**Lexical Frequency Effects in the Psychological Manifestation of Morpheme Structure Constraints**

Robin Melnick  
San José State University  
robin@melnick.us

---

**Which is ‘closer’ to English?**

\[
\begin{array}{ll}
\text{/frʌm/} & \text{/zwʌg/} \\
\text{'from'} & \text{'zweig'}
\end{array}
\]

---

**Word-Distance Perception Metrics**

- Greenberg and Jenkins (1964)
  - Fewer productive substitution patterns → more distant
- Chomsky and Halle (1968)
  - Max # distinctive segmental features in violation of phonotactic constraints
- Vitz and Winkler (1973)
  - Metric includes allowance for insertions / deletions
- Breen and Jackson (1998)
  - Distinguishing features weighted + segmental transitions contribute
- McAllister and Brodda (2002)
  - Distance measured in speech comprehension
  - Insertions/deletions, seg. features, comparison of syllable onset, nucleus, coda, initial seg. (alliteration), stress pattern, rhyme (nucleus + coda)…

---

**Ohala & Ohala 1986**

- Experiments with subject ratings of distance perception
- Yielded evidence that G&J substitution-pattern metric more effectively predictive than C&H distinctive-feature measure
- Still, certain ‘disturbing’ (O&O 1986:247) results did not follow predictions
- Some intuitive judgment involved in their G&J calculations
- Might a lexical frequency effect be in play here?

---

**Lexical Frequency in Phonology**

- Connine et al. (1993)
  - LF increases speech-recognition accuracy, but only with slower response times
- Makashay (2001)
  - Likely to ‘correct’ metathesized clusters w/ higher LF
- Davidson (2006)
  - Does non-word-initial cluster LF in English enhance ability to reproduce non-English words with same clusters in word-initial position? (No!)
- Zhao and Jurafsky (2007)
  - LF affects tone in Cantonese: low-freq words hyper-articulated, higher pitch

---

**Greenberg & Jenkins Scoring**

<table>
<thead>
<tr>
<th>Total valid patterns:</th>
<th>/frʌd/</th>
<th>/gteʃt/</th>
<th>/mbjʒ/</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>S = 17 – total =</td>
<td>2</td>
<td>5</td>
<td>15</td>
</tr>
</tbody>
</table>

---

**Corpora and Tools**

- American National Corpus, 2nd Release
  - 20M+ words
  - Actually several different corpora
    - Switchboard, Call-Home, Charlotte, NY Times, Berlitz Travel Guides, Slate Magazine, ICIC
    - Corpus of Fundraising Texts, MICASE (Michigan Corpus of Academic Spoken English), various non-fiction, various fiction, medical research articles, anonymized posts to websites
- Software to calculate frequencies for
  - Spoken, written, combined

---

**LF for Phonetic Sequence?**

- What’s the lexical frequency of [kʰɾʌf]?  
  - Ignore predictable elements
    - Stress, aspiration
  - Three-stage process
    - Convert to (modified) ARPAbet → k r ah ch (manual)
    - Software converts to orthography by comparison to CMUDict
    - Software calculates associated LF from ANC data

---

**LF for a Nonce Form?**

- Start with G&J substitution method, CCVC
- Program finds ALL possible substitutions, corresponding LF for each
- Aggregate the LF scores → Substitute LF (SLF)
- Hypothesis: weighted scoring better
  - Words that are “farther away” should count less …
  - Weighted Substitute LF (WSLF)

\[
WSLF = \sum_{d=1}^{n} SLF_d \times 10^{(1-d)}
\]
Is WSLF an effectively predictive metric?

Yes…

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Predictions</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WSLF</td>
<td>Closeness Rank</td>
</tr>
<tr>
<td>frʌd</td>
<td>0.023840</td>
<td>1</td>
</tr>
<tr>
<td>steɪf</td>
<td>0.017156</td>
<td>2</td>
</tr>
<tr>
<td>gteɪt</td>
<td>0.010599</td>
<td>3</td>
</tr>
<tr>
<td>prym</td>
<td>0.003431</td>
<td>4</td>
</tr>
<tr>
<td>mbyʒ</td>
<td>0.000001</td>
<td>5</td>
</tr>
</tbody>
</table>

Pearson's correlation: $r = -0.941$, $p = 0.0171$

Is WSLF independent of G&J?

No!

As G&J increases, WSLF increases, with strong correlation: $r = -0.9049$, $p < 0.001$.

Might LF have been a factor in O&O 1986?

Yes!

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Predictions</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>G&amp;J prediction</td>
</tr>
<tr>
<td>klǫb</td>
<td>klæb</td>
<td>A &gt; B</td>
</tr>
<tr>
<td>θlɛz</td>
<td>θlɛd</td>
<td>A &gt; B</td>
</tr>
<tr>
<td>friz</td>
<td>frʌd</td>
<td>A &gt; B</td>
</tr>
<tr>
<td>flıt</td>
<td>flʌg</td>
<td>A &gt; B</td>
</tr>
<tr>
<td>sθɛf</td>
<td>sθip</td>
<td>A &gt; B</td>
</tr>
</tbody>
</table>

Is G&J effective independent of LF?

Not significantly…

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Predictions</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G&amp;J distance</td>
<td>WSLF distance</td>
</tr>
<tr>
<td>smɪl</td>
<td>3</td>
<td>5.8</td>
</tr>
<tr>
<td>sreɪf</td>
<td>4</td>
<td>6.1</td>
</tr>
<tr>
<td>stɛf</td>
<td>5</td>
<td>5.2</td>
</tr>
<tr>
<td>stɛl</td>
<td>6</td>
<td>5.8</td>
</tr>
<tr>
<td>stɪv</td>
<td>7</td>
<td>6.7</td>
</tr>
<tr>
<td>stɪð</td>
<td>8</td>
<td>7.0</td>
</tr>
<tr>
<td>stɪʒ</td>
<td>10</td>
<td>7.0</td>
</tr>
</tbody>
</table>

$r = 0.6772$, $p = 0.0947$$0.7766$, $p = 0.0400$

Is WSLF effective independent of G&J?

Yes.

Subjects heard pairs of words with equal G&J, but differentiated WSLF, rating 11 of 12 pairs as predicted by WSLF, with statistically significant preference in 9 of 11 successes. In the other two successful-prediction cases, the judgment agreed with the prediction, but the scoring variation is not statistically significant. (In one of these, the WSLF difference itself was not statistically significant.) Only one pair was rated opposite from its prediction, though the scoring here was the closest of all the pairs and is not statistically significant. Overall, the aggregate results are extremely statistically significant.

Does single highest-LF potential substitute matter?

No.

- Peak Substitute Lexical Frequency (PSLF)
- Programmatically selected 184 (0.1% of all possible) CCVC sequences
- Just 32 PSLF English words cover the 184 test forms
- A single word from LF = 0.33% accounted for fully 27 of the 184!

Discussion

- Some subjects mentioned ‘rhyming’ as a test strategy. (The Mueller 2006 model includes rhyming effects.)
- Two subjects wondered about “accent” – a lay cover term for stress, pitch?
- Follow up with another language?
- Try a different phonetic dictionary, since CMUDict is limited by its use of ARPAbet?
- WSLF as formulated was just a first-order approximation. Follow up with further tuning?

Conclusions

- Evidence that perceptions of phonological ‘distance’ are significantly affected by the frequency of ‘similar’ words (i.e., potential comparison targets)
- Implications for cognition: Evidence for a gradient (not categorical) view of phonotactic constraints
- Some weighted measure of LF should be included in predictive metrics for phonological similarity

Acknowledgements

- Thanks to Manjari Ohala and Dan Silverman for their generous comments on earlier drafts of the paper
- Thanks to the organizers and sponsors of the LSA 2008 Summer Meeting, especially the Dept. of Linguistics here at The Ohio State University, for their support
- Funded in part by an ’07-08 California State University Sally Casanova Pre-Doctoral Fellowship
- Paper available on request to robin@melnick.us