**Information-Theoretic account of the epenthetic vowel in Korean**

Sung-Hoon Hong

Hankuk University of Foreign Studies & Indiana University

(hongshoon@hufs.ac.kr)

---

### Introduction


- Their idea is that less marked segments or words have less information content.
- Hume, Hume & Bromberg proposed Information Content (IC) (context-sensitive, more precisely) to measure the markedness of individual segments.
- The epenthetic vowel of a language has the lowest context-sensitive IC values among all vowels in that language.
- Goldsmith proposed Phonological Complexity to address the wellformedness of bigger linguistic units such as words.
- The representational harmony of a word is maximized by minimizing its information complexity (Goldsmith 2002).
- Where there is morphophonemic alternation, the selection of a right phoneme in morphophonemic alternations.

### Information Theory

The amount of information (or information content) is measured by how much variation and difference there is for an element, the lower the probability and the higher the information content of the element.

**Essential formulas:**

(1) \[ I(C|x) = - \log P(x|C) \]

(2) \[ MI(C|y) = \sum \log \frac{P(x|y|C)}{P(x|\overline{C})} \]

### Morphophonemic alternation in Information Theory

Goldsmith (2001, 2002)'s representation theory of a word is maximized by minimizing its Phonological Complexity (PC), which is the average IC value of the elements that comprise the word.

\[ PC(x) = \sum \log P(x|y|C) \]

Regarding morphophonemic alternation, Goldsmith(2002-04) suggests that the selection of a correct phoneme leads to PC minimization.

- Morphophonemic alternation to be tested: the occurrence of \( \overline{v} \) vs. \( \overline{w} \) where \( \overline{w} \) is a word entry, \( \overline{v} \) is a potential candidate.

- We have only considered 'phonemic' level. Our information about unigrams and bigrams were extracted from a phonetically transcribed corpus. An analysis based on a phonetically transcribed corpus might suggest a different result.

### Conclusion

By showing that Korean epenthetic vowel is indeed lowest in context-sensitive IC, this research supports Hume & Bromberg’s proposal that context-sensitive IC can be a measure for segmental markedness.

**References**


Table 1. Context-sensitive IC values of Korean vowels

| Vowel  | Frequency | IC(x|y) | MI(C|y) | PC(x) |
|--------|-----------|--------|--------|-------|
| i      | 5555993   | 4.4252 | 14.7001 | -21.2883 | -2.1659 |
| e      | 2216827   | 5.6729 | 6.0951 | -4.1774 | 6.1530 |
| a      | 1031531   | 4.2735 | 4.9005 | 10.6369 | 19.8535 |
| o      | 3914951   | 4.8425 | 18.2835 | -2.9674 | 21.2995 |
| u      | 5979700   | 4.2372 | 13.8783 | 4.1957 | 22.5064 |
| n      | 2674416   | 5.3902 | 16.7618 | 0.4094 | 22.5644 |
| m      | 9770880   | 3.6350 | 22.0351 | 3.0478 | 29.1378 |
| p      | 1761557   | 6.0600 | 19.4430 | 15.1943 | 40.8577 |

Table 2. The comparison between the PC values of a form with \( /i/ \) (PC1) and a form without \( /i/ \) (PC2)

<table>
<thead>
<tr>
<th></th>
<th>PC1</th>
<th>PC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-final stems</td>
<td>0.2089</td>
<td><em>0.6454</em></td>
</tr>
<tr>
<td>V-final stems</td>
<td>0.6454</td>
<td><em>0.2089</em></td>
</tr>
</tbody>
</table>

Why are non-alternating suffixes not amenable to PC minimization?

- There may be several potential reasons.
- Most non-alternating suffixes begin with an obstruent, while most alternating suffixes start with a sonorant.

Using PC for a measure of phonotactic markedness is partially successful, and a precise formula is yet to come to guarantee the selection of a right phoneme in morphophonemic alternations.

---

**Access to the full text is available through the university’s institutional repository.**

---

**Information-Theoretic account of the epenthetic vowel in Korean**

Sung-Hoon Hong

Hankuk University of Foreign Studies & Indiana University

(hongshoon@hufs.ac.kr)

---

**Introduction**


- Their idea is that less marked segments or words have less information content.
- Hume, Hume & Bromberg proposed Information Content (IC) (context-sensitive, more precisely) to measure the markedness of individual segments.
- The epenthetic vowel of a language has the lowest context-sensitive IC values among all vowels in that language.
- Goldsmith proposed Phonological Complexity to address the wellformedness of bigger linguistic units such as words.
- The representational harmony of a word is maximized by minimizing its information complexity (Goldsmith 2002).
- Where there is morphophonemic alternation, the selection of a correct phoneme leads to complexity minimization.

**Essential formulas:**

(1) \[ I(C|x) = - \log P(x|C) \]

(2) \[ MI(C|y) = \sum \log \frac{P(x|y|C)}{P(x|\overline{C})} \]

**Morphophonemic alternation in Information Theory**

Goldsmith (2001, 2002)'s representation theory of a word is maximized by minimizing its Phonological Complexity (PC), which is the average IC value of the elements that comprise the word.

\[ PC(x) = \sum \log P(x|y|C) \]

Regarding morphophonemic alternation, Goldsmith(2002-04) suggests that the selection of a correct phoneme leads to PC minimization.

- Morphophonemic alternation to be tested: the occurrence of \( \overline{v} \) vs. \( \overline{w} \) where \( \overline{w} \) is a word entry, \( \overline{v} \) is a potential candidate.

- We have only considered 'phonemic' level. Our information about unigrams and bigrams were extracted from a phonetically transcribed corpus. An analysis based on a phonetically transcribed corpus might suggest a different result.

- The calculation of PC is here based on 'forward' conditional probability. Other methods of probability calculation might yield a different outcome.

**Conclusion**

By showing that Korean epenthetic vowel is indeed lowest in context-sensitive IC, this research supports Hume & Bromberg’s proposal that context-sensitive IC can be a measure for segmental markedness.

Using PC for a measure of phonotactic markedness is partially successful, and a precise formula is yet to come to guarantee the selection of a right phoneme in morphophonemic alternations.