Context Effects in Spoken Word Recognition of English CVCCVC words and nonsense words

Robert Felty
University of Michigan
robfelty@umich.edu

Acknowledgments

I would like to thank the members of my committee for their support and input

- José Benkí (co-chair, Otolaryngology)
- Patrice Beddor (co-chair, Linguistics)
- Robin Queen (Linguistics and German)
- Robert Kyes (German)

I would also like to thank those who helped me prepare and pilot the stimuli for the experiments and the Linguistics department at the University of Konstanz for letting me use their facilities, and for their hospitality.

Background: Lexical Access

- The field of Lexical Access seeks to determine how the mental lexicon affects language processing.
- Two classes of models differ in their predictions of how morphologically complex words are stored in the lexicon and accessed.

Background: Models

- Associative Models
  - Claim that words are stored whole in the lexicon
  - Examples: TRACE, MERGE
- Combinatorial Models
  - Claim that morphemes are stored separately and combined during lexical access
  - Also known as morphological decomposition models

Background: Context Effects

Previous research has found several different context effects which play a role in word recognition

I will be focusing on the following context effects:

- Lexical status (word or nonword)
- Lexical frequency (how often a word occurs)
- Neighborhood Density (how similar a word is to other words)

Background: Previous Research

- Using a Lexical Decision task, and a Cross-modal Priming task, Clahsen et al. (2001) found a difference in processing of German inflected adjectives.

	-m dominant adjectives				-s dominant adjectives		
	Stem form	-m	-s		Stem form	-m	-s
ruhig	838	51	13	rein	783	14	38

Background: Previous Research

- Using a Lexical Decision task, and a Cross-modal Priming task, Clahsen et al. (2001) found a difference in processing of German inflected adjectives.

	-m dominant adjectives				-s dominant adjectives		
	Stem form	-m	-s		Stem form	-m	-s
ruhig	838	51	13	rein	783	14	38

Background: Previous Research

- Using a Lexical Decision task, and a Cross-modal Priming task, Clahsen et al. (2001) found a difference in processing of German inflected adjectives.

	-m dominant adjectives				-s dominant adjectives		
	Stem form	-m	-s		Stem form	-m	-s
ruhig	838	51	13	rein	783	14	38

Research Questions

- Is the mental lexicon organized in a combinatorial or an associative way?
- That is, are morphemes stored separately in the lexicon and then combined to form words during lexical access, or are words stored whole in the lexicon?
- What influence does phonetics have in the processing of multimorphemic words?

Task / Subjects

- Open Response Speech-In-Noise Task
- 2 different Signal to Noise Ratios (SNRs) used – -5dB and 0 dB
- signal dependent (but uncorrelated) noise (see Schroeder, 1968)
- 30 Native American English speakers participated

Materials

- 150 CVCCVC words
  - 74 monomorphemic
    bandage [bændɪʤ] *toxic* [taksɪk] *hectic* [hɛktɪk]
  - 76 bimorphemic
    mending [mɛndɪŋ] *painted* [peɪntɪd] *senses* [sɛnsɪz]
- 150 CVCCVC pseudowords
  nutvit [nʊtvɪt] *nisren* [nɪsrɪn] *tulsid* [tʊlsɪd]
- single male talker

Analysis: Confusion

1. Convert spelling to phonemes
2. For each SNR (0 or -5), Block (word or nonword), and position (C1, C2 etc.) make a confusion matrix
3. For each subject, calculate the mean word score (p_w) and phoneme score (p_p)

Analysis: J-factor

- The j-factor model provides a measure of context effects.
- The j-factor model assumes that phonemes are the basic unit of speech, and that phonemes are perceived independently (which has been shown to hold true most of the time).
- The probability of correctly identifying a given word (or nonword) can be calculated as the product of the probabilities of its constituent phonemes.

Analysis: J-factor

$$p_W = p_{C1} p_{V1} p_{C2} p_{C3} p_{V2} p_{C4} \quad (0)$$

$$p_W = p_p^j \quad (0)$$

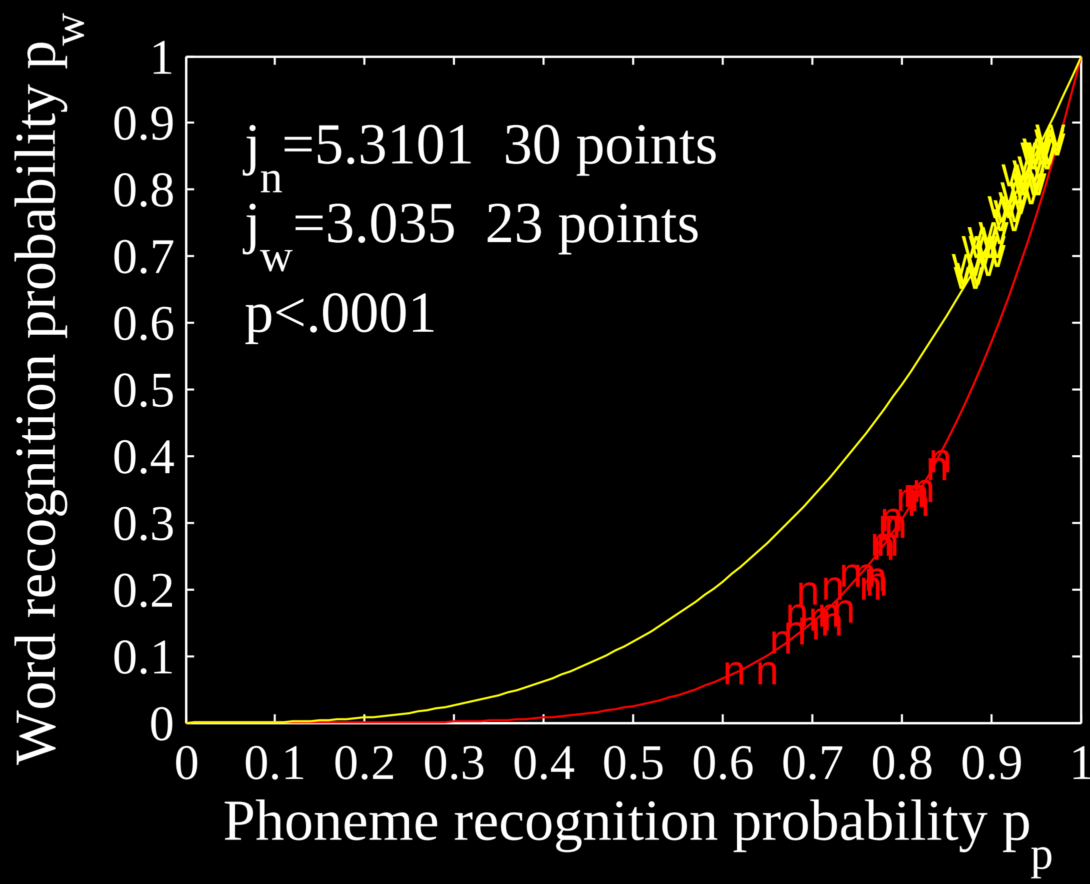
$$j = \frac{log(p_W)}{log(p_p)} \quad (0)$$

Predictions

- Nonwords – $j = 6$, which suggests that phonemes are being predicted independently of one another
- Words – $j < 6$, which suggests that lexical status is affecting perception.
- Frequency – As lexical frequency increases, j should decrease
- Density – As density increases, j should increase

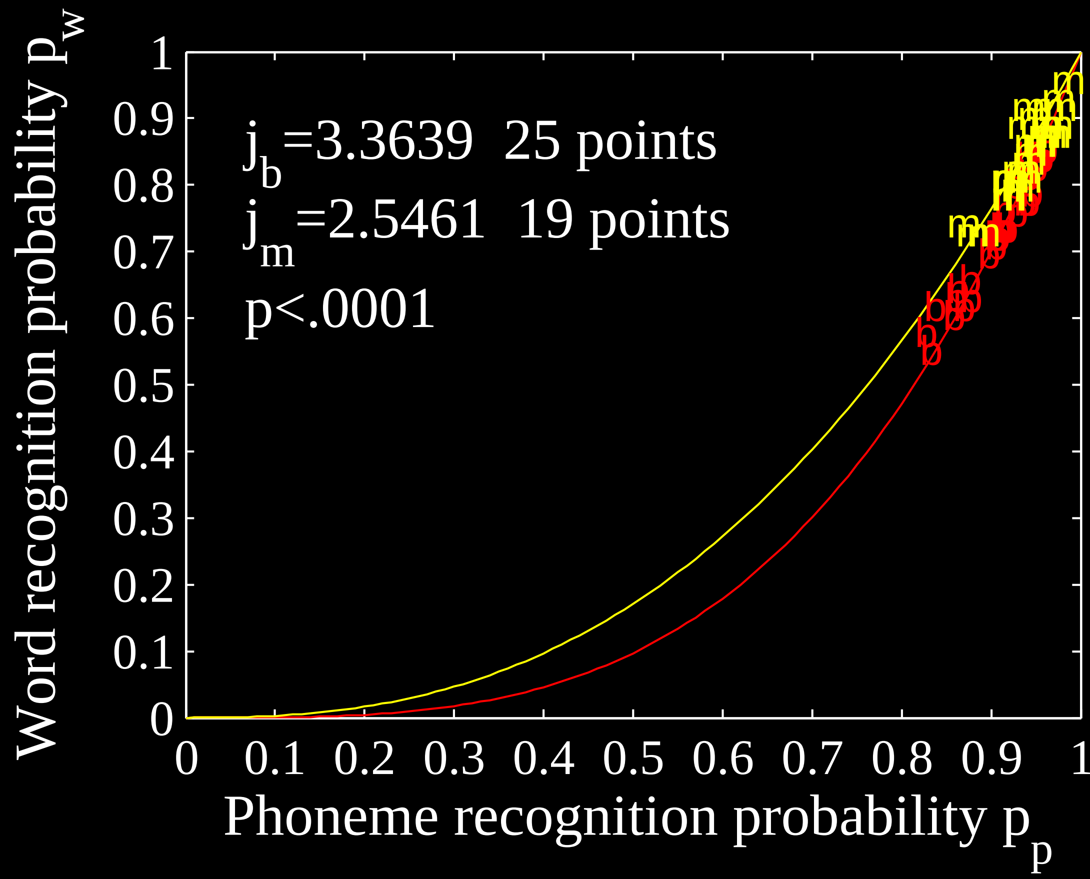
Results: Subject Analysis

- As expected, there is a significant difference in j between words and nonwords
- j for nonwords is slightly smaller than expected



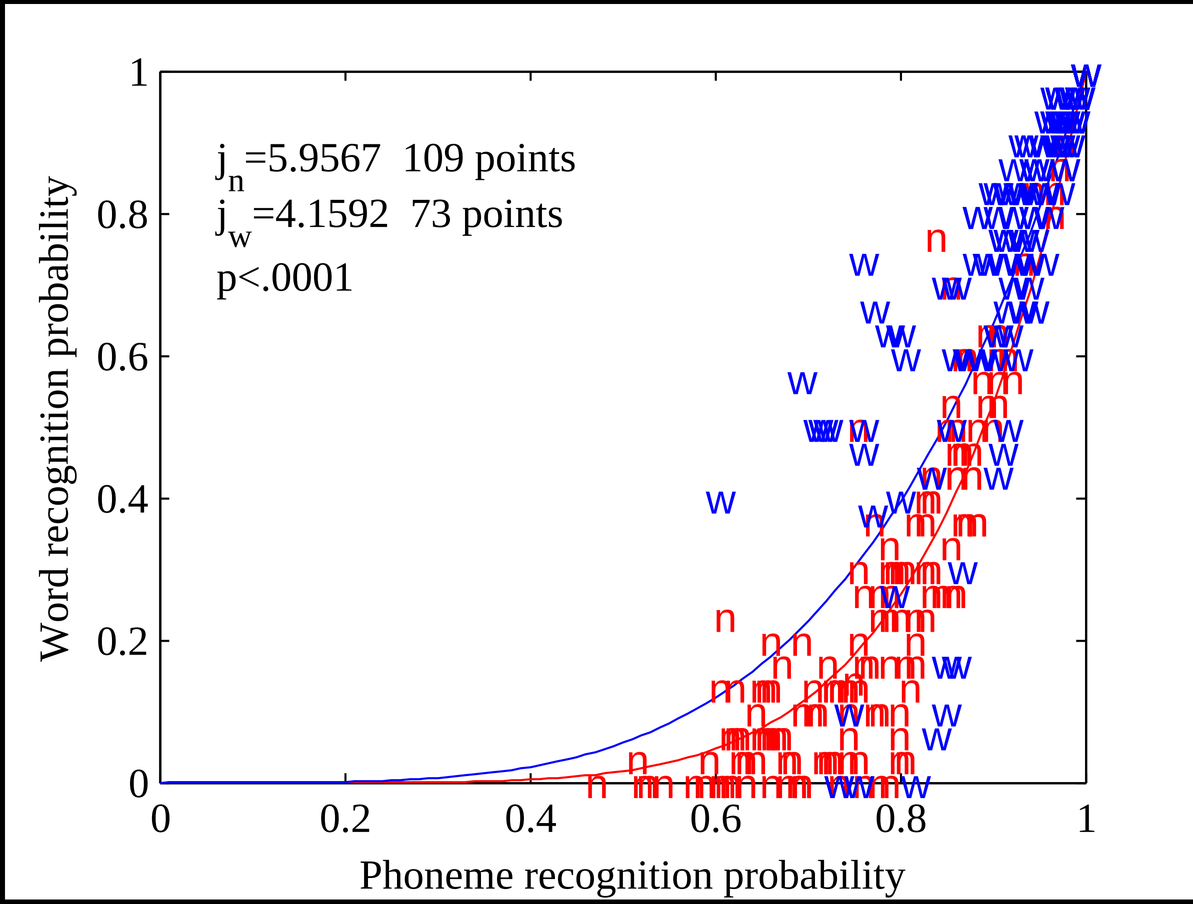
Results: Subject Analysis

- Monomorphemes and bimorphemes also differ significantly in *j*
- This indicates a greater context effect for monomorphemes than bimorphemes



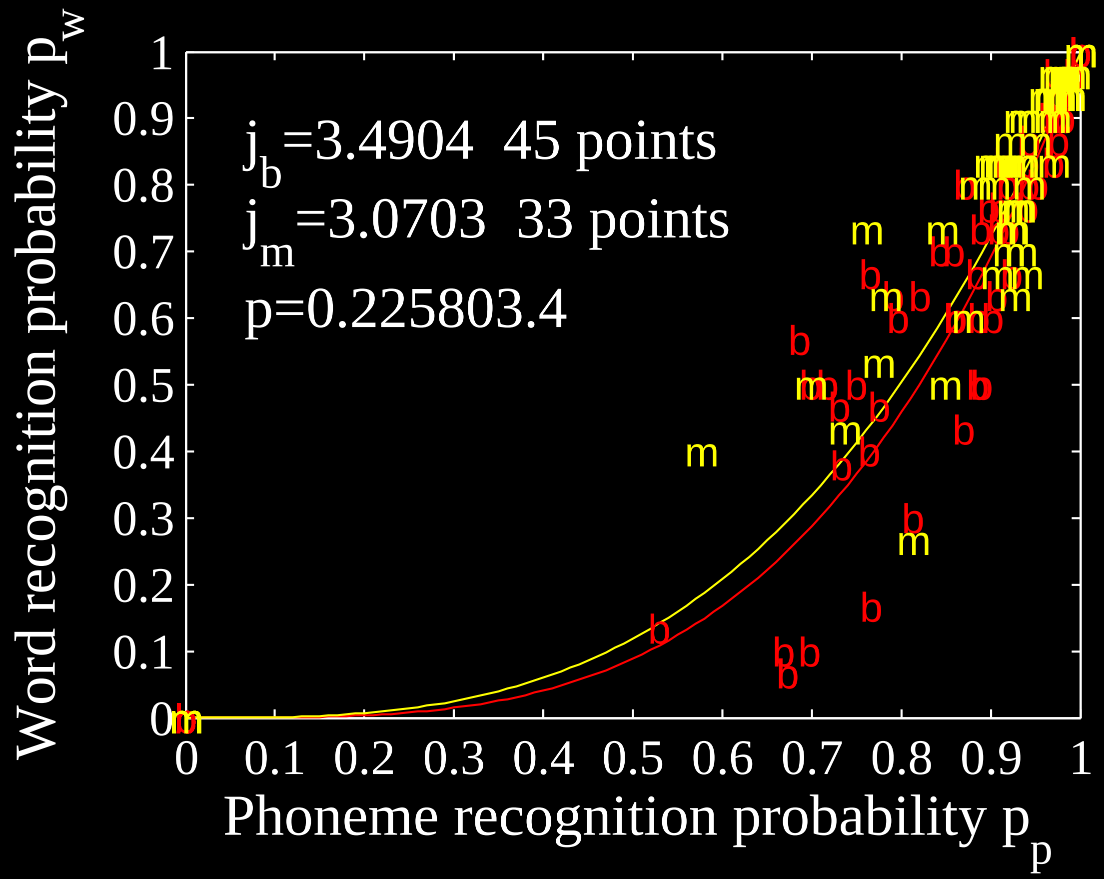
Results: Items Analysis

- The items analysis is consistent with the subject analysis
- There is more variation in the items analysis, since individual words cannot be phonemically balanced, as is the case for the subjects analysis



Results: Items Analysis

- The items analysis of bi- and monomorphemes is also consistent with the subjects analysis

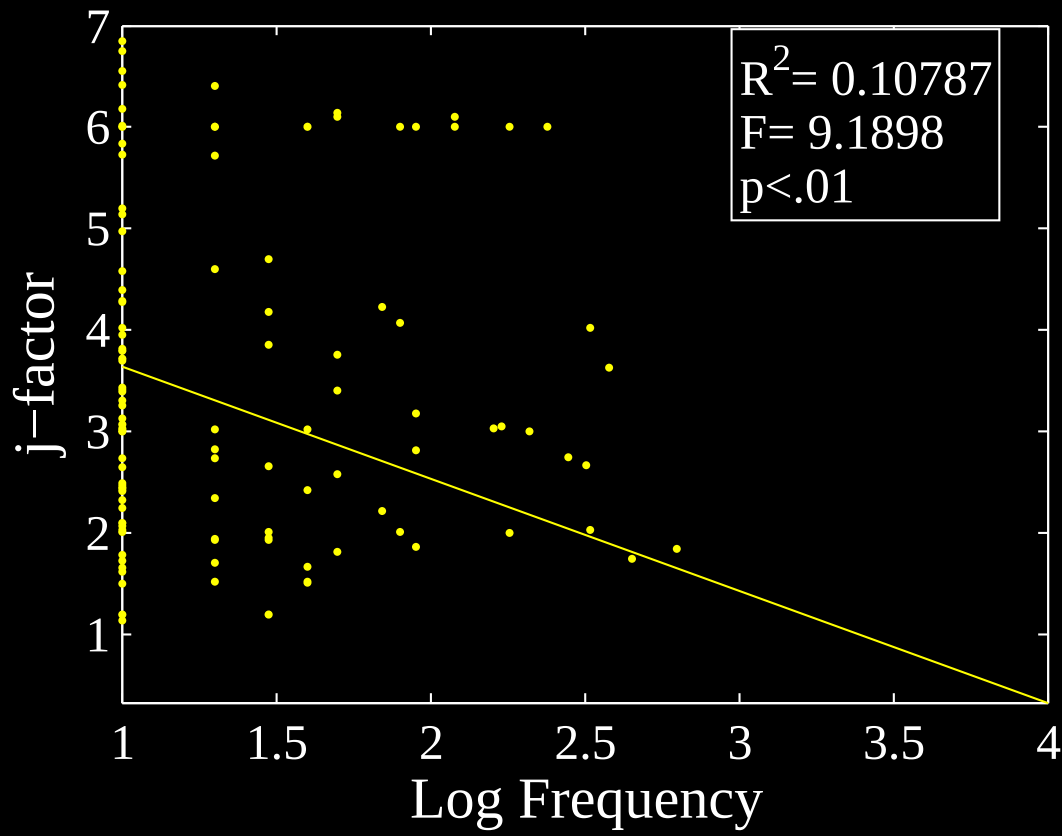


Results: J-factor summary

	mean	lower C.I.	upper C.I.
nonwords	5.31	5.18	5.44
words	3.035	2.91	3.16
bi	3.36	3.20	3.53
mono	2.55	2.31	2.78

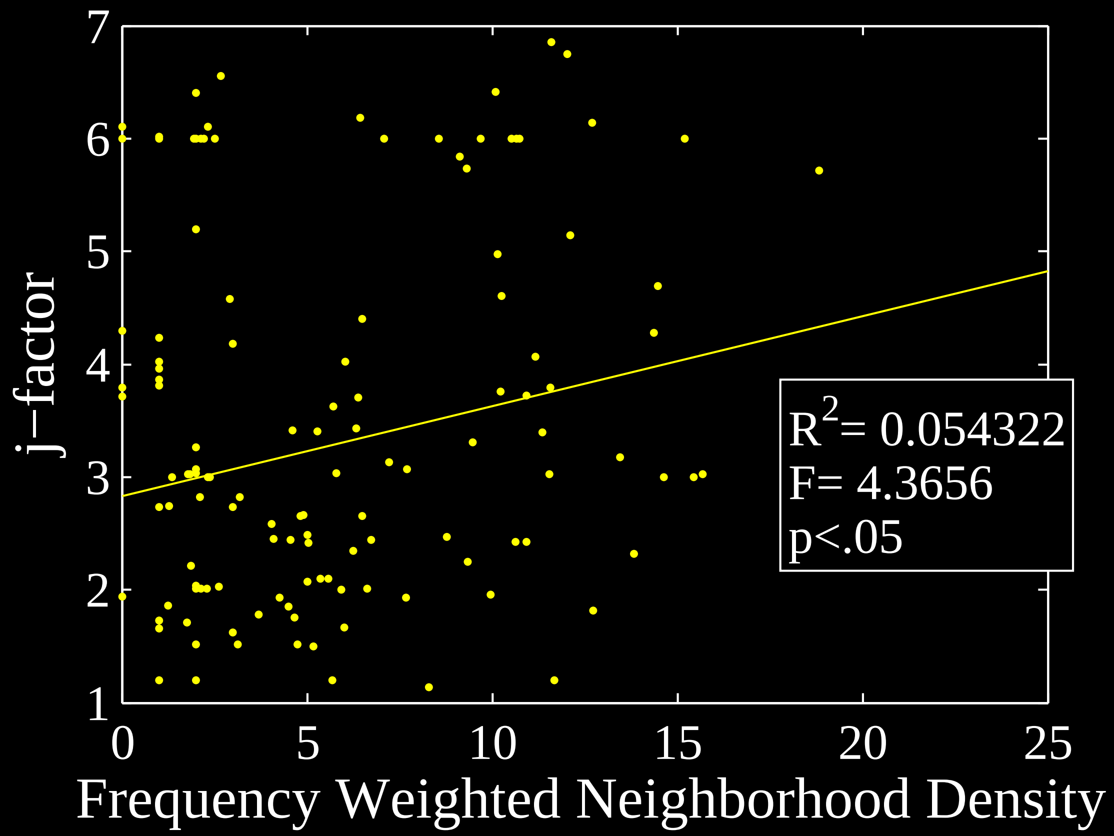
Results: Frequency

- Linear regression shows a significant correlation between Frequency and j-factor
- However, it only accounts for app. 10% of the variation



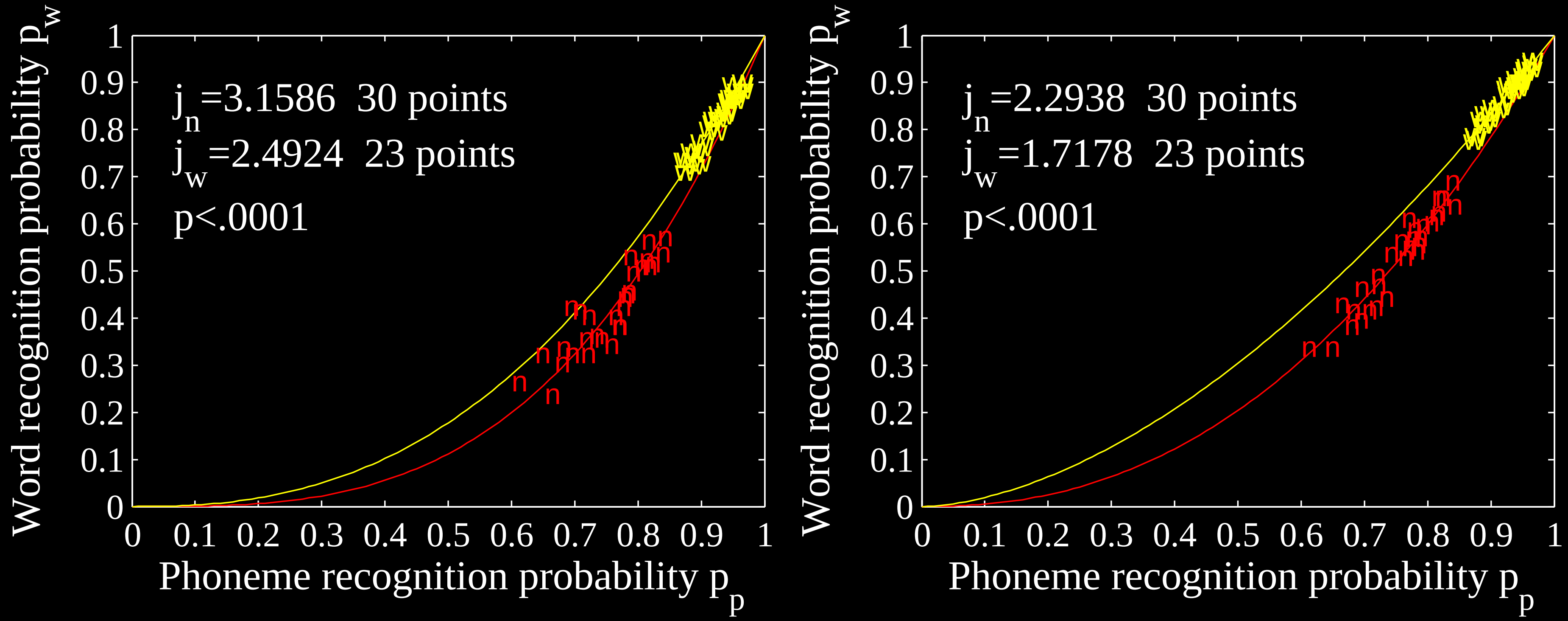
Results: Neighborhood Density

- Neighborhood density is also significant, but only accounts for 5% of the variation found
- The trend is in the right direction



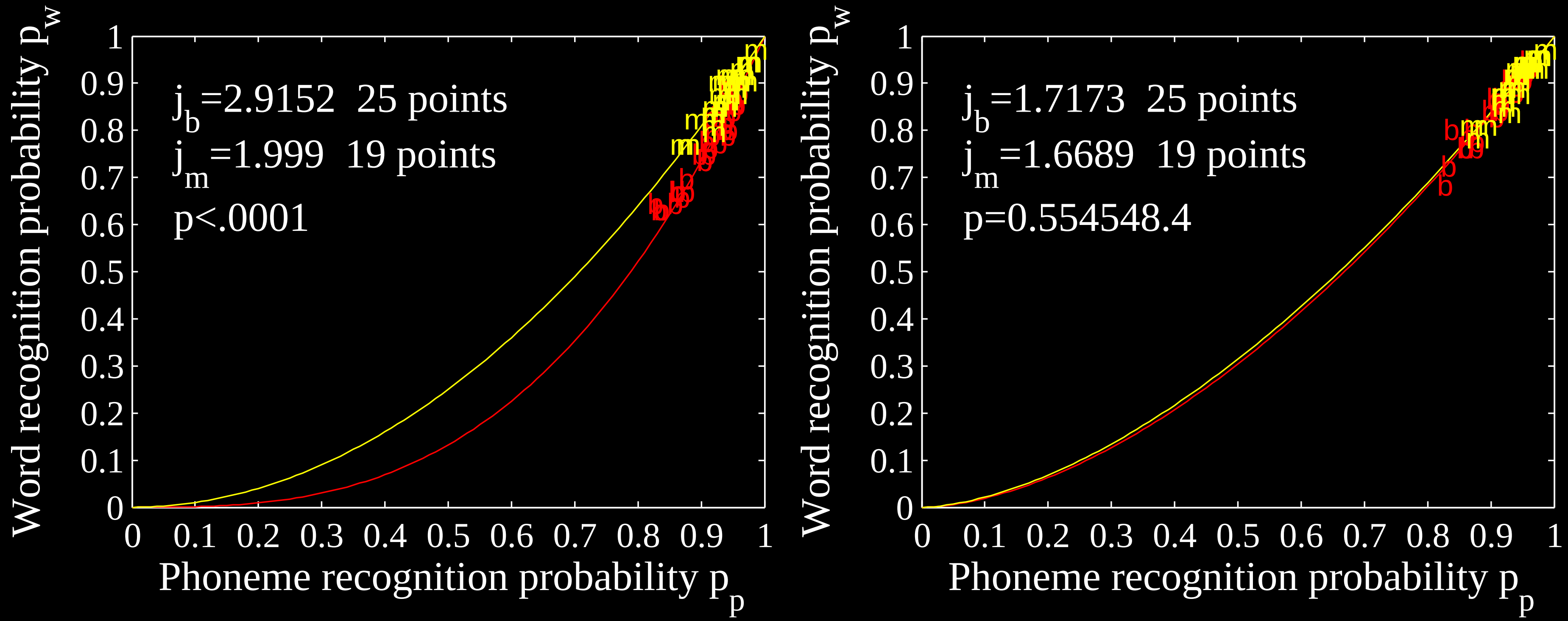
Discussion: Words and Nonwords

Why is *j* for nonwords less than 6?



Discussion: Mono- and Bimorphemes

Where does the difference between mono- and bimorphemes arise?



Discussion: J-factor syllable summary

	mean	lower C.I.	upper C.I.
		first syllable	
nonwords	3.15	2.98	3.26
words	2.49	2.30	2.54
bi	2.91	2.62	3.06
mono	2.00	1.79	2.12
		second syllable	
nonwords	2.29	2.19	2.33
words	1.72	1.64	1.77
bi	1.72	1.64	1.81
mono	1.74	1.52	1.75

Conclusions

- The j-factor results for CVCCVC words are mostly consistent with the previous results using CVC stimuli
- The difference in j of mono- and bimorphemes supports a combinatorial model of lexical access.

Future Research

- Do other languages exhibit a similar difference in mono- and bimorphemes?
- Specifically, will a more highly inflecting language such as German show an even greater difference between mono- and bimorphemes, and will it be in the same direction?

German Experiments

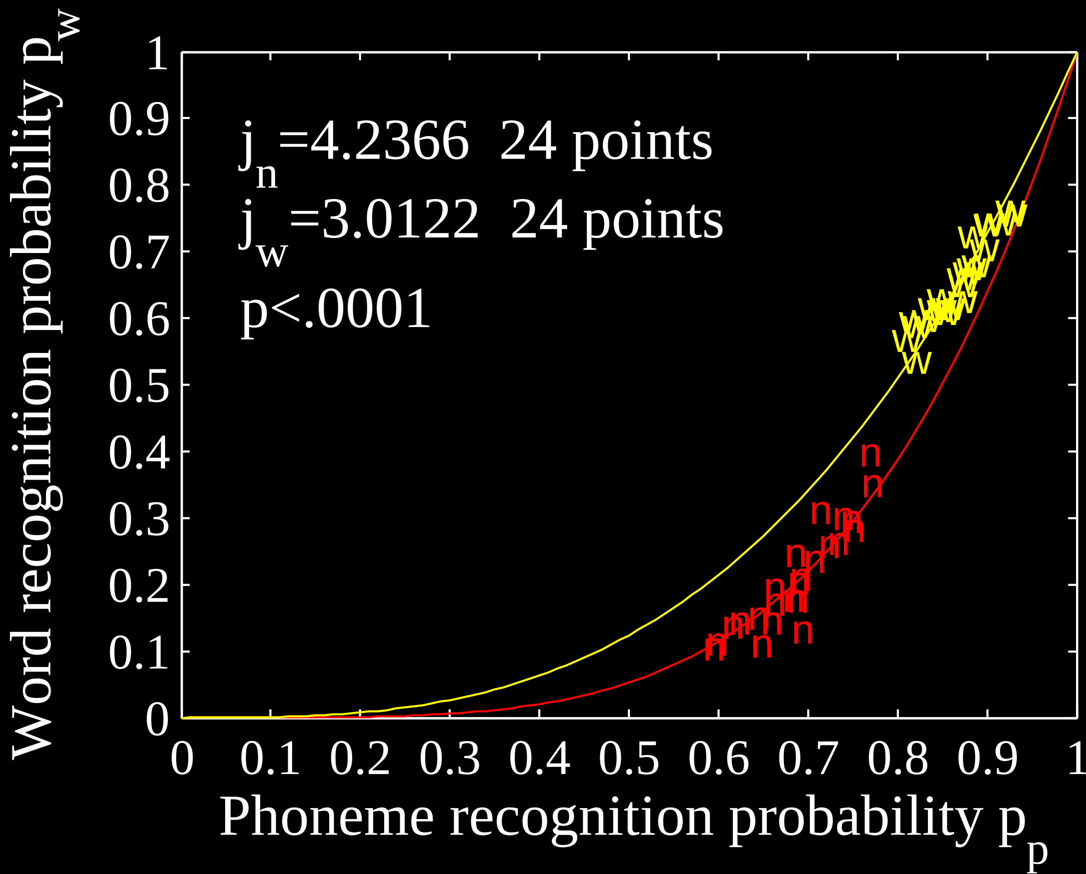
- Task is the same as in the first experiment
- 24 (so far) native Speakers of German took part
- S/Ns were 2 dB and 7 dB

German Experiments: Materials

- 150 CVCCVC words
  - 75 monomorphemic
    Laster [bændɪʤ] *dunkel* [dʊŋkəl] *hektik* [hɛktɪk]
  - 75 bimorphemic
    Feindes [faɪndəs] *bestem* [bɛstəm] *derber* [dɛrbər]
- 150 CVCCVC pseudowords
  nemschen [nɛmʃən] *tulker* [tʊlkər] *bomgech* [bɔmgəx]
- single male talker

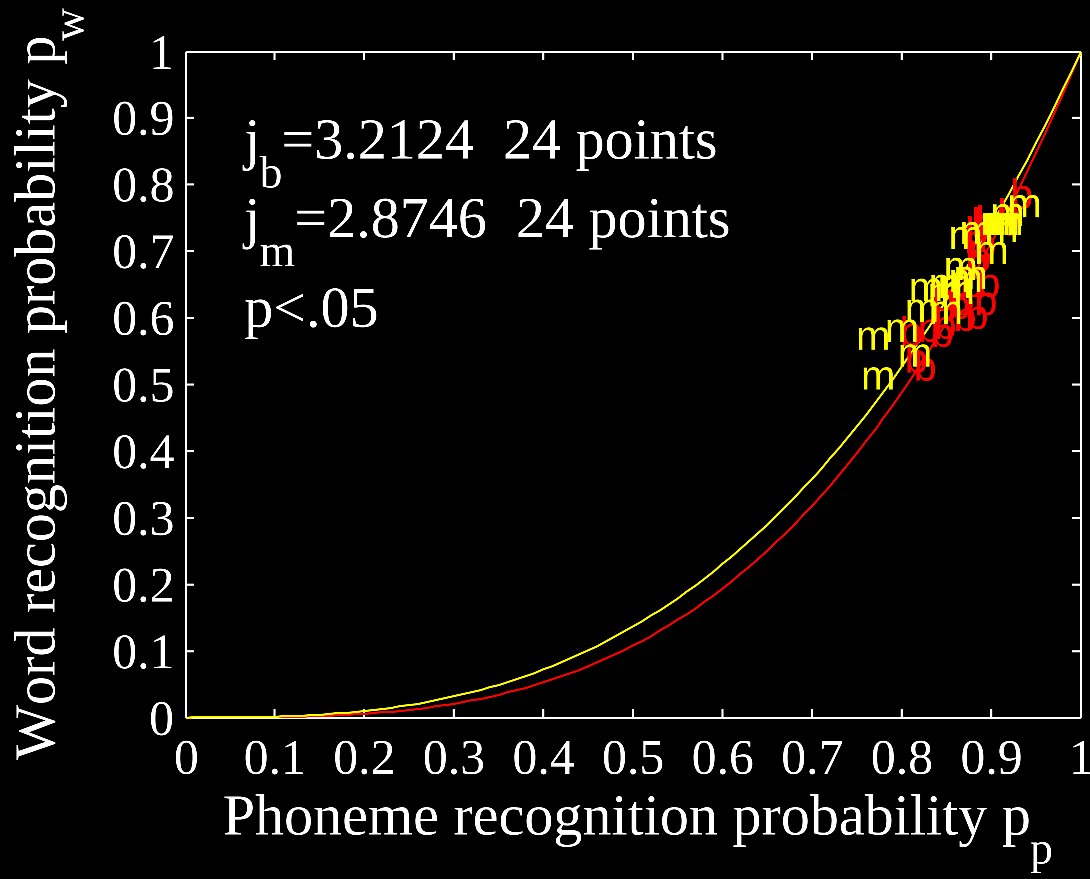
German Results: Subject Analysis

- As expected, there is a significant difference in *j* between words and nonwords
- *j* for nonwords is much smaller than expected



German Results: Subject Analysis

- As expected, there is a significant difference in j between monomorphemes and bimorphemes



German Results: Preliminary Conclusions

- Data from German shows similar pattern for words vs. nonwords and mono- vs. bimorphemes compared to the English data
- Difference in mono- and bimorphemes supports a combinatorial model of lexical access

Remaining Questions

- Why is j for German nonwords much lower than expected?
- Do the factor results for words suggest a bias for words as Nearey has suggested, or is the basic unit of speech perception for words larger than a phoneme?
- Can the difference between mono- and bimorphemes be explained by associative models?

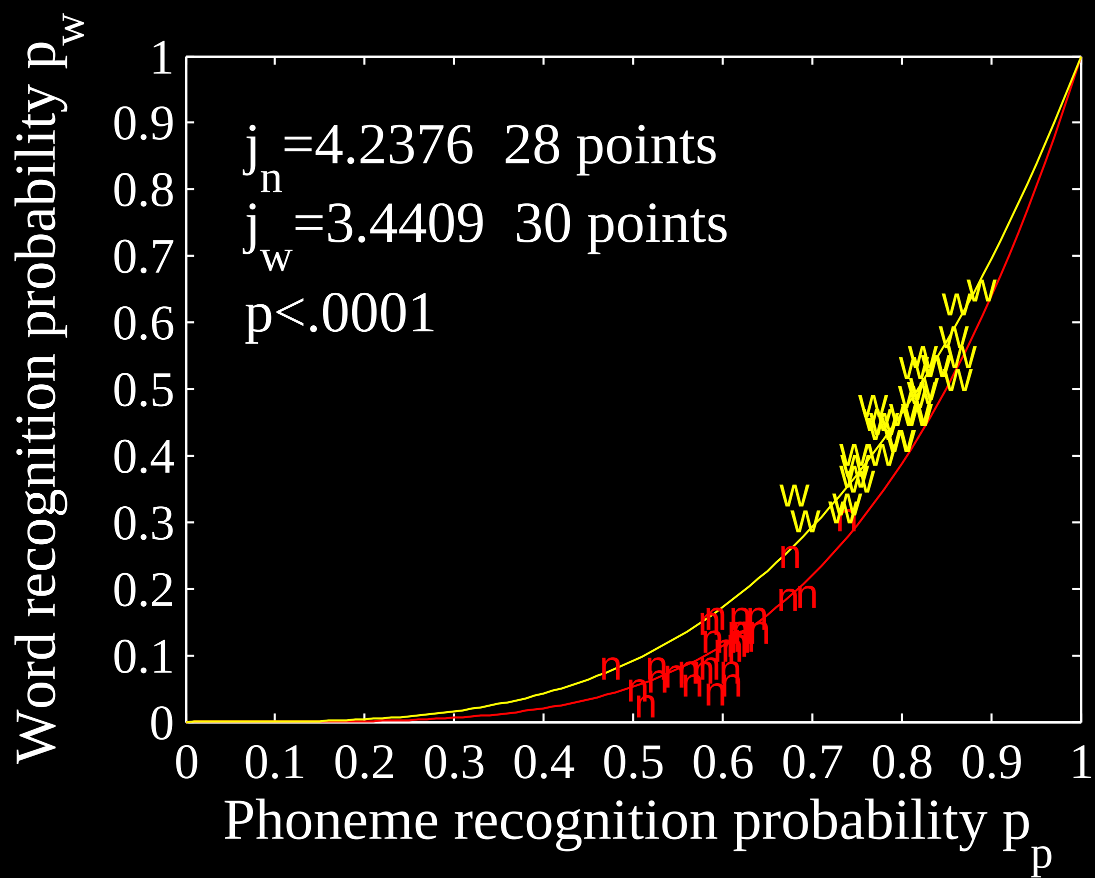
Thanks

Context Effects in Spoken Word Recognition of English CVCCVC words and nonsense words

Robert Felty
University of Michigan
robfelty@umich.edu

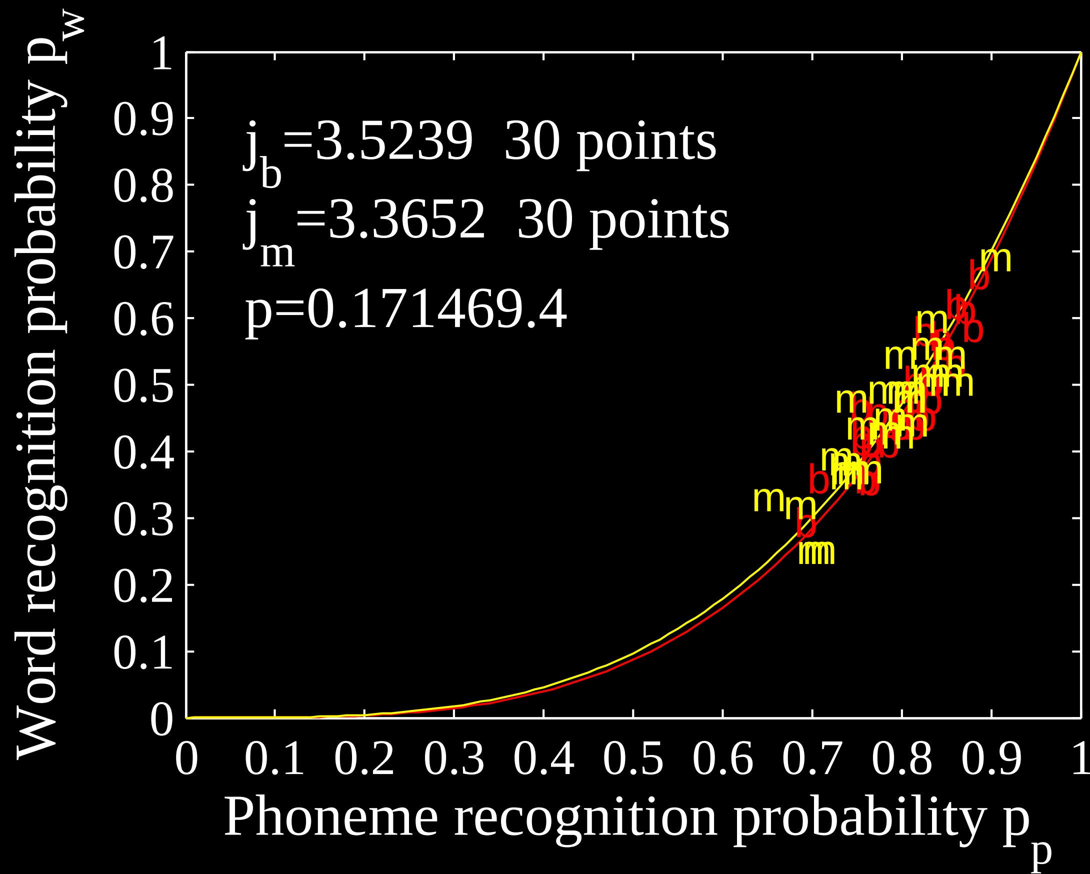
American / German Results: Word vs. Nonword

- As expected, there is a significant difference in j between words and nonwords
- j for nonwords is much smaller than expected



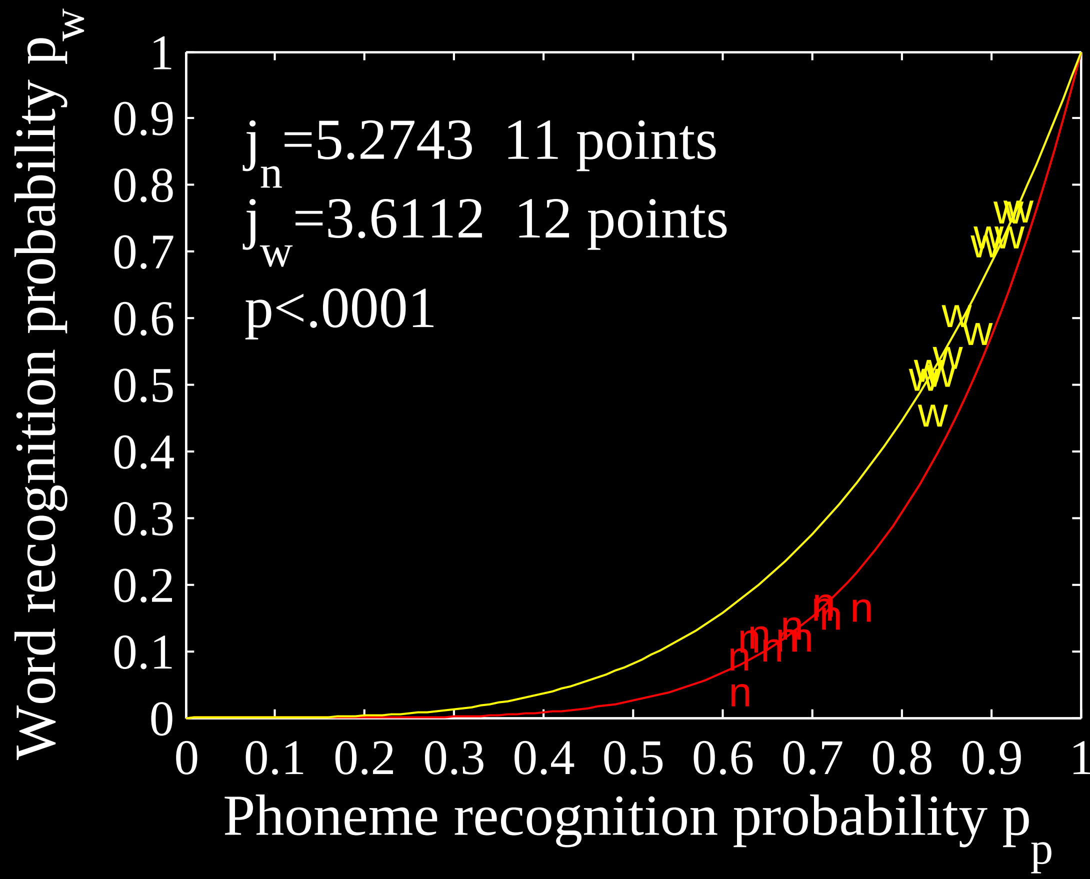
American / German Results: Mono vs. Bi

- No difference between mono- and bi-morphemes found



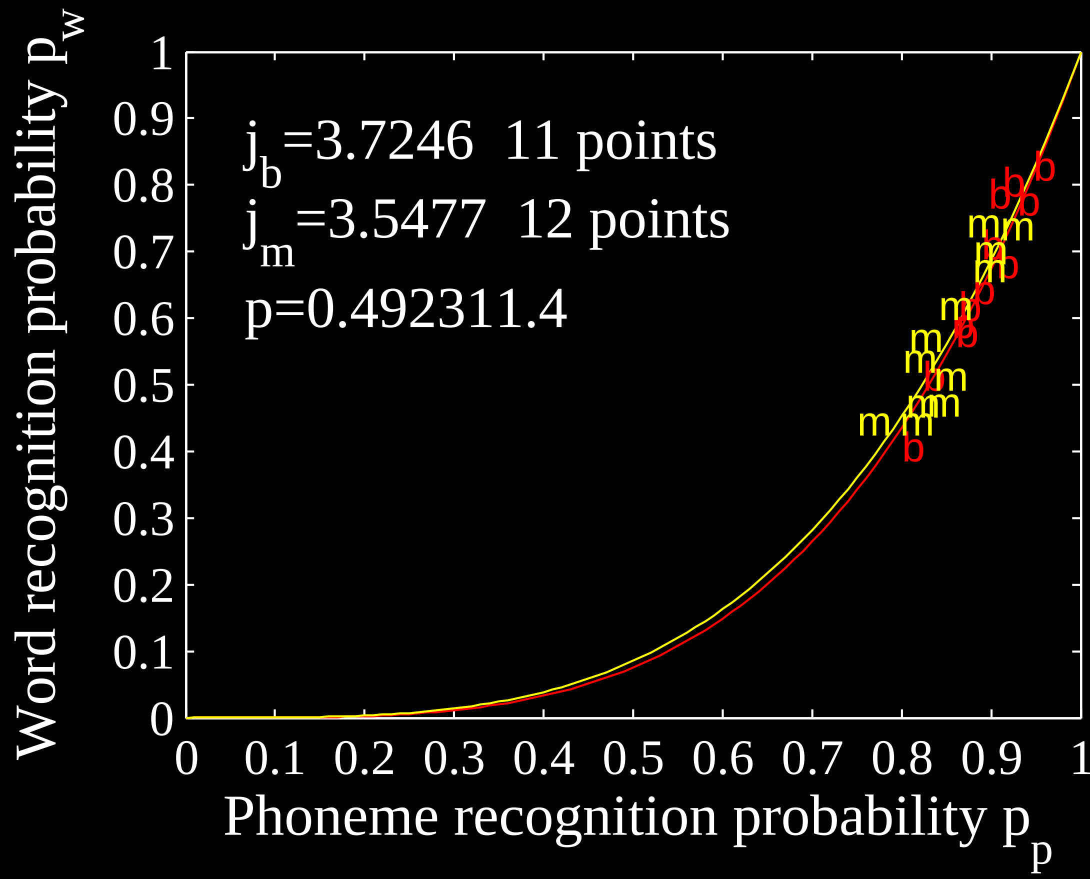
German / English Results: Word vs. Nonword

- As expected, there is a significant difference in *j* between words and nonwords



German / English Results: Bi vs. Mono

- No difference between bi and mono yet, but with more subjects it looks like there could be a small difference



Open Response Data: Model

How does one deal with open response data?

- give as much credit as possible
- be consistent

Open Response Data: Examples

- typos
  - metathesis typo *biulded* – scored as bɪldəd
  - letters next to each other on keyboard
- real words in non words *bahbone* – scored as babwʊn
- misspellings *concious* for *conscious*

p_p by position