Cross-linguistic differences in lexical access and spoken word recognition

Robert Felty robfelty@umich.edu
University of Michigan
March 9th, 2007

0.1 Acknowledgments

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

- José Benkí (co-chair, Communications Sciences and Disorders, Michigan State University)
- Patrice Beddor (co-chair, Linguistics)
- Andries Coetzee (Linguistics)
- 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.

1 Background

1.1 Research Goals

- Learn more about the role of morphology in the mental lexicon
- 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?
- Extend previous research using open response spoken word recognition to bisyllabic words
- Compare context effects across two phonologically similar, yet morphologically diverse languages

1.2 Morphology

- The study of Lexical Access seeks to determine how the mental lexicon affects language processing.
- The role of morphology in the lexicon is studied widely in lexical access research
- Results from cross-linguistic research suggest that morphology plays different roles in lexical access based on the type of morphological system of the language
- Two classes of models differ in their predictions of how morphologically complex words are stored in the lexicon and accessed.
- 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
  - Examples: Taft (1988); Taft and Forster (1975)

1.3 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.

Table 1 Example from Clahsen et al. (2001)

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

1.4 Qualitative Predictions

- A highly inflectional language (German) will show a greater effect of morphological complexity than a language with little inflectional morphology (English)
- Other context effects such as lexical frequency and neighborhood density will have a smaller effect on non-native listeners than native listeners, given that their lexicons are not as developed

2 Method

2.1 Task / Subjects

- Open Response Speech-In-Noise Task
  - Participants respond via keyboard input
- 2 different Signal to Noise Ratios (SNRs) used for each experiment
- signal dependent (but uncorrelated) noise (see Schroeder, 1968)
- Two separate experiments
  - Experiment 1 — 30 native speakers of English
  - Experiment 2 — 32 native speakers of German

2.2 English Materials

- 150 CVCCVC words
  - 74 monomorphemic *basket* /bæskɪt/ *compass* /kəmpəs/ *random* /ɹændəm/
  - 76 bimorphemic *mending* /mɛndɪŋ/ *painted* /peɪntɪd/ *senses* /sɛnsɪz/

- 150 CVCCVC nonwords *nutvit* /nʊtvɪt/ *nisren* /nɪsrɪn/ *tulsid* /tʊlsɪd/
- single male talker

2.3 German Materials

- 150 CVCCVC words
  - 75 monomorphemic *dunkel* /dʊŋkəl/ *selten* /zɛltən/ *hektik* /hɛktɪk/
  - 75 bimorphemic *Feindes* /faɪndəs/ *bestem* /bɛstəm/ *derber* /dɛrbər/
- 150 CVCCVC nonwords *nemschen* /nɛmʃən/ *mofkem* /mɔfkəm/ *bomgech* /bɔmɡəx/
- single male talker

3 Analysis

3.1 Confusion

1. Convert spelling to phonemes
2. For each SNR, 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)

3.2 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; see Fletcher, 1953; Allen, 1994).
- The probability of correctly identifying a given word (or nonword) can be calculated as the product of the probabilities of its constituent phonemes, as shown in equation 1.

$$p_w = p_{C1} p_{V1} p_{C2} p_{C3} p_{V2} p_{C4} \tag{1}$$

where p_w is the probability of correctly identifying a word (or nonword). Assuming that phonemes are perceived independently, (1) can be rewritten as:

$$p_w = p_p^j \tag{2}$$

where j is the number of phonemes, and p_p is the geometric mean of the probabilities of each constituent phoneme. Rewriting (2), the quantity j can be empirically determined from confusion matrices by:

$$j = \frac{log(p_w)}{log(p_p)} \tag{3}$$

3.2.1 Previous J-factor results

- 3 studies have used the j-factor model with CVC English stimuli (Boothroyd and Nittrouer, 1988; Olsen et al., 1997; Benkí, 2003)
- All have found $j_{nonword} \approx 3$ and $j_{word} \approx 2.5$
- 1 study using CVC Mandarin stimuli (Benkí et al., in preparation) did not find a difference between words and nonwords

3.3 Quantitative Predictions

- Nonwords — $j = 6$; interpretation is that phonemes are being predicted independently of one another
- Words — $j < 6$; interpretation is that lexical status is affecting perception.
- Morphology — $j_{bi} > j_{mono}$; interpretation is that monomorphemes have more context than bimorphemes
- Frequency — $j_{word} \propto \frac{1}{\text{frequency}}$; interpretation is that frequency provides a facilitatory effect
- Neighborhood density — $j_{word} \propto$ density; interpretation is that density provides an inhibitory effect

4 Experiment One Results — English listeners

4.1 English — Lexical Status

- As expected, there is a significant difference in j between words and nonwords (see Figure 1, page 5)
- j for nonwords is slightly smaller than expected

4.2 English — Morphology

- After removing confounds with lexical frequency and neighborhood density, no significant difference was found between monomorphemes and bimorphemes

4.3 English — Lexical Frequency

- Words were grouped into low and high frequency groups via median splits
- As predicted, high frequency words have a lower j, indicating a facilitatory effect of frequency

4.4 English — Neighborhood Density

- Words were also grouped into sparse and dense neighborhoods via median splits
- As predicted, an increase in density causes an inhibitory effect

5 Experiment Two Results — German listeners

5.1 Item Exclusion

- Initial results for German had much lower than expected j-scores
- Additional analysis revealed that this was due to stimuli containing post-vocalic /ʀ/ which frequently does not behave as an independent phoneme
- Results for lexical status and morphology shown here have excluded words containing post-vocalic /ʀ/
  94 nonwords and 79 words (36 monomorphemic and 43 bimorphemic)
- Lexical frequency and neighborhood density effects did not seem to be affected by this, so they are shown with the full set of stimuli

5.2 German — Lexical Status

- As predicted, j_{word} is significantly lower than $j_{nonword}$ (see Figure 2, page 5)
- j for nonwords is slightly smaller than expected

5.3 German — Morphology

- As predicted, j_{mono} was significantly lower than j_{bi}
- This indicates a greater context effect for monomorphemes than bimorphemes

5.4 German — Lexical Frequency

- Effects of lexical frequency were also significant
- However, the effect is opposite of that predicted — we find an inhibitory effect

5.5 German — Neighborhood Density

- Neighborhood density is also significant
- As predicted, an increase in density causes an inhibitory effect

6 Discussion

6.1 Summary of Results

Table 2 J-factor analysis summary

	Lexical Status	Morphology	Log wordform frequency	Log lemma frequency	phonological neighborhood density	phonetic neighborhood density
English	2.07***	0.09	0.51**	0.47**	-0.47**	-0.90***
German	1.45***	0.78***	-0.69***	-0.98***	-0.29*	-1.02***

***$p < .001$, **$p < .01$, *$p < .05$

6.2 Cross-linguistic effects

- One of the major differences found between the English and German results is the effect of morphology
- The interpretation for this is that German has a much richer inflectional morphology, and therefore morphology plays a larger role in the structure of the lexicon
- Similar cross-linguistic differences have been reported by Marslen-Wilson (2001).
- In comparing Polish, Arabic, English, and Chinese they have obtained different results in terms of how morphology is processed and represented in the lexicon.

Marslen-Wilson (2001) find that:

- In English, complex words such as *darkness* are represented by their constituent morphemes, and are combined during lexical access. English also exhibits stem-priming, e.g. the stem in *darkness* and *darkly* prime *dark*. This is not the case for semantically opaque words such as *department*, which does not prime *depart*.
- Polish also exhibits affix priming, e.g. *kotek/ogródek* 'a little cat' / 'a little garden' – the diminutive affix in the prime facilitates perception of the target and suffix interference (e.g. *pis-anie/pis-arz* 'writing'/'writer' – no facilitation is found in such pairs, despite facilitation of inflectional endings).
- Morphology seems to play an even larger role in Arabic, which has root priming even for semantically opaque words.
- Chinese has virtually no inflectional or derivational morphology
- Compounding is very active in Mandarin Chinese, and bimorphemic compounds account for up to 70% of all word forms in the language.
- However Marslen-Wilson and colleagues find no evidence for morphological decomposition in Mandarin compounds.
- Vannest et al. (2002) also find similarly various results in a comparison of English and Finnish derivational morphology.
- Research on Finnish inflectional morphology has shown support for combinatorial-like processing (e.g. Laine et al., 1999), Vannest et al.
- But they find less evidence for morphological decomposition with derivational morphology than for English.
- They hypothesize that words with derivational affixes are stored separately in Finnish in order to decrease the amount of morphological processing that the Finnish speaker must perform.

6.3 Interaction of Phonetics and Morphology

- It is possible that differences in mono- and bimorphemic stimuli could be partially due to acoustics or response bias.
- The final consonants in the bimorphemic stimuli were restricted to the phonemes /ʀ s m n/, which, along with /ə/ constitute all of the possible inflectional endings for nouns and adjectives in German.
- /m/ and /n/ are known to be highly confusable with one another.
- In addition, /n/ occurs as an inflectional ending much more frequently than /m/.
- In order to investigate this further, a Signal Detection Theory (SDT) analysis was carried out.
- SDT measures the sensitivity of distinguishing two stimuli, using the metric, d'.
- SDT also provides a measure of bias, c, which indicates whether one is more or less likely to respond with a particular phoneme.
  - Positive values of c indicate a bias towards a response;
  - negative values indicate a bias against a response.
- To carry out the SDT analysis, the original confusion matrices for each S/N were transformed into 2x2 submatrices. An SDT analysis was then applied to each submatrix.

Table 3 Signal Detection Theory analysis of /m/ and /n/ submatrix in final position. For this analysis /m/ is considered to be the target stimulus. Positive c' indicate a bias towards /n/.

1. in the absence of lexical context effects (nonword condition), /m/ and /n/ are highly confusable, with a small bias towards /n/
2. /m/ and /n/ are perceived as most distinct in the monomorphemic condition,
3. bias towards /n/ is greatest in the bimorphemic case.

	d'	c
Nonwords		
lower S/N (2 dB)	-0.182	0.555
higher S/N (7 dB)	0.664	0.743
Bimorphemes		
lower S/N (2 dB)	1.616	0.984
higher S/N (7 dB)	1.913	0.556
Monomorphemes		
lower S/N (2 dB)	3.514	0.239
higher S/N (7 dB)	4.733	-0.060

6.4 Conclusions

- The j-factor results for CVCCVC words are mostly consistent with the previous results using CVC stimuli (Boothroyd and Nittrouer, 1988; Olsen et al., 1997; Benkí, 2003)
- One striking new result is that j_{word} does not scale linearly with word length
- The influence of morphology on spoken word recognition is language dependent
- The processing differences between mono- and bimorphemic found in this study present a challenge to theories of lexical access which assume whole word storage.
- Listeners are particularly sensitive to lexico-statistical information when presented with highly confusable stimuli

6.5 Future Research

- Further investigate effects of word length on spoken word recognition using stimuli of a variety of lengths
- Determine the time course of these effects using speech-in-noise tasks which also incorporate a measure of time course (either behavioral or neurological)

References

Allen, Jont. 1994. How do humans process and recognize speech?, *IEEE Transactions of Speech and Audio Processing*, 2(4), 567–577.

Benkí, José. 2003. Quantitative evaluation of lexical status, word frequency and neighborhood density as context effects in spoken word recognition, *Journal of the Acoustical Society of America*, 113(3), 1689–1705.

Benkí, José, J. Myers, and Terrance Nearey. in preparation. Lexical frequency effects in Mandarin.

Boothroyd, Arthur and Susan Nittrouer. 1988. Mathematical treatment of context effects in phoneme and word recognition, *Journal of the Acoustical Society of America*, 84, 101–114.

Clahsen, Harald, Sonja Isenbeiss, Meike Hadler, and Ingrid Sonnenstuhl. 2001. The mental representations of inflected words: an experimental study of adjectives and verbs in German, *Language*, 77(3), 510–543.

Fletcher, Harvey. 1953. *Speech and Hearing in Communication*, New York: Krieger.

Laine, M, S. Vainio, and J. Hyönä. 1999. Lexical access routes to nouns in a morphologically rich language, *Journal of Memory and Language*, 40, 109–135.

Marslen-Wilson, William. 2001. Access to lexical representations: Cross-linguistic issues, *Language and Cognitive Processes*, 16(5/6), 699–708.

Olsen, Wayne, Dianne Van Tasell, and Charles Speaks. 1997. Phoneme and word recognition for words in isolation and sentences, *Ear and Hearing*, 18(3), 175–188.

Schroeder, M. 1968. Reference signal for signal quality studies, *Journal of the Acoustical Society of America*, 44, 1735–1736.

Taft, Marcus. 1988. A morphological decomposition model of lexical representation, *Linguistics*, 26, 657–667.

Taft, Marcus and Kenneth Forster. 1975. Lexical storage and retrieval of prefixed words, *Journal of Verbal Learning and Verbal Behavior*, 14, 638–647.

Vannest, Jennifer, Raymond Bertram, Juhani Järvikivi, and Jussi Niemi. 2002. Counterintuitive cross-linguistic differences: More morphological computation in English than in Finnish, *Journal of Psycholinguistic Research*, 31, 83–106.

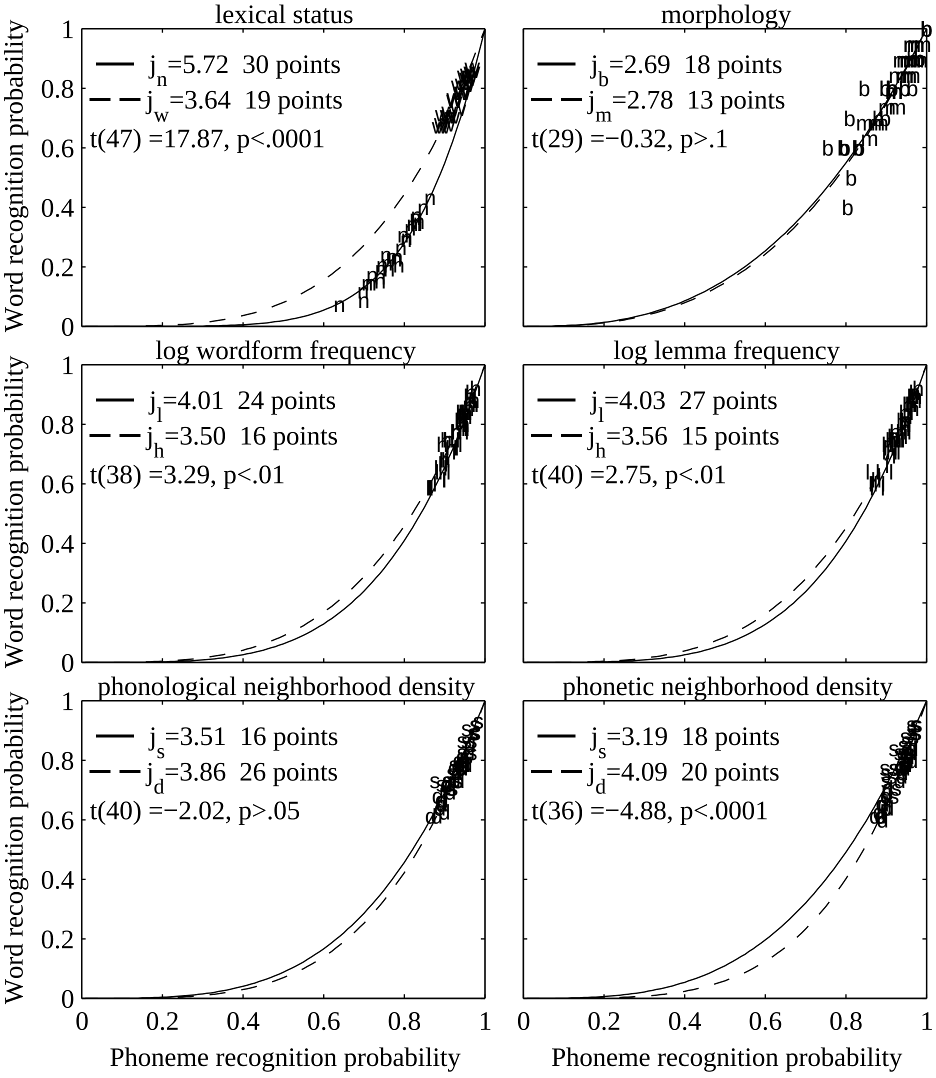

Figure 1: Experiment 1 (English listeners) J-factor Analysis by subject — Each plot compares two subsets of results from the subject analysis. Curves represent $y = x^j$. Frequency and neighborhood density plots show only word results. P-values given are from 2-sample t-tests; before computing the statistics, all points lying in the floor or ceiling ranges ($>$ .95 or $<$ .05) were removed, but are still shown on the plot.

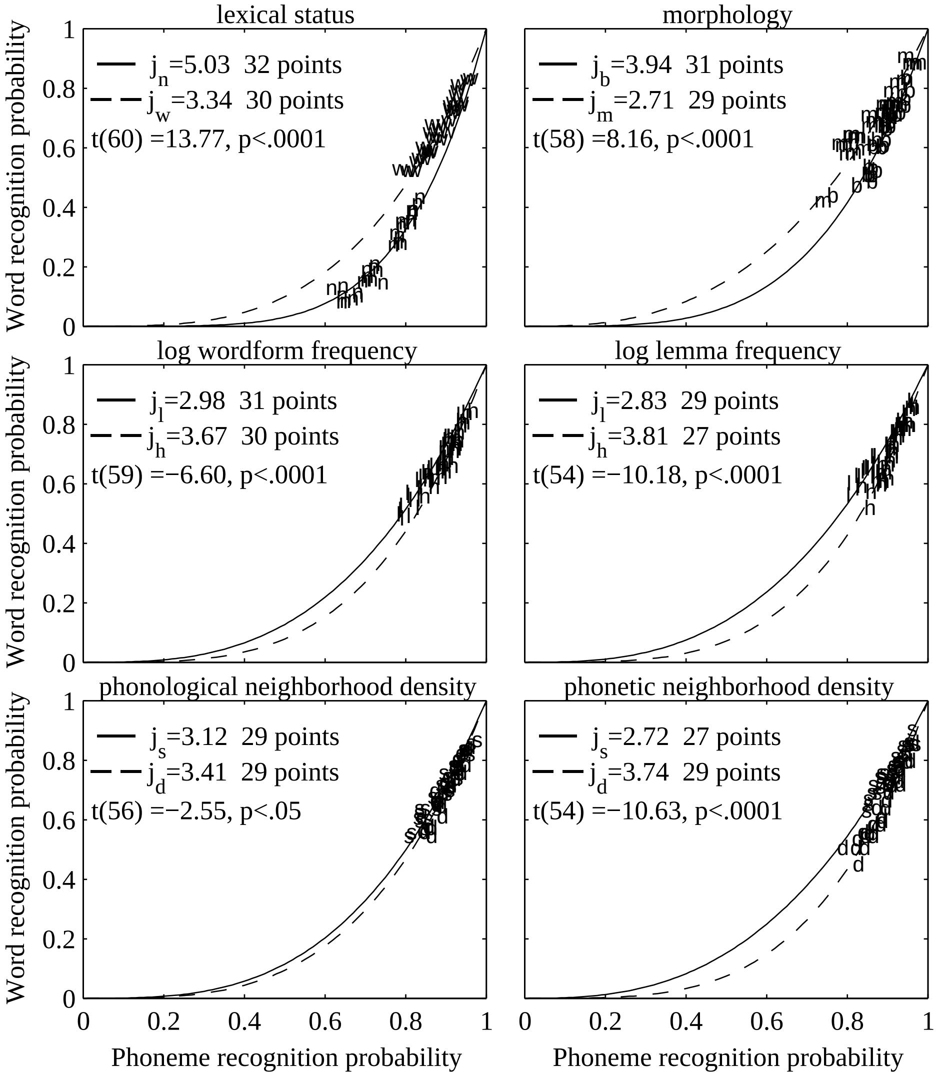

Figure 2: Experiment 2 (German listeners) J-factor Analysis by subject — Each plot compares two subsets of results from the subject analysis. Curves represent $y = x^j$. The frequency and density plots display only word results. P-values given are from 2-sample t-tests; before computing the statistics, all points lying in the floor or ceiling ranges ($>$ .95 or $<$ .05) were removed, but are still shown on the plot.

7 Sample Confusion Matrices

Table 4 English — V1 nonwords S/N = -5 dB. Numbers given are percentages. Total number of presentations for each phoneme is given in the final column.

	i	ɪ	eɪ	ɛ	æ	oʊ	ɑ	ə	ɚ	ɔɪ	aʊ	aɪ	null	other	Total
i	**79**	7		7									7		14
ɪ	5	**79**		9		1		1				3	3		392
eɪ		9	**64**	11	9			4					2	2	56
ɛ		8		**85**	2			2					3		518
æ				27	**62**		4	1			3		3		350
oʊ		1	1			**34**	25	23	1	1	3		4	7	182
ɑ					16	2	**46**	25			2		1	8	252
ə		1		8		12	10	**64**			1		3	2	196
ɔɪ										**64**	14	7	14		14
aʊ				20	17	2	2	6			**46**		5	1	84
aɪ		12		14	12			2			5	**52**		2	42

mean p_p= 62
min p(oʊ)= 34
max p(ɛ)= 85

Table 5 German — V1 nonwords S/N = 2 dB. Numbers given are percentages. Total number of presentations for each phoneme is given in the final column.

	i	ɪ	y	ʏ	u	ʊ	e	ɛ	œ	ɔ	a	ɔɪ	aɪ	ɔl	null	other	Total
ɪ	7	**64**		5	1	4	7	6	1	3					1	2	560
ʊ		5	1	4	2	**73**	2	1	1	7					1	3	480
ɛ		4				2	1	**79**	4	4	2	1			1	3	400
œ		3		5		2	1	43	**22**	19		3			1	2	176
ɔ								2	1	**85**	3	6	1		1	3	384
a								1	1	5	**87**	1	4			1	288
ɔɪ				1				9	9	16	24	**33**	4	1		4	80
aɪ											28		**72**				32

mean p_p= 64
min p(œ)= 22
max p(a)= 87

Table 6 English — C4 nonwords S/N = -5 dB. Numbers given are percentages. Total number of presentations for each phoneme is given in the final column.

	d	g	ʤ	t	k	ʧ	f	s	ʃ	v	z	l	m	n	ŋ	nd	rd	null	other	Total
d	**93**			2										2				4		308
ʤ		1	**90**															8		84
t	17			**77**	3			2										2		294
k				4	**94**													2		308
s								**97**			2							1		266
ʃ									**93**									7		14
v	7			4				7		**39**	25							14	4	28
z	1		2					37			**54**							6		224
m													**73**	25				2		238
n	1												6	**85**	1	5		3		182
ŋ	1	1											3	7	**82**			5		154

mean p_p= 80
min p(v)= 39
max p(s)= 97

Table 7 German — C4 nonwords S/N = 2 dB. Numbers given are percentages. Total number of presentations for each phoneme is given in the final column.

	p	t	k	s	x	l	ʀ	m	n	ŋ	null	other	Total
k	3	4	**86**		5						1		240
s	1			**97**							1	1	496
x			4		**91**						1	4	128
l	1					**72**	6				19		432
ʀ	2					1	**78**				12	7	624
m	3					3	2	**21**	60	5	3	3	240
n	1					2	2	27	**57**	3	3	5	240

mean p_p= 72
min p(m)= 21
max p(s)= 97