

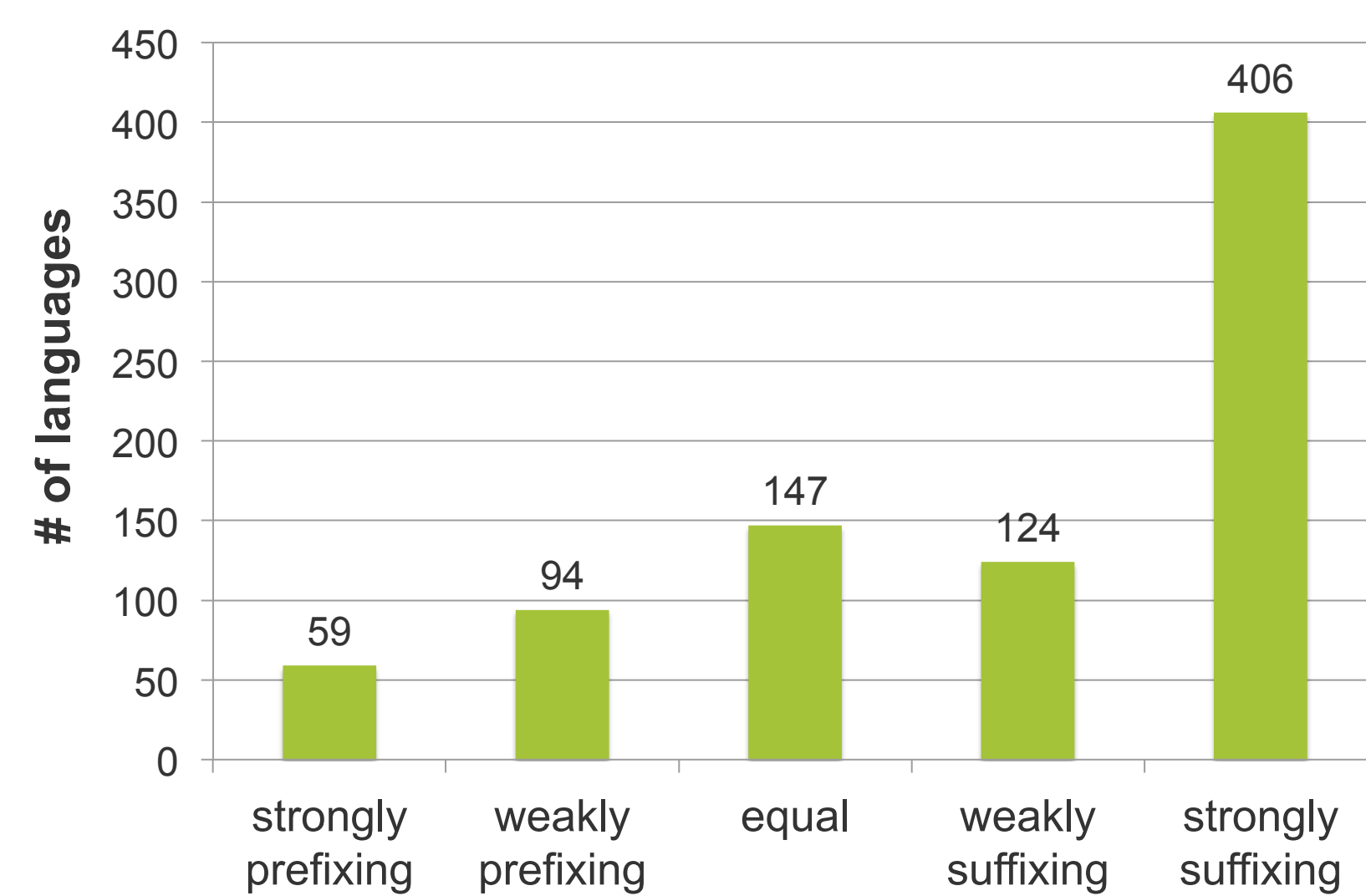
# Suffixing Preferences as a Consequence of Probabilistic Reasoning

Yoshihiko Asao

yoshihik@buffalo.edu

## Introduction

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



(Dryer and Haspelmath 2011)

## Previous Psycholinguistic Accounts

Greenberg (1957)

- The listener prefers to hear stems first, which convey more substantial information (among other suggestions)

Cutler et al. (1985)

- The parser needs the lexical information carried by a stem earlier than the grammatical information carried by an affix

Unclear why asymmetry is only found in morphology

## Our Approach

Prefixes are **more difficult to detect** than suffixes

This prediction follows from **distributional facts** such as length and frequency and the assumption that human parser utilizes **probabilistic knowledge**

Our argument is demonstrated by simulations based on the following three typologically diverse languages:

**English** (strongly suffixing)

- Frequency information taken from **Google Web NGram**
- Phonological and morphological information taken from **CELEX2 Lexical Database**

**Reverse English** (strongly prefixing)

- The same as English except phonemes are given in the reverse order

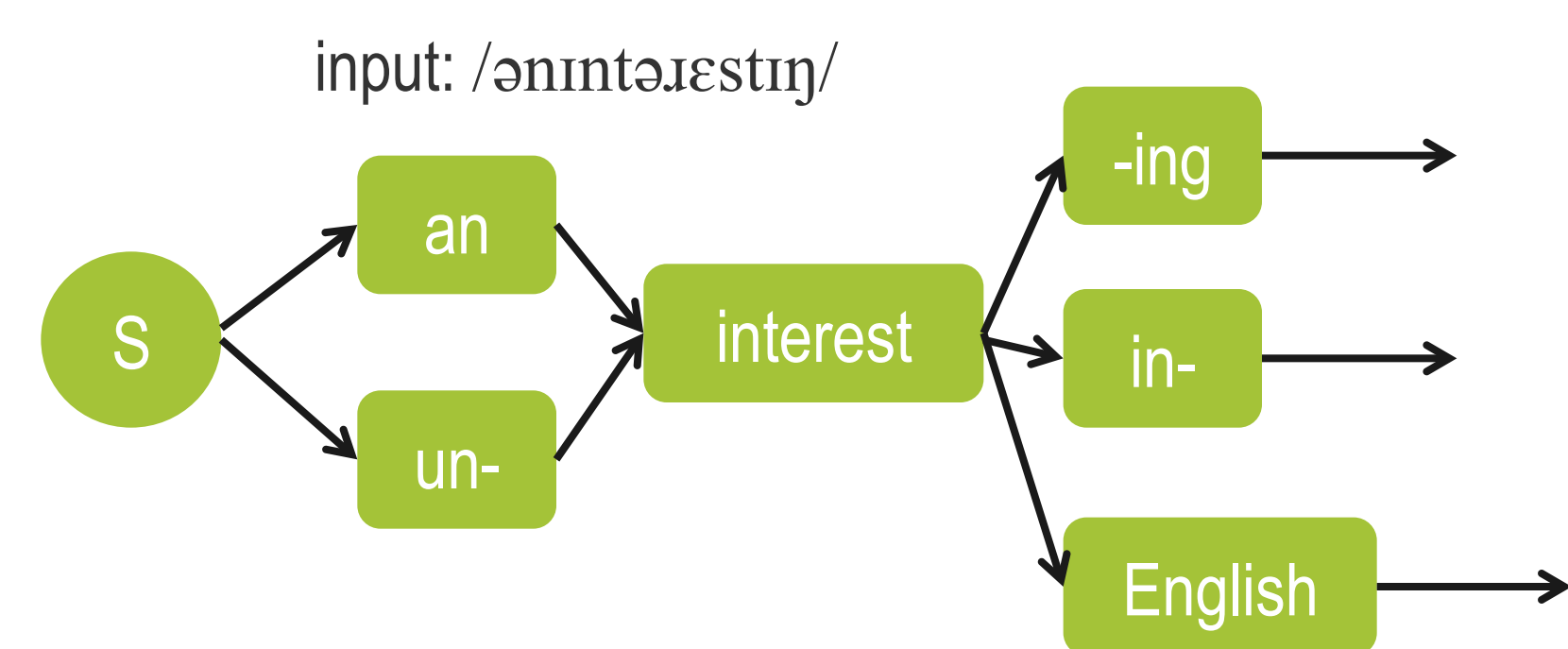
**Walman** (weakly prefixing)

- A Papuan language (data courtesy of Matthew Dryer)

## 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/

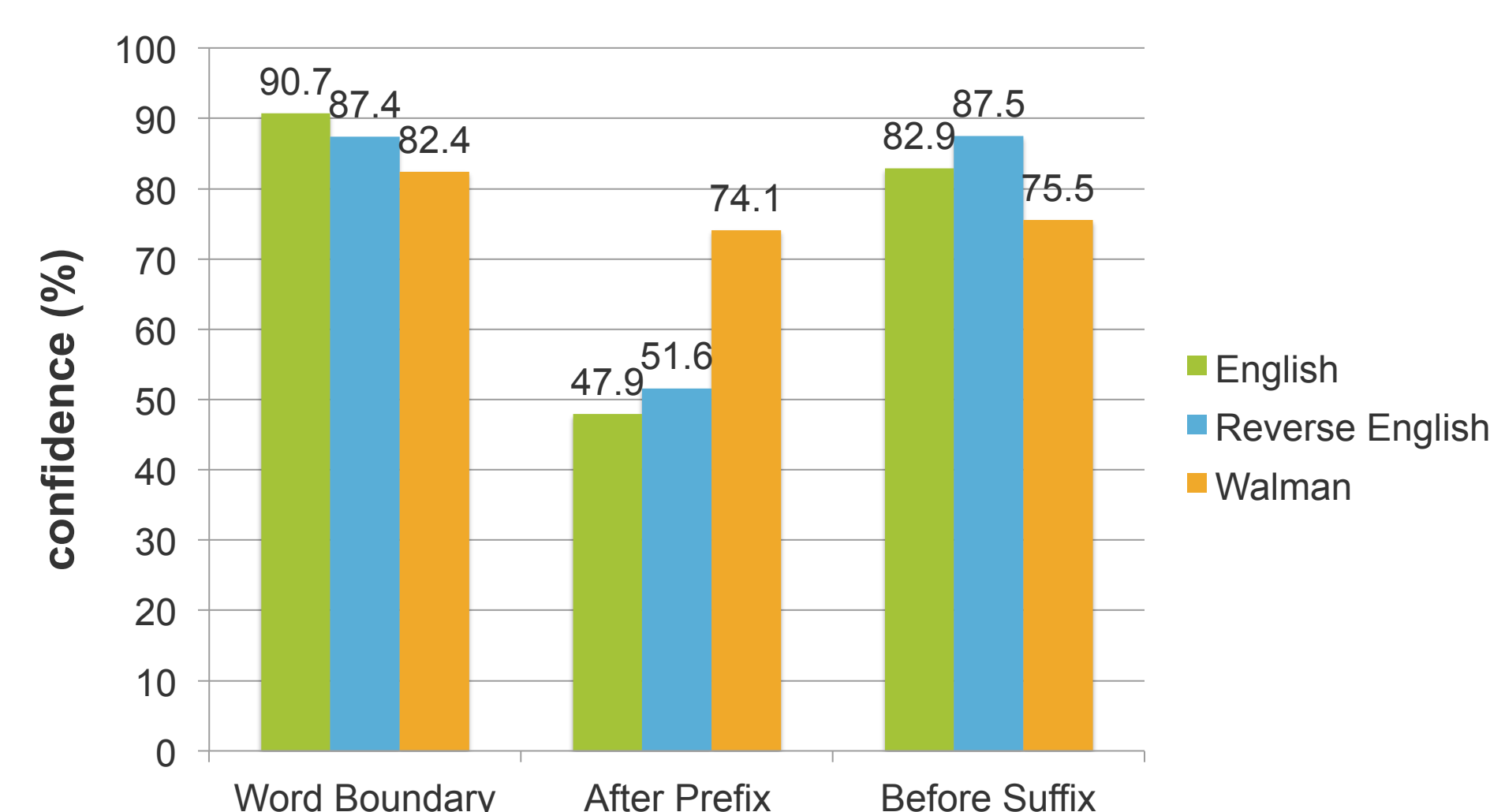
Morpheme	Confidence (%)
he	93.9%
heat	3.6%
heath	0.7%
heal	0.5%
Hebrew	0.5%
...	...

← **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

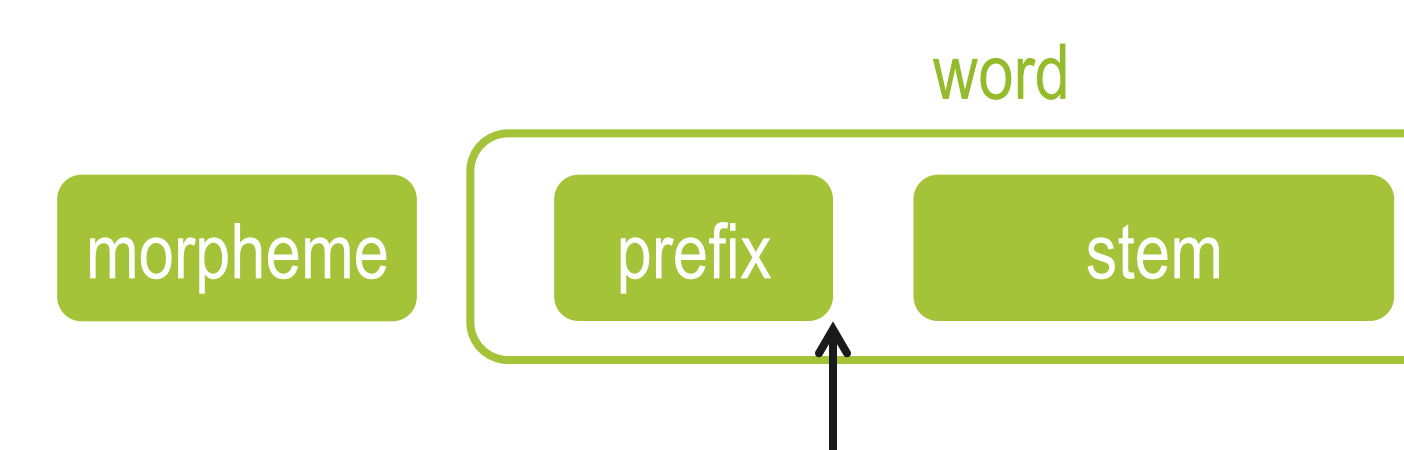


## Discussion

### 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



A boundary after prefix is hard to identify because:

**Length:** A prefix is short

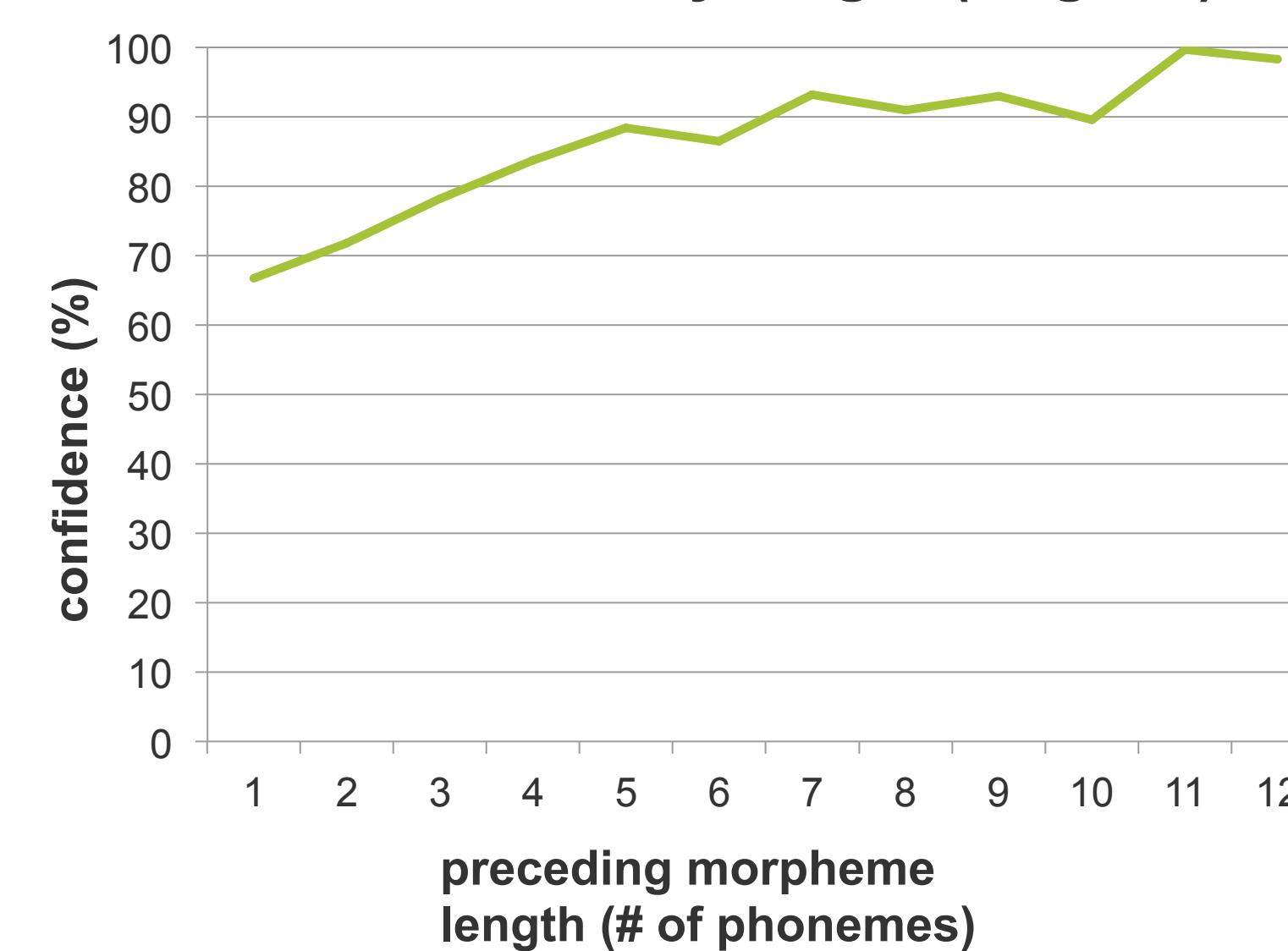
**(Un)predictability:** A prefix is typically the first morpheme after a word boundary

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

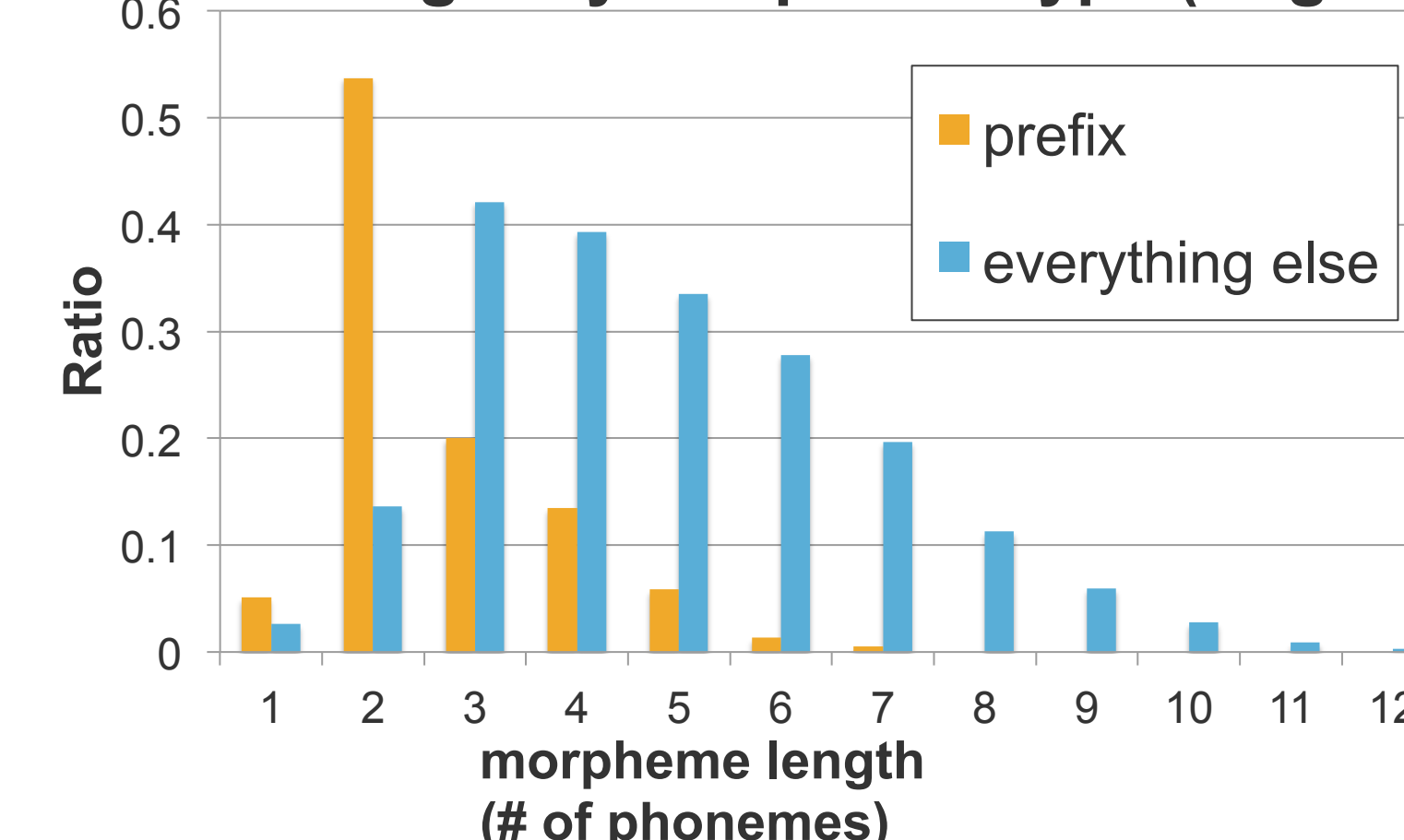
### Length:

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

### confidence by length (English)



### length by morpheme type (English)



## Discussion (cont.)

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

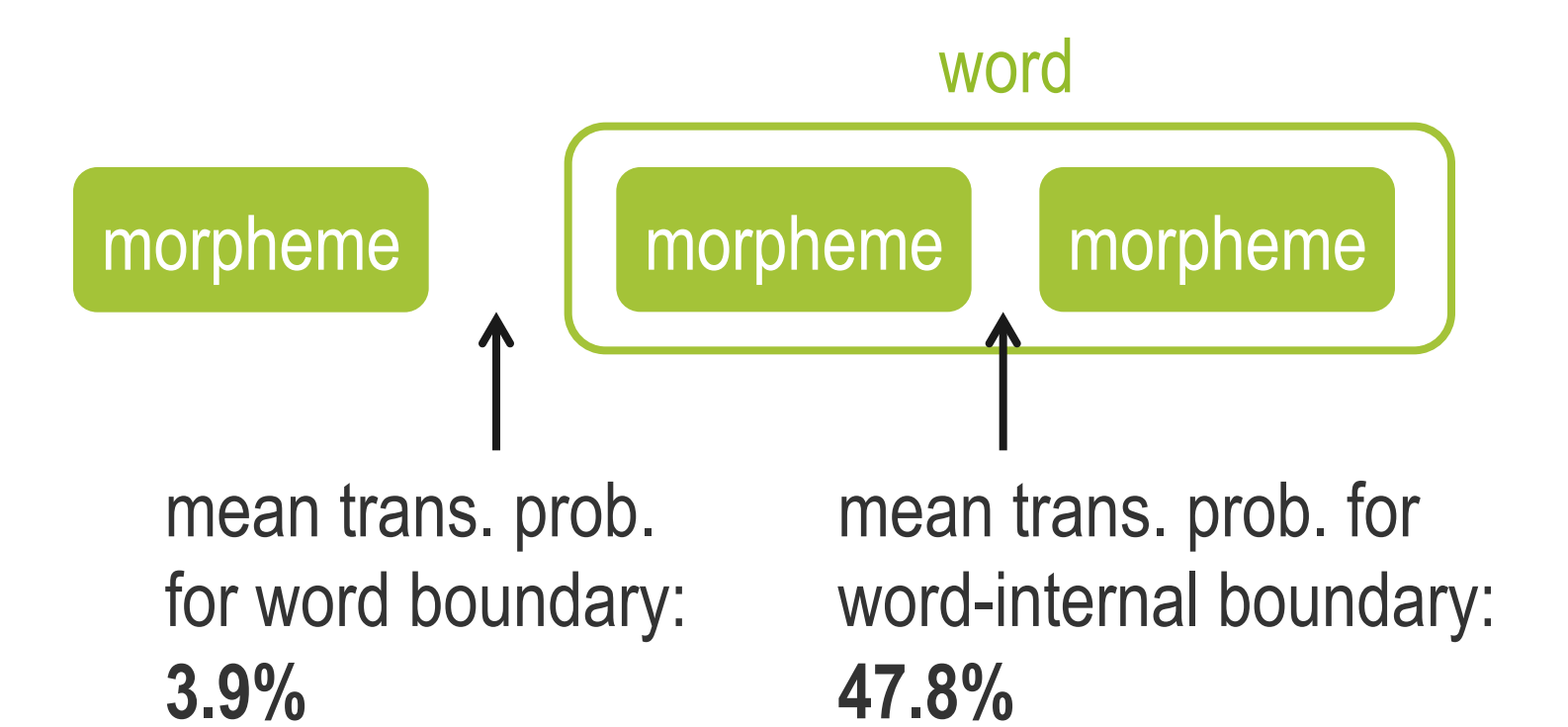
If the plural is a prefix (s-pipe):  
/s p a j p/

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 more likely to be available for a word boundary than for a word-internal morpheme boundary

## Conclusion

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

## References

- Cutler, Hawkins and Gilligan (1985). The suffixing preference: a processing explanation. *Linguistics*, 23, 723-758.
- Dryer and Haspelmath (2011). *World Atlas of Linguistic Structures Online*.
- Greenberg (1957). *Essays in Linguistics*.
- Greenberg (1963). Some universals of grammar with a particular reference to the order of meaningful elements. *Universal of Language: Volume 2*.