Cross-linguistic differences in lexical access and spoken word recognition

Robert Felty
University of Michigan
robfelty@umich.edu

March 9th, 2007
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.
Background
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.
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)
Using a Lexical Decision task, and a Cross-modal Priming task, Clahsen et al. (2001) found a difference in processing of German inflected adjectives.

Example from Clahsen et al. (2001)

<table>
<thead>
<tr>
<th>-m dominant adjectives</th>
<th>-s dominant adjectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stem form</td>
<td>-m</td>
</tr>
<tr>
<td>ruhig</td>
<td>838</td>
</tr>
</tbody>
</table>
Using a Lexical Decision task, and a Cross-modal Priming task, Clahsen et al. (2001) found a difference in processing of German inflected adjectives.

Example from Clahsen et al. (2001)

<table>
<thead>
<tr>
<th>-m dominant adjectives</th>
<th>-s dominant adjectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stem form</td>
<td>-m</td>
</tr>
<tr>
<td>ruhig</td>
<td>838</td>
</tr>
</tbody>
</table>
Using a Lexical Decision task, and a Cross-modal Priming task, Clahsen et al. (2001) found a difference in processing of German inflected adjectives.

Example from Clahsen et al. (2001)

<table>
<thead>
<tr>
<th>-m dominant adjectives</th>
<th>-s dominant adjectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stem form</td>
<td>-m</td>
</tr>
<tr>
<td>ruhig</td>
<td>838</td>
</tr>
</tbody>
</table>
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.
Method
## Task / Subjects

**Open Response Speech-In-Noise Task**
- Participants respond via keyboard input

**English Materials**
- 2 different Signal to Noise Ratios (SNRs) used for each experiment
- Signal dependent (but uncorrelated) noise (see Schroeder, 1968)

**German Materials**
- Two separate experiments
  - Experiment 1 — 30 native speakers of English
  - Experiment 2 — 32 native speakers of German
English Materials

- 150 CVCCVC words
  - 74 monomorphemic:
    - basket /bæskɪt/
    - compass /kəmpəs/
    - random /rændəm/
  - 76 bimorphemic:
    - mending /ˈmɛndoʊŋ/
    - painted /ˈpeɪntɪd/
    - senses /ˈsɛnsɪz/

- 150 CVCCVC nonwords:
  - nutvit /nʌtvɪt/
  - nisren /ˈnɪsrɛn/
  - tulsid /ˈtʊlsɪd/

- single male talker
German Materials

150 CVCCVC words

- 75 monomorphemic:
  - dunkel /dɔŋkəl/
  - selten /zɛltən/
  - hektik /hɛktɪk/

- 75 bimorphemic:
  - Feindes /faɪndəs/
  - bestem /bɛstəm/
  - derber /dɛrber/

150 CVCCVC nonwords:

- nemschen /nɛmsʃən/
- mofkem /mɔfkəm/
- bomgech /bɔmgɛx/

Single male talker
Analysis
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 \( p_p \)
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.
## J-factor

### Background

The J-factor is a measure used in computational linguistics to evaluate the performance of machine translation systems. It is calculated based on the confusion matrix of a system.

### Method

The J-factor predictions are calculated as follows:

1. \( p_w = p_{C1}p_{V1}p_{C2}p_{C3}p_{V2}p_{C4} \)

2. \( p_w = p^j_p \)

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

### Analysis

- **English Results**
- **German Results**

### Discussion

Robert Felty

Cross-linguistic differences . . . – 14
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 $j_{\text{nonword}} \approx 3$ and $j_{\text{word}} \approx 2.5$. 1 study using CVC Mandarin stimuli (Benkí et al., in preparation) did not find a difference between words and nonwords.
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 — $j_{word} \propto \text{density}$; interpretation is that density provides an inhibitory effect.
Quantitative Predictions

Hypothetical Results

Word recognition probability vs. Phoneme recognition probability

\[ j_n = 5.9995 \quad 5 \text{ points} \]
\[ j_w = 1 \quad 5 \text{ points} \]
\[ p < .0001 \]

Robert Felty
Experiment One Results

English Results

- English listeners

German Results

Discussion

Robert Felty

Cross-linguistic differences . . . – 16
As expected, there is a significant difference in $j$ between words and nonwords. $j$ for nonwords is slightly smaller than expected.

Phoneme recognition probability

Word recognition probability

$- j_n = 5.72 \quad 30 \text{ points}$

$- j_w = 3.64 \quad 19 \text{ points}$

$t(47) = 17.87, p < .0001$
After removing confounds with lexical frequency and neighborhood density, no significant difference was found between monomorphemes and bimorphemes.
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.

$$j_l = 4.01 \quad 24 \text{ points}$$

$$j_h = 3.50 \quad 16 \text{ points}$$

$$t(38) = 3.29, p < .01$$
Words were also grouped into sparse and dense neighborhoods via median splits. As predicted, an increase in density causes an inhibitory effect.

- $j_s = 3.19$, 18 points
- $j_d = 4.09$, 20 points

$t(36) = -4.88$, $p < .0001$
Experiment Two Results

German listeners
Initial results for German had much lower than expected j-scores.

Additional analysis revealed that this was due to stimuli containing post-vocalic /\ö/ which frequently does not behave as an independent phoneme.

Results for lexical status and morphology shown here have excluded words containing post-vocalic /\ö/.

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.
German — Lexical Status

- As predicted, $j_{word}$ is significantly lower than $j_{nonword}$
- $j$ for nonwords is slightly smaller than expected

Phoneme recognition probability

Word recognition probability

$t(60) = 13.77, p < .0001$

Robert Felty

Cross-linguistic differences ... – 23
As predicted, $j_{mono}$ was significantly lower than $j_{bi}$. This indicates a greater context effect for monomorphemes than bimorphemes.

\[
\text{Phoneme recognition probability} \quad \text{Word recognition probability}
\]

- $j_b = 3.94$ 31 points
- $j_m = 2.71$ 29 points

$t(58) = 8.16, p < .0001$
Effects of lexical frequency were also significant. However, the effect is opposite of that predicted — we find an inhibitory effect.

Robert Felty

Cross-linguistic differences ... – 25
Neighborhood density is also significant

As predicted, an increase in density causes an inhibitory effect

\[ t(54) = -10.63, p < .0001 \]
Discussion

Robert Felty

Cross-linguistic differences . . . – 27
## Summary of Results

J-factor analysis summary

<table>
<thead>
<tr>
<th>Lexical Status</th>
<th>Morphology wordform frequency</th>
<th>Log lemma frequency</th>
<th>phonological neighborhood density</th>
<th>phonetic neighborhood density</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>2.07*** 0.09</td>
<td>0.51**</td>
<td>-0.47**</td>
<td>-0.90***</td>
</tr>
<tr>
<td>German</td>
<td>1.45*** 0.78***</td>
<td>-0.69***</td>
<td>-0.98***</td>
<td>-0.29*  -1.02***</td>
</tr>
</tbody>
</table>

***p < .001, **p < .01, *p < .05
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.
Cross-linguistic effects

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.
Cross-linguistic effects

- 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.
Cross-linguistic effects

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.
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 /r 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.
Interaction of Phonetics and Morphology

- 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.
Interaction of Phonetics and Morphology

1. in the absence of lexical context effects (nonword condition), /m/ and /n/ are highly confus-able, 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.

<table>
<thead>
<tr>
<th></th>
<th>(d')</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nonwords</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lower S/N (2 dB)</td>
<td>-0.182</td>
<td>0.555</td>
</tr>
<tr>
<td>higher S/N (7 dB)</td>
<td>0.664</td>
<td>0.743</td>
</tr>
<tr>
<td><strong>Bimorphemes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lower S/N (2 dB)</td>
<td>1.616</td>
<td>0.984</td>
</tr>
<tr>
<td>higher S/N (7 dB)</td>
<td>1.913</td>
<td>0.556</td>
</tr>
<tr>
<td><strong>Monomorphemes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lower S/N (2 dB)</td>
<td>3.514</td>
<td>0.239</td>
</tr>
<tr>
<td>higher S/N (7 dB)</td>
<td>4.733</td>
<td>-0.060</td>
</tr>
</tbody>
</table>
Conclusions

- The j-factor results for CVCCVC words are mostly consistent with the previous results using CVC stimuli.
- One striking new result is that $j_{\text{word}}$ 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 confusible stimuli.
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)
hanks
References


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


Signal-dependent Noise

(1) \[ s_{noisy} = s + \alpha \cdot \pm 1 \cdot s \]

where \( \pm 1 \) is determined randomly on a sample per sample basis, and \( \alpha \) is defined as:

(2) \[ \alpha = \sqrt{\frac{1}{10^{\frac{SNR_{dB}}{10}}}} \]
This method has the advantage that S/N is constant for the entire utterance, rather than using an average as with additive (broadband) noise.

Noise is generated on the fly. The resulting noise sounds very similar to broadband noise, and previous experiments using signal-dependent noise find very similar results to broadband noise.

The method has cross-linguistic differences.

This method has the advantage that S/N is constant for the entire utterance, rather than using an average as with additive (broadband) noise.

Noise is generated on the fly. The resulting noise sounds very similar to broadband noise, and previous experiments using signal-dependent noise find very similar results to broadband noise.

The method has cross-linguistic differences.
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 *babwɔn*
- misspellings *concious* for *conscious*
Can $j$ be larger than $n$?

Consider the words *hot* and *hut*.

The raw CELEX frequency of *hot* is 2498 and *hut* has a frequency of 396.

Consider the following hypothetical spoken word recognition results for *hot* and *hut*:

<table>
<thead>
<tr>
<th></th>
<th>$p_{C1}$</th>
<th>$p_V$</th>
<th>$p_{C2}$</th>
<th>$p_P$</th>
<th>$p_W$</th>
<th>$j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>.90</td>
<td>.8</td>
<td>2.12</td>
</tr>
<tr>
<td>hut</td>
<td>.9</td>
<td>.2</td>
<td>.9</td>
<td>.54</td>
<td>.1</td>
<td>3.74</td>
</tr>
</tbody>
</table>

The same bias for *hot* appears as a bias against *hut*. A result of $j > n$ does not make sense for subjects.
Can $j$ be larger than $n$?

<table>
<thead>
<tr>
<th>Item</th>
<th>freq</th>
<th>dens $p_p$</th>
<th>$p_w$</th>
<th>$j$</th>
<th>$p_{C1}$</th>
<th>$p_{V1}$</th>
<th>$p_{C2}$</th>
<th>$p_{C3}$</th>
<th>$p_{V2}$</th>
<th>$p_{C4}$</th>
<th>errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>hosted</td>
<td>1</td>
<td>1.11</td>
<td>.71</td>
<td>.1</td>
<td>6.74</td>
<td>.13</td>
<td>.97</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>posted, coasted,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hasted, toasted</td>
</tr>
<tr>
<td>chances</td>
<td>2.5</td>
<td>4.91</td>
<td>.92</td>
<td>.8</td>
<td>2.67</td>
<td>.83</td>
<td>1</td>
<td>.97</td>
<td>.87</td>
<td>.87</td>
<td>chancing, cancers,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>cancer, Candice</td>
</tr>
</tbody>
</table>

Robert Felty

Cross-linguistic differences ... – 38