexicons



Modeling lexicons



Modeling speaker-specific lexicons?



Our approach

* do words activate different semantic associates depending on speaker?
  * large-scale free association task
* if so, how quickly do speaker-specific semantic associations appear in spoken word recognition?
  * semantic priming task

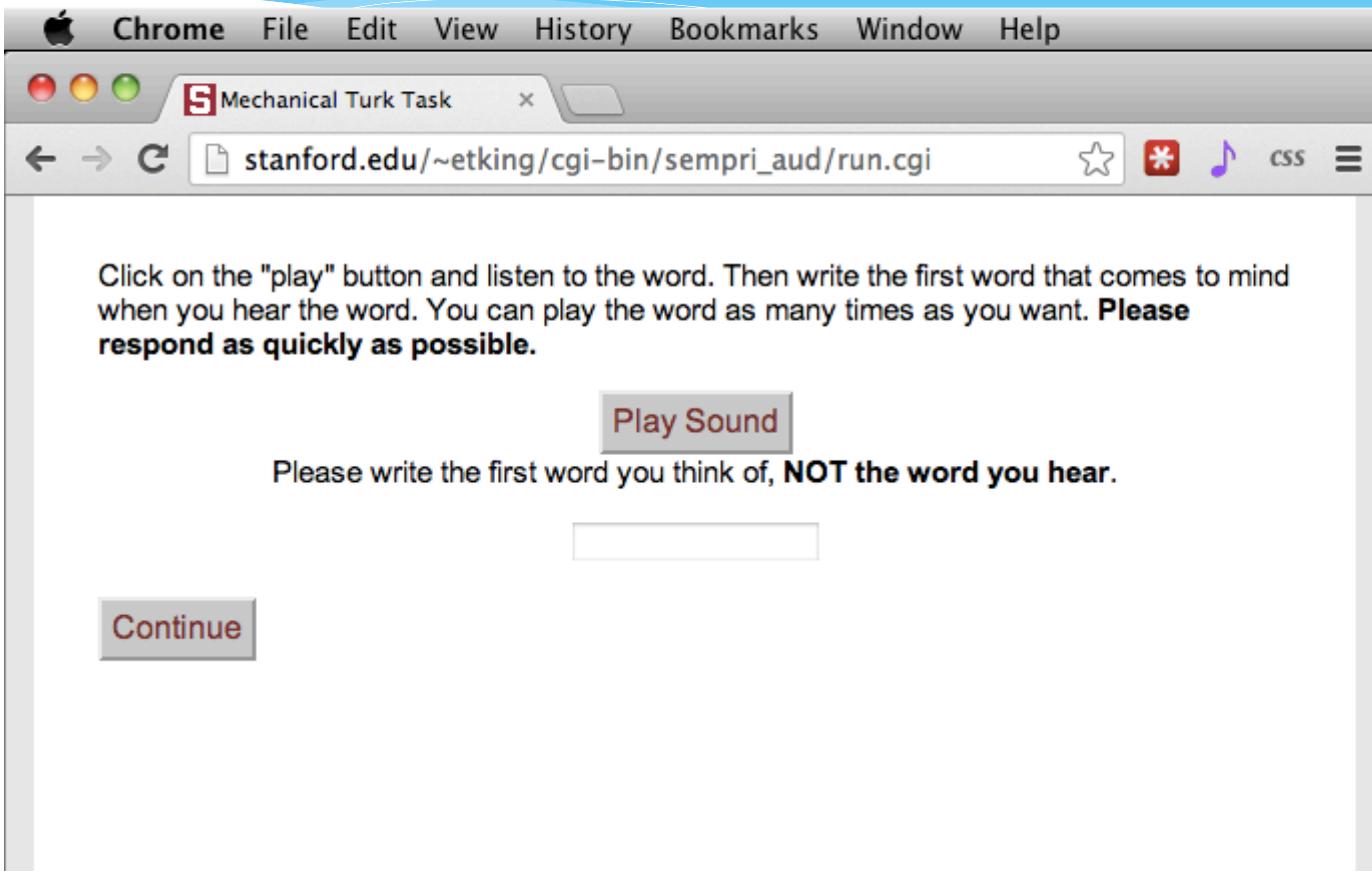
Free association task

* do listeners respond with different semantic associates depending on speaker sex?
  * 262 randomly chosen prompt words, recorded by male and female speakers
  * 191 Mechanical Turk participants heard all prompt words by either male or female speaker
    * responded to prompt "Write the first word that comes to mind"

Free association task – stimuli

Jay	Em
male	female
early 80s	late 30s
African American	White
Southern (MS)	Northern (NY)

Free association task



Free association results

* responses to each prompt word ranked by *association strength* – the percentage of participants to provide that response
  * (responses grouped by lexical stem)

* we examine
  * *top associates* – the strongest response to each prompt

Free association – top associates

* 61 prompt words (22%) produced different top associates across speakers
  * higher than random intra-speaker baseline (p<

  * ACADEMY: *school* (Jay) *awards* (Em)
  * PRETTY: *ugly/girl* (Jay) *beautiful* (Em)
  * CONFERENCE: *call* (Jay) *meet* (Em)

Semantic priming

* free association shows some differences in speaker-specific word association
  * do these differences appear in on-line word recognition?
* cross-modal semantic priming
  * listeners hear a prime, then see a target word (semantically related or unrelated to the prime) or nonword
  * listeners decide whether the target is a word
  * faster recognition to semantically related targets than to unrelated targets

Semantic priming

* our design crosses relatedness with speaker congruence, using 24 words from our free association results

* four types of trials:
  * related, speaker match
    * ACADEMY (Jay) -> *school*, ACADEMY (Em) -> *award*
  * related, speaker mismatch
    * ACADEMY (Jay) -> *award*, ACADEMY (Em) -> school
  * unrelated: ACADEMY -> whistle
  * nonword : ACADEMY -> *troded*

Semantic priming

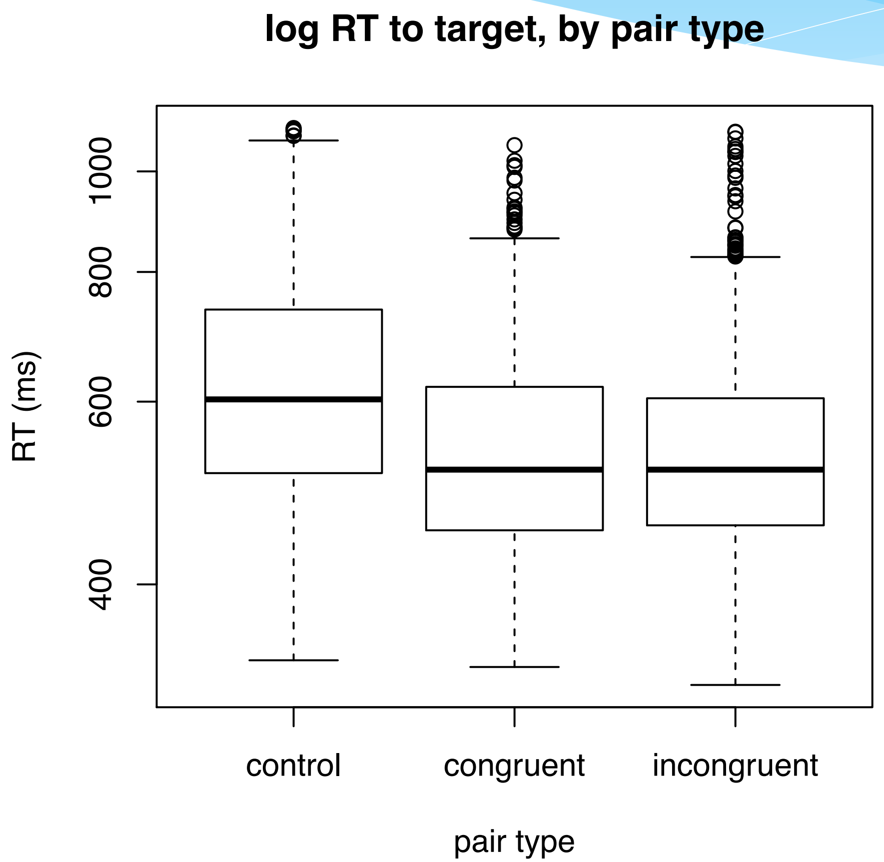
* 48 subjects participated in a lab setting
* counterbalanced lists
  * no subject heard a prime or saw a target more than once
  * all subjects heard both speakers throughout the experiment

Semantic priming – results

* reaction times were analyzed in two ways:
  * *categorical* : sex match or mismatch
  * *gradient* : as a function of speaker-specific association strength

* reaction times from incorrect lexical decisions, and those less than 300ms or greater than 2 std deviations above the mean, were excluded

Semantic priming – results



Semantic priming – results

* overall semantic priming is significant (related v. control)
  * 85-88ms ($p < 0.001$)

* sex-congruent slightly faster than incongruent, but not significant
  * 2.4ms ($p > 0.7$)

Semantic priming – results

* lack of effect may be due to variance in association strength
  * though all words were top associates, according to free association task, strengths ranged from 5% to 49%

Jay		Em	
girl	15	beautiful	20
ugly	15	pink	12
beautiful	13	**ugly**	10
		girl	8

Semantic priming – results

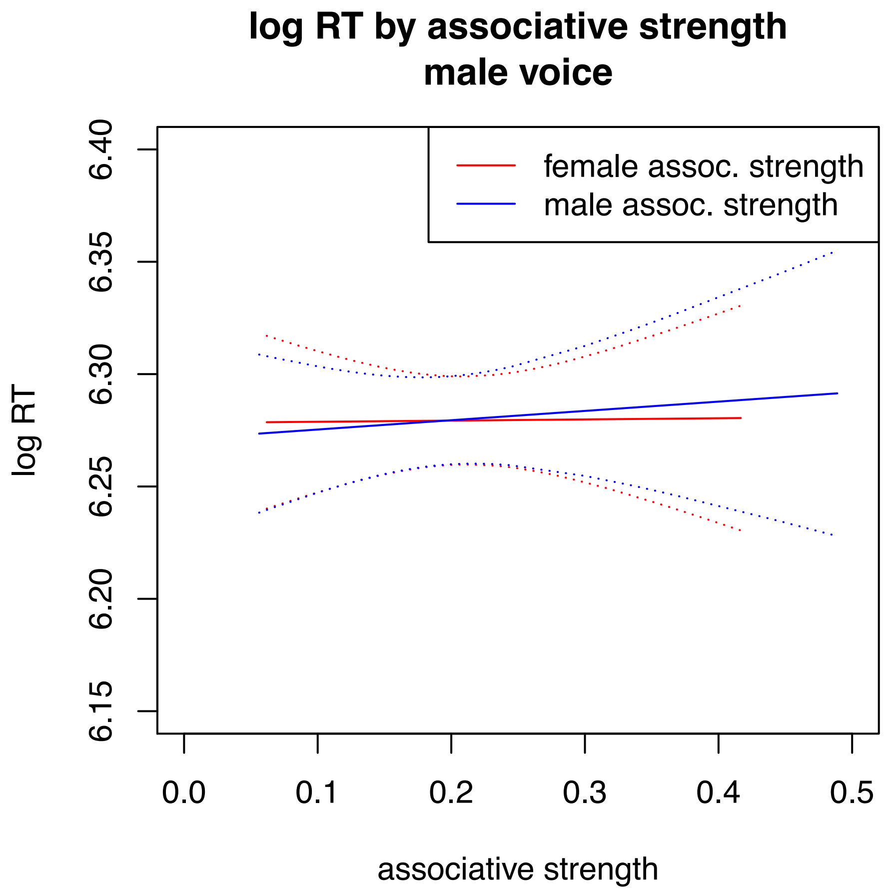
* lack of effect may be due to variance in association strength
  * though all words were top associates, according to free association task, strengths ranged from 5% to 49%

Jay		Em	
girl	15	**beautiful**	20
ugly	15	pink	12
beautiful	13	ugly	10
		girl	8

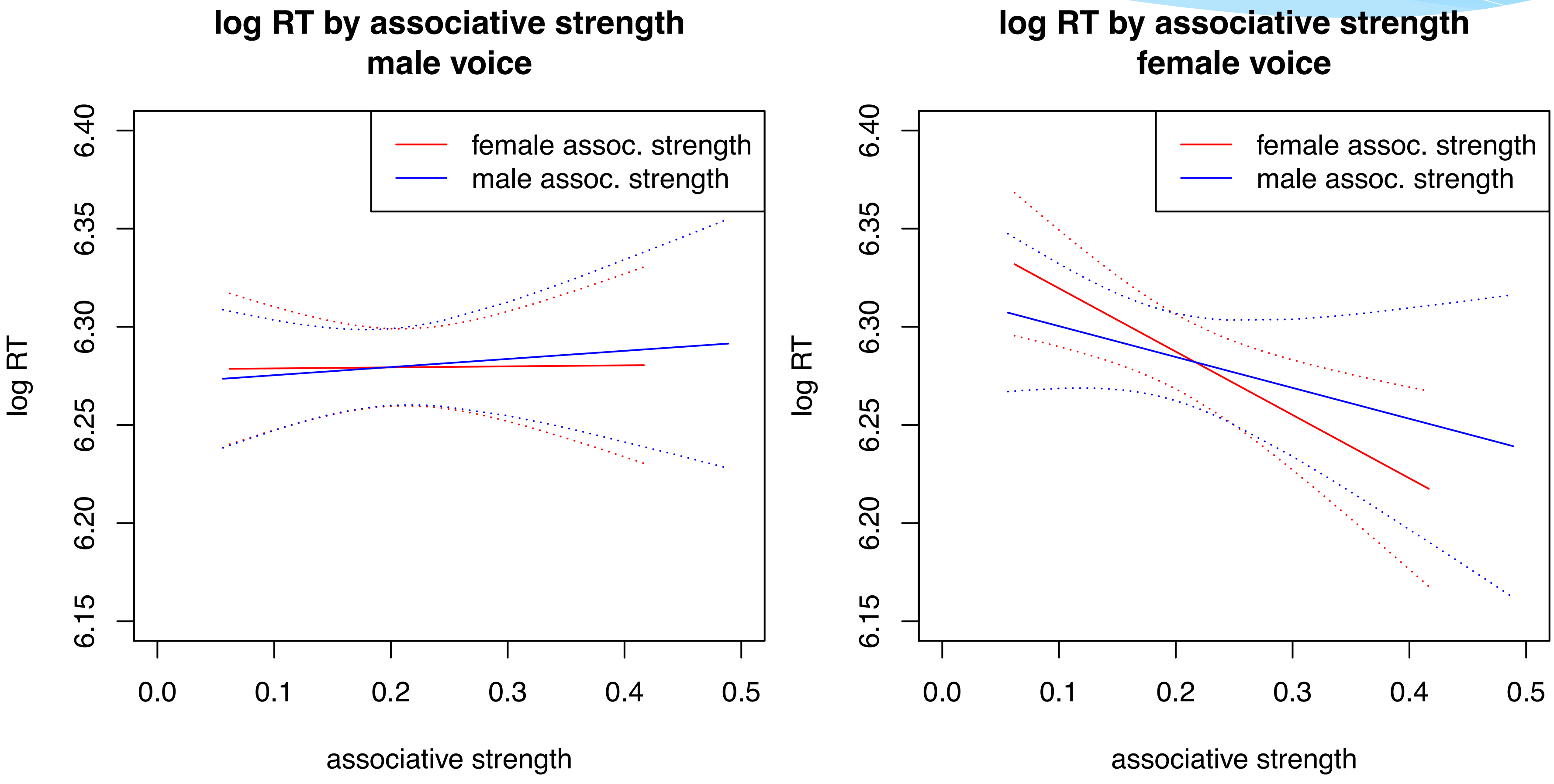
Semantic priming – results

* so, we test reaction time as a function of strength
  * split data into Jay and Em voice primes
  * compare each speaker’s association strength within voice

Semantic priming – results



Semantic priming – results



Semantic priming

* Em primes:
  * Em strength: $p < 0.01$
  * Jay voice: n.s. ($p = 0.19$)
* Jay primes:
  * Em strength: n.s. ($p = 0.99$)
  * Jay strength: n.s. ($p = 0.71$)

* Jay's voice appears to prime all related words equally well, while Em's voice shows a gradient effect of her association strength

General Discussion

* a subset (22%) of 262 random words prompt different free association responses depending on speaker sex

* in an on-line priming task, listeners' reaction times fall significantly as a function of association strength to female primes, when the primes are spoken in a female voice

General Discussion

* why only an effect for female-congruent targets?
  * no convincing answer yet
  * but, various free association studies, significantly different lexical richness in responses dependent on speaker characteristics (King & Sumner, submitted)
    * reponses to female speakers tend to show stronger top associates, and comparatively weaker non-top associates

General Discussion

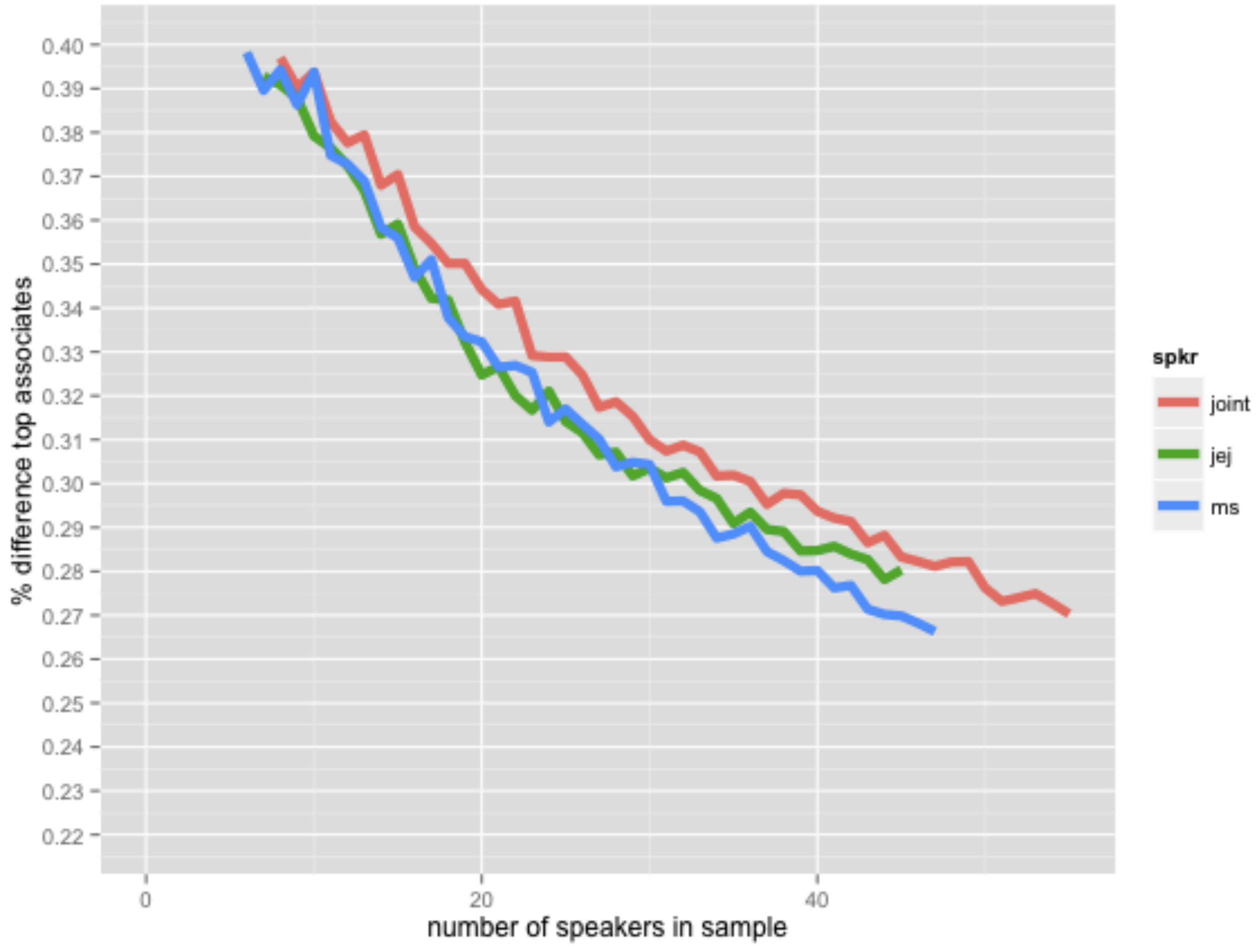
* how to model speaker-specific semantic associations?
  * can we find these differences outside of free association?
    * corpus tasks?

* how to model the interaction in a word processing model?
  * do these differences arise because of usage/experience?
  * or are we tapping into stereotypes?

Conclusion

* different speakers prompt different semantic associates in a free association task
  * 22% of words are different, significantly higher than expected
* these differences are reflected in on-line processing
  * but only for our younger, female, speakers
* evidence for an interaction between acoustic variation and semantic association

Free association – difference baselines



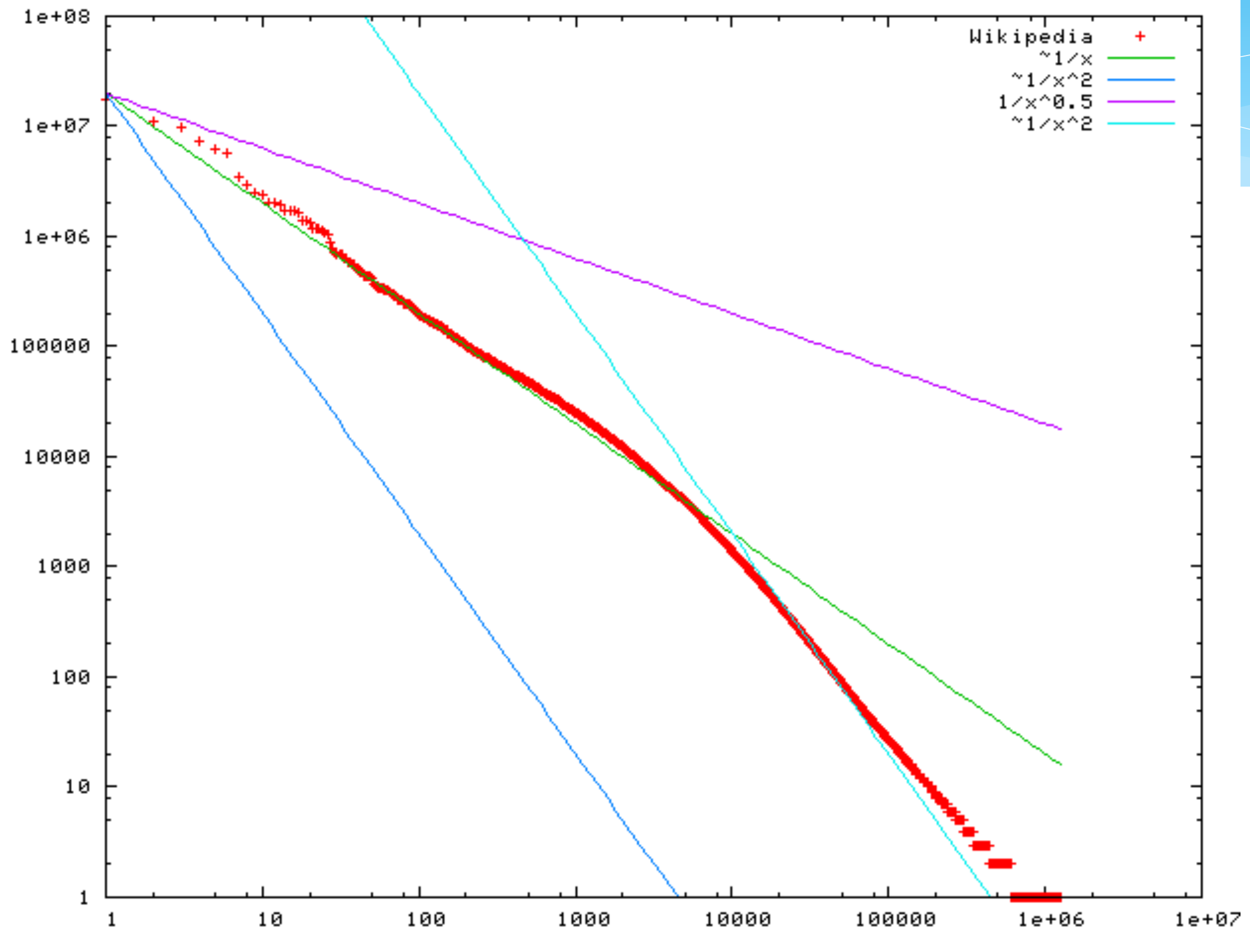
Free association – difference baselines

```
> with(joint[joint$setsize>40&joint$setsize<50,],summary(lm(difference_proportion ~ spkr)))

Call:
lm(formula = difference_proportion ~ spkr)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.067918 -0.014977 -0.000284  0.014984  0.068992 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.2854750  0.0007258 393.301  < 2e-16 ***
spkrjej     -0.0033223  0.0012146  -2.735  0.00628 ** 
spkrms      -0.0141991  0.0010974 -12.939  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02178 on 2097 degrees of freedom
Multiple R-squared:  0.07667, Adjusted R-squared:  0.07579 
F-statistic: 87.07 on 2 and 2097 DF,  p-value: < 2.2e-16
```

Free association – structure

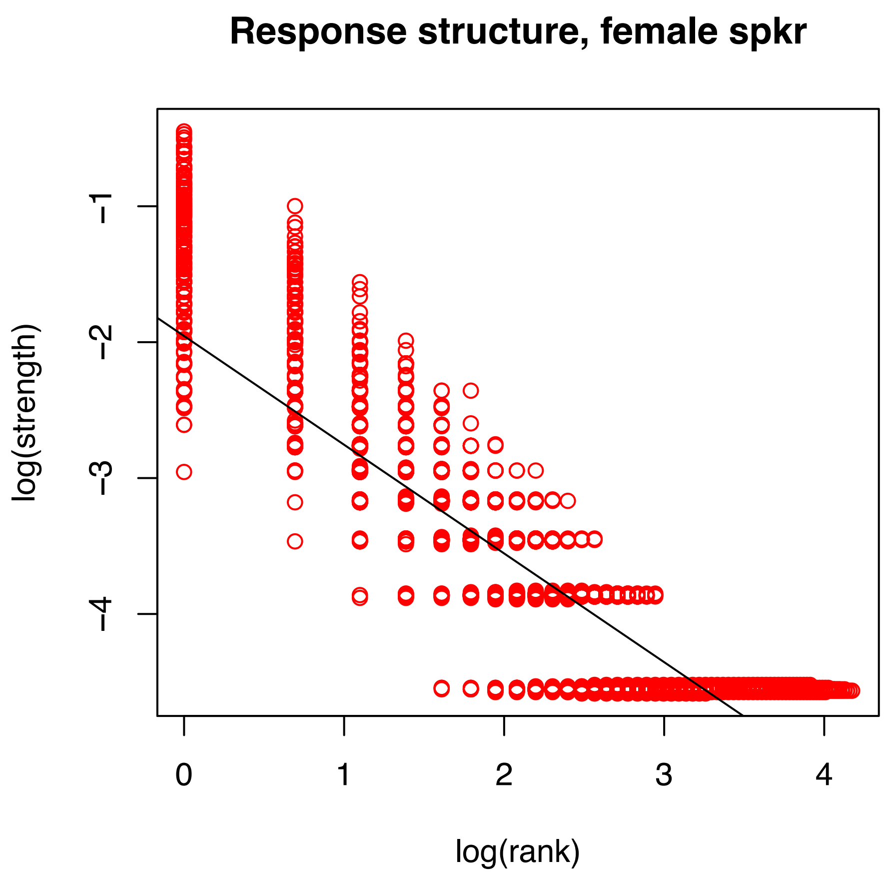
* overall structure of responses: Zipf's Law
  * the *n*th most common response should occur 1/*kn* times as often as the most common (for some constant *k*)
  * linear function in log-log space
    * (log frequency ~ log sorted rank)

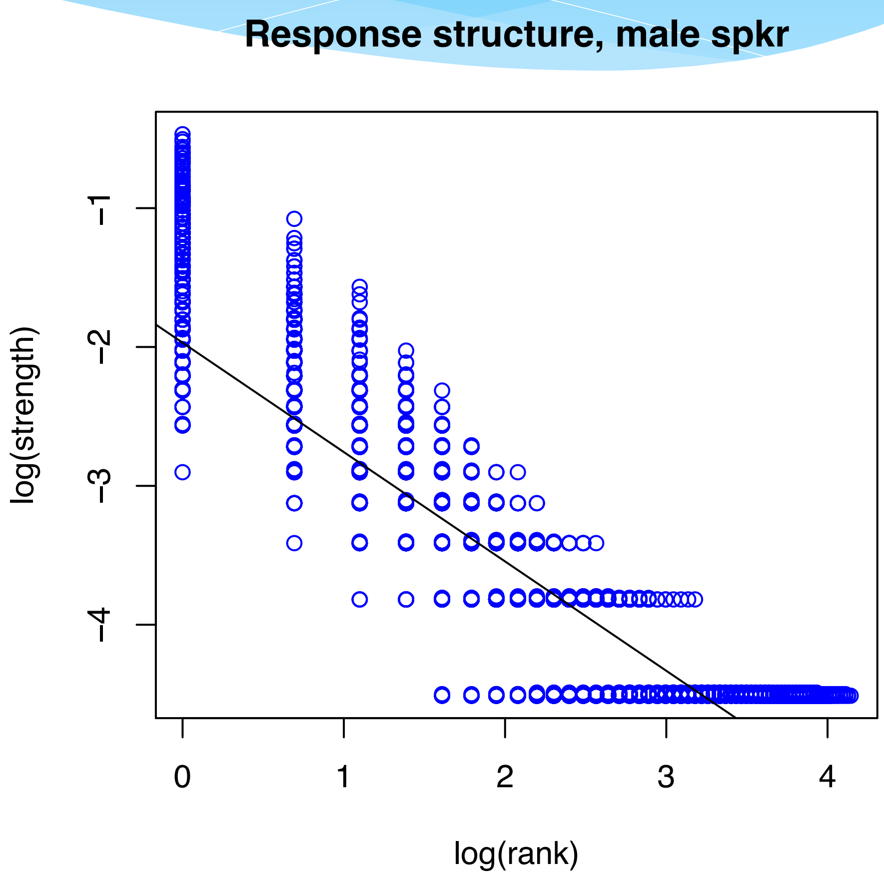
Free association – structure



en.wikipedia.org

Free association – structure





Free association – structure

```
> pvals.fnc(zipf.model.spkr.interaction)
$fixed
                 Estimate MCMCmean HPD95lower HPD95upper  pMCMC Pr(>|t|)
(Intercept)       -1.9160  -1.9170    -1.9422    -1.8905 0.0001   0.0000
log(rank)         -0.8168  -0.8163    -0.8238    -0.8082 0.0001   0.0000
spkrms             0.0216   0.0215    -0.0080     0.0522 0.1586   0.1624
log(rank):spkrms  -0.0164  -0.0164    -0.0273    -0.0059 0.0030   0.0027

$random
    Groups        Name Std.Dev. MCMCmedian MCMCmean HPD95lower HPD95upper
1    prime (Intercept)   0.1291     0.1189   0.1191     0.1089     0.1301
2 Residual               0.3317     0.3320   0.3321     0.3286     0.3354
```