Hidden knowledge of syllable gap wellformedness

James Kirby and Alan C. L. Yu

Phonology Lab
Department of Linguistics
University of Chicago

Experimental Approaches to Optimality Theory
Ann Arbor
19 May 2007
Outline

1. Background: phonotactic gaps
2. The study
3. Results
4. Discussion
5. Summary
Not all gaps are the same

- Phonologists traditionally recognize a distinction between *accidental* and *systematic* gaps in the lexicon (Fischer-Jørgensen 1952; Halle 1962)

- A grammar is deemed descriptively and explanatorily adequate if and only if it ...
  1. admits possible words (attested words and accidental gaps), and
  2. rules out impossible words (systematic gaps).
Not all gaps are the same

- Accidental gaps *don’t violate any phonotactic restrictions*
- Systematic gaps *entirely absent from the language because of well-defined phonotactic constraints...*
- ...or whatever the phenomenon of interest happens to be (e.g. OCP-PLACE violations: Frisch *et. al.* 2004)
Not all gaps are the same

- Traditional grammatical approaches presume a **categorical** distinction between systematic vs. accidental gaps.
  - all systematic gaps are equally ill-formed;
  - all accidental gaps are equally well-formed.

- Predicts wellformedness judgments to be categorical as well.
Not all gaps are the same

- But: not all unattested words are judged identically.
- Acceptability of unattested words is **gradient** (Ohala & Ohala 1986; Coleman & Pierrehumbert 1997; Frisch *et al.* 2000; Albright & Hayes 2003; Hay *et al.* 2004, etc.)
- Acceptability reflected in **statistical properties of the lexicon** (*n*-gram probabilities, neighborhood density, O/E ratios, etc.)
Inner structure of systematic gaps?

- Accidental gap acceptability is gradient.
- What about systematic gaps?
- Prevailing view seems to be that systematic gaps are categorically ill-formed.
Certain OT theorists have proposed nonce word wellformedness is determined by the relative ranking of markedness constraints (Boersma 1997; Hayes 2000; Zuraw 2000, 2002; Boersma & Hayes 2001; cf. Anttila 1997a, 1997b; Anttila & Andrus 2006).

- Statistical patterns in the lexicon may be reflected in the analysis in the form of stochastically ranked constraints.
- Gradient well-formedness reflects probability of the winning output.
Gradedness is not possible without variation.

“Our basic premise, then, is that intermediate well-formedness judgments often result from grammatically encodeable patterns in the learning data that are rare, but not vanishingly so, with the degree of ill-formedness related monotonically to the rarity of the pattern.” (Boersma & Hayes 2001:73)
Systematic gap → categorical illformedness

- Frisch et. al. (2004); Coetzee & Pater (2006): strength of OCP-P\textsubscript{LACE} restriction correlated with segment similarity (operationalized as observed/expected values)
- Zero-frequency items are assigned O/E = 0.0
- Doesn’t make a prediction about the range of acceptability within the set of zero-frequency items (but cf. Wilson & Hayes 2006, Pater 2007)
Difference in judgments are **task-specific** (Berent & Shimron 1997; Coetzee to appear)

Gradience predicted only when comparing two nonwords

In non-comparative wordlikeness judgment tasks, judgments were categorical, *with no intra-group variation*

Within-group acceptability variation was observed only in the comparative task

Prediction: gradience only in comparative wordlikeness tasks.
Other studies focus on gaps which do not violate language phonotactics

- Frisch, Large, & Pisoni (2000); Bailey & Hahn (2001): designed not to contain any over phonotactic violations
- Albright & Hayes (2003): contained only one obvious phonotactic violation [bzarʃk]
- Iverson & Salmons (2005): investigate subtypes of accidental gaps

These studies thus have very little to say about acceptability of systematic gaps.
Research questions

Few (no) studies have examined variation within systematic gaps.

1. Is systematic gap acceptability gradient?
2. Is it task-specific?
3. Is it supported by lexical statistics?
Our study: Cantonese

- Why Cantonese?
- Well understood historical phonotactic gaps.
- Highly restricted syllable phonotactics
- Related work on Mandarin (Myers 2002; Myers & Tsay 2004, 2005)
Cantonese phonotactics

- Yue dialect spoken in Hong Kong, Guangdong province, diaspora. (C)(G)V(V)(C) syllable structure
- 19 onsets: /p pʰ t tʰ ts tsʰ k kʰ kʷ kʰw m n ɳ f s h l j w/
- 6 codas: /p t k m n ɳ/
- 11 diphthongs: /ai ei au eu ei εu əy ɔi ui iu ou/
- 6 tones: /55 25 33 21 23 22/
- Due to various documented sound changes, several syllable combinations are not possible
Systematic gaps

Labial gaps
- Labial onsets do not occur in syllables with labial codas (*pap)
- Labial codas do not occur with rounded vowels (*um)
- Labial onsets do not occur with front rounded vowels (*my)

Onset-tone gaps
- Aspirated onsets do not occur with 22 tone (*pʰa22)
- Unaspirated onsets do not occur with 11 or 23 tones (*pa11)

Coronal-vowel gaps
- Coronal onsets and codas may not co-occur with the nuclei /o, u/ (*ton)
- Coronal onsets also do not occur with the vowel /u/ (*tup)
Experimental corpus

432 items conforming to a CV(C) template, derived from all possible combination of

- eight onset phonemes /f, p, pʰ, m, s, t, tʰ, n/;
- three vowel phonemes /a:, i:, u:/;
- three codas /m, n/ and ∅;
- six tones /55, 25, 33, 21, 23, 22/;

This resulted in 162 attested syllables and 270 nonwords.
Nonwords

Of the nonwords,

- 61 fill labial dissimilation gaps;
- 36 fill onset-tone gaps;
- 42 fill coronal gaps;
- 27 syllables filled two types simultaneously, and 1 all three.

The remaining 103 syllables were judged to be accidental gaps because they did not violate any phonotactic constraints per se.
Ten native speakers of Cantonese were presented with a randomized series of items from the corpus and given two tasks per stimulus:

**Lexical decision task**
"Is this a word of Cantonese?" (yes/no)

**Wordlikeness rating task**
“How good a word of Cantonese is this?” (1-7)

Items judged on a 7-point scale, with 1 indicating “very poor - highly unlikely to be a real word of Cantonese” and 7 indicating “very good - a highly prototypical Cantonese word”
Results

The study

Results

Discussion

Summary

Background: phonotactic gaps

The study

Results

Discussion

Summary

Results

<table>
<thead>
<tr>
<th>gap type</th>
<th>Zscorearcsin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Att</td>
<td>−0.6</td>
</tr>
<tr>
<td>Tone</td>
<td>−0.4</td>
</tr>
<tr>
<td>Acc</td>
<td>−0.2</td>
</tr>
<tr>
<td>Lab</td>
<td>0.0</td>
</tr>
<tr>
<td>Mul</td>
<td>0.2</td>
</tr>
<tr>
<td>Cor</td>
<td>0.4</td>
</tr>
<tr>
<td>Cor=Coronal</td>
<td></td>
</tr>
<tr>
<td>Mul=Multiple</td>
<td></td>
</tr>
<tr>
<td>Lab=Labial</td>
<td></td>
</tr>
<tr>
<td>Acc=Accidental</td>
<td></td>
</tr>
<tr>
<td>Tone=Onset−Tone</td>
<td></td>
</tr>
<tr>
<td>Lex=Lexical</td>
<td></td>
</tr>
</tbody>
</table>

The figure shows the Zscorearcsin values for different gap types: Att, Tone, Acc, Lab, Mul, and Cor. The bar graph indicates the distribution of Zscorearcsin values for each gap type.
Wilcoxon rank sum tests ($\alpha = 0.016$)

Onset-tone
- $\mu = 0.018$
- $\sigma = 0.578$

Accidental
- $\mu = 0.28$
- $\sigma = 0.46$

Significant
- $p = 0.0015$
Wilcoxon rank sum tests ($\alpha = 0.016$)

Att Tone Acc Lab Mul Cor
Lex=Lexical
Tone=Onset−Tone
Acc=Accidental
Lab=Labial
Mul=Multiple
Cor=Coronal

Accidental
- $\mu = 0.28$
- $\sigma = 0.46$

Labial
- $\mu = -0.302$
- $\sigma = 0.367$

Not significant
- $p = 0.0825$
Wilcoxon rank sum tests ($\alpha = 0.016$)

Labial
- $\mu = -0.302$
- $\sigma = 0.367$

Coronal
- $\mu = -0.713$
- $\sigma = 0.24$

Significant
- $p < 0.001$
Lexical statistics

- Phonotactic probability
  - Speakers are sensitive to sequential segment probabilities
  - Nonwords with high phonotactic probability should be judged as “more wordlike”
  - Operationalized as average conditional probability, joint n-gram probability, n-gram frequency stated in terms of natural classes...

- Neighborhood density
  - Hearing novel words activates a set of stored exemplars
  - The more exemplars activated, the more the nonword resembles an existing word
  - Operationalized as edit distance, weighted edit distance (GNM), others?
Phonotactic probability

- Phonotactic probability operationalized as average bigram log probability:

  \[ P(W) \approx \frac{1}{\text{length}(W)} \sum_{i=1}^{\text{length}(W)} -\log_2 p(w_i|w_{i-1}) \]

- Zero-frequency bigrams were assigned a small non-zero probability
Lexical neighborhood density

- Neighborhood density (Greenberg & Jenkins 1964): number of lexical neighbors which differ by $k$ changes (substitution, deletion, addition)
- Calculated using the Chinese Character Database (Kwan et. al. 2003) which covers a total of 13,060 character entries
- Weighted by token frequency in Hong Kong Cantonese Adult Language Corpus (HKCAC: Leung & Law 2001), containing 140,000 monosyllables drawn from around 8 hours of speech
Phonotactic probability

- Demonstrated as predictive of *nonword* acceptability judgments (Coleman & Pierrehumbert 1997; Frisch, Large, & Pisoni 2000; Vitevitch & Luce 1998; etc.)
- Vitevitch & Luce 1998: claimed as *dominant* predictor
Phonotactic probability

**Results**

- **Not significant**
  - $R_a^2: -0.002$
  - $F(1, 430) = 0.1387$
  - $p = 0.71$

- **Words**
  - $\mu = 12.34$
  - $\sigma = 2.79$

- **Nonwords**
  - $\mu = 13.11$
  - $\sigma = 3.83$
Neighborhood density

- Lexical neighborhood density: dominant in the processing of real words (Vitevitch & Luce 1998; Bailey & Hahn 2001)
- Less effective a predictor for nonword wordlikeness judgments?
Neighborhood density

Significant
- $R_a^2 = 0.277$
- $F(1, 430) = 166$
- $p < 0.001$

Words
- $\mu = 179.45$
- $\sigma = 67.32$

Nonwords
- $\mu = 96.57$
- $\sigma = 63.07$
Multiple regressions

**Entire corpus**
- $R^2_a = 0.3429$
- $F(2, 429) = 113.4$
- $p < 0.001$

**Words**
- $R^2_a = 0.052$
- $F(2, 159) = 4.43$
- $p = 0.013$
- Only density is significant

**Nonwords**
- $R^2_a = 0.214$
- $F(2, 267) = 37.71$
- $p < 0.001$
- Both factors significant
Phonotactics and neighborhood density

Phonotactics/wordlikeness trend weaker overall than density/wordlikeness trend.
Discussion

- There appears to be *intra-group variation* (cf. Frisch & Zawaydeh)
- There appears to be intra-group variation *in a non-comparative wordlikeness task* (cf. Coetzee)
- There appears to be *degrees of ill-formedness among zero-frequency items* (cf. Boersma & Hayes)
Towards an explanation

- Why did we observe intra-group variation?
- Why is neighborhood density apparently such a good regressor, and phonotactic probability such a poor one?
Towards an explanation

Maybe still a function of task type?

- Inclusion of lexical fillers may cause lexical density to emerge as the dominant cue of wordlikeness (Bailey & Hahn 2001; Shademan 2006)
- However, Frisch et al. 2000 found that, in the presence of grammatical probability and lexical similarity effects, grammatical probability was a better predictor of well-formedness judgments...
- and Shademan (2006) claims phonotactic log probability was an invariantly good predictor; lexical density just emerged as a better predictor when real-word fillers were introduced.
- Not so for Cantonese - why?
Towards an explanation

- English, which permits complex onsets and codas, allows for a far greater number of logically possible monosyllables ($n > 158,000$) than does Cantonese ($n = 5,130$ [19 initials $\times$ 45 finals $\times$ 6 tones])
- English also makes use of a much smaller proportion of these possibilities ($10,000$ monosyllables $\approx 6\%$) than does Cantonese ($1,900$ monosyllables, $\approx 36\%$)
Towards an explanation

- This may explain why neighborhood density correlates with wordlikeness so well in Cantonese: most nonwords are similar in at least one segment/toneme to existing lexemes.
- The fact that most nonwords have lexical neighbors may underlie the emergence of lexical neighborhood density as a predictor of wordlikeness.
Towards an explanation

- Why wasn’t phonotactic probability a good regressor?
- Most sequences violate phonotactics, so expected probabilities will be low
- Might improve with a better model?
Future directions

- Wilson & Hayes (2006) adjust O/E values to allow the size of E to reflect the acceptability of non-occurring segment pairs
- Pater (2007) outlines a mechanisms making use of the GLA in a Harmonic Grammar framework that can assign gradient acceptability values to zero-frequency sequences
Future directions

- Explore more sophisticated neighborhood models: Bybee suggests high type frequency neighbors might actually NOT count as part the neighborhood for purposes of the lexical density calculation (dissociation effect)
- Explore more sophisticated phonotactic models (backoff, smoothing)
- Replicate study with stimulus set balanced for lexical statistics
Gradient acceptability effects emerge even among nonwords which roundly violate phonotactic constraints.

In Cantonese, acceptability seems to be correlated most strongly with lexical neighborhood density.

Correlation of lexical statistics to wordlikeness influenced not only by task, but also by the phonotactic and lexical properties of a given language.


For Further Reading II


