Suffixing Preferences as a Consequence of Probabilistic Reasoning
Yoshihiko Asao
yoshihik@buffalo.edu

Introduction

Suffixes are more common than prefixes in the world’s languages (Greenberg 1965)

Previous Psycholinguistic Accounts

Greenberg (1957)
- The listener prefers to hear stems first, which convey more substantial information (among other suggestions)
- The parser needs the lexical information carried by a stem earlier than the grammatical information carried by an affix

Cutler et al. (1985)
- Greenberg (1963). Some universals of grammar with a particular reference

Observations

- Prefixes are difficult to identify in English
- Prefixes are difficult to identify in Reverse English too, indicating that the effect in English cannot be simply attributed to the low frequency of prefixes
- Differences are small in Walman
- Prefixes are at least as difficult as other types of morphemes

Procedure

Create a morpheme-level bigram model of a language without distinguishing word boundaries from word-internal morpheme boundaries

For a given phonetic input, the parser finds the most likely message based on the bigram model

At each morpheme boundary, calculate how confident the parser is about the existence of the boundary

message: He likes me.
partial input: /hi/
he 93.9% boundary confidence: 93.9%

The simulation program parsed 1,000 randomly generated sentences for each language

Results

- Morpheme boundaries were classified into three types: word boundary, boundary after prefix, and boundary before suffix
- Mean confidence was calculated for each type

Discussion

Observations

- Prefixes are difficult to identify in English
- Prefixes are difficult to identify in Reverse English too, indicating that the effect in English cannot be simply attributed to the low frequency of prefixes
- Differences are small in Walman
- Prefixes are at least as difficult as other types of morphemes

Three factors behind the difficulty with prefixes

Length: The end of a short morpheme is difficult to identify because short morphemes are more likely to match a part of other morphemes by chance

Predictability: Prefixes are less predictable from preceding context than suffixes because:
- There is more uncertainty across a word boundary than within a word
- Most prefixes are preceded by a word boundary, while suffixes are always preceded by a word-internal morpheme boundary

Additional Cues: Phonotactic and suprasegmental cues are less likely to be available

Discussion (cont.)

The boundary can be identified immediately
- If the plural is a suffix (pipe-s):
  /p aj p s/
- If the plural is a prefix (s-pipe):
  /s p aj p/

The boundary can only be identified retrospectively

Predictability:
- Prefixes are less predictable from preceding context than suffixes because:
  - There is more uncertainty across a word boundary than within a word
  - Most prefixes are preceded by a word boundary, while suffixes are always preceded by a word-internal morpheme boundary

Additional Cues:
- Phonotactic and suprasegmental cues are less likely to be available for a word boundary than for a word-internal morpheme boundary

Conclusion

Sufffixing preferences may follow from the probabilistic nature of the human parser, without making assumptions specific to morphological processing

References

Linguistics. 23, 723-756.
