Cuing Infants in: From Universal to Language-specific Cues in Word Segmentation

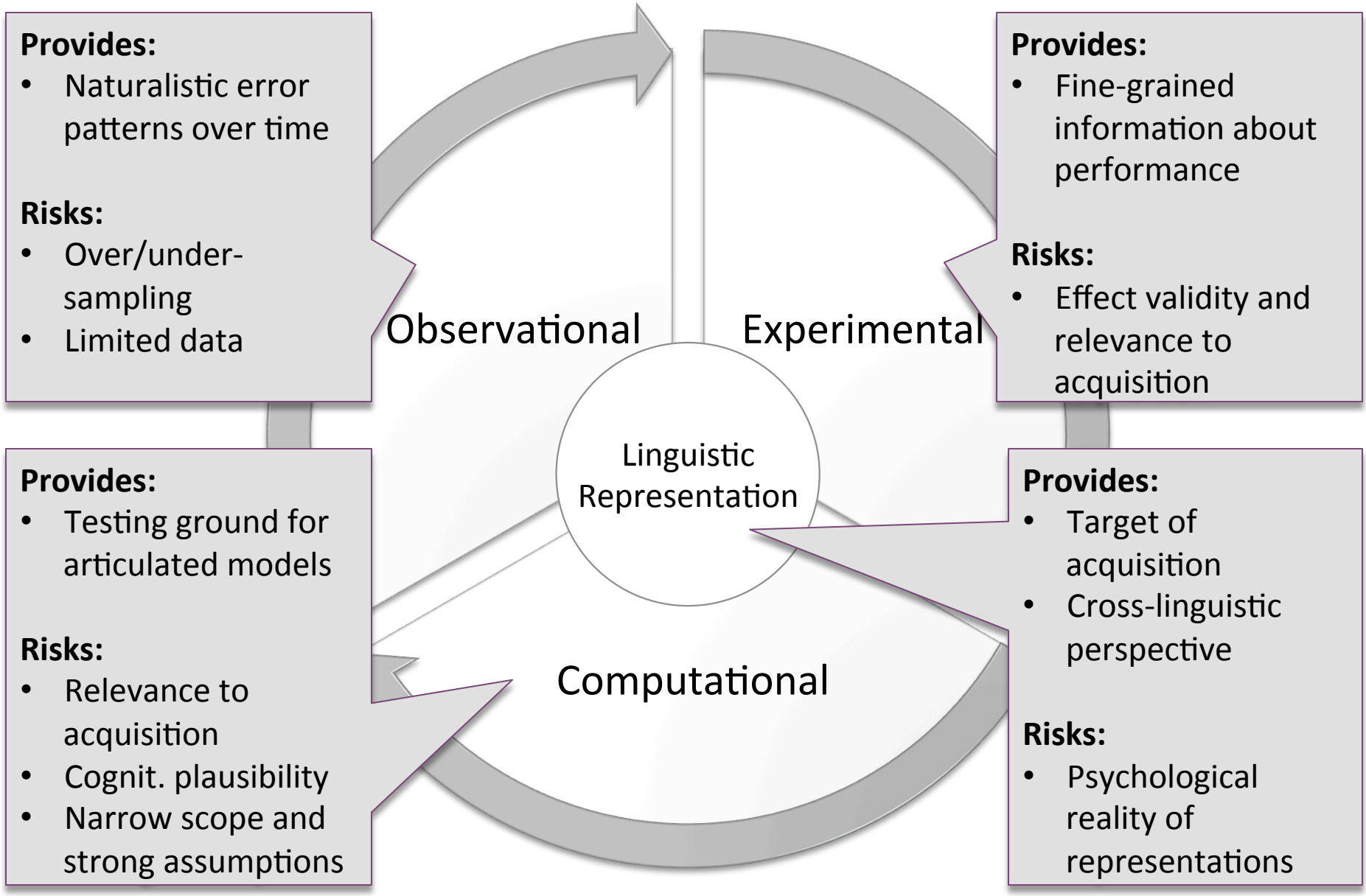
Constantine Lignos and Charles Yang

University of Pennsylvania
Department of Computer and Information Science
Institute for Research in Cognitive Science

ICIS 2012

6/8/2012

Towards understanding language acquisition



I. Modeling infant word segmentation

Our modeling goal:

- Interest: how we can explain the behavior of children as they learn to segment words
- Word segmentation is not an end in itself: provides useful *units* (Peters, 1983) for learning and understanding
  - Lexicon, morphosyntax, phonology

Build the *simplest* model that:

- Aligns with infants' capabilities
- Replicates infants' behavior in a principled fashion
- Performs well at the task

Developmental patterns in segmentation

- Undersegmentation at younger ages (Brown, 1973; Clark, 1977; Peters, 1977, 1983)
- Oversegmentation at older ages (Peters, 1983)
- With time, attending to multiple cues, most popularly:
  - Words in isolation not sufficient (Aslin et al., 1996), but probably helpful (Brent and Siskind, 2001; Lew-Williams et al., 2011, this session; Johnson, this session)
  - Bootstrapping from known words (Bortfeld et al., 2005; Dahan and Brent, 1999)
  - Dominant stress patterns (Jusczyk et al., 1999)
  - More easily identify novel words at beginning and ends of utterances at 8 months (Seidl & Johnson, 2006)

Modeling development

- One perspective: simple, language- (possibly domain-) general system fuels early lexicon growth
  - AKA “proto-lexicon”, “initial cohort”
  - Possibly TPs (Saffran et al. 1996 et seq.) or chunking (Hewlett & Cohen, 2011; Perruchet & Vintner, 1998)
- With good lexical candidates, more rich language-specific cues can be learned:
  - Stress patterns
  - Morphology
  - Phonotactics

Modeling assumptions

- Learner is given syllabified input
  - As with artificial language learning (Saffran et al, 1996 et seq.)
  - Younger infants treat syllables holistically (Bertoncini and Mehler, 1981; Bijeljac-Babic et al., 1993; Jusczyk and Derrah, 1987)
- Able to map acoustic signal to strong/weak stress on syllables (Johnson & Jusczyk, 2001)

Overview of our algorithm

- Segmenter has a lexicon of potential words it builds over time
  - Starts empty, words are added based on segmentation of each utterance
  - Each word has a score
- Operates online
  - Processes one utterance at a time
  - Cannot remember previous utterances or how it segmented them, only lexicon
- Operates left-to-right in each utterance to insert word boundaries between syllables

Model in a nutshell

1. Use utterance boundaries to help find initial words.
2. Bootstrap from known words.
3. Reward the words that appear to lead to better segmentations, penalize the ones that lead us astray.

- We'll (quickly) work through some examples
  - Orthography for easy reading, input is syllabified phonemes from CHILDES (Brown's Adam, Eve, and Sarah):

    B.IH0.G|D.R.AH1.M
    HH.AO1.R.S
    DH.OW0.Z|AA0.R|CH.EH1|K.ER0.Z

(Lignos, 2011)

In the beginning…

- Just add whole utterances to the lexicon
- Gets words in isolation for free, but often more than one word

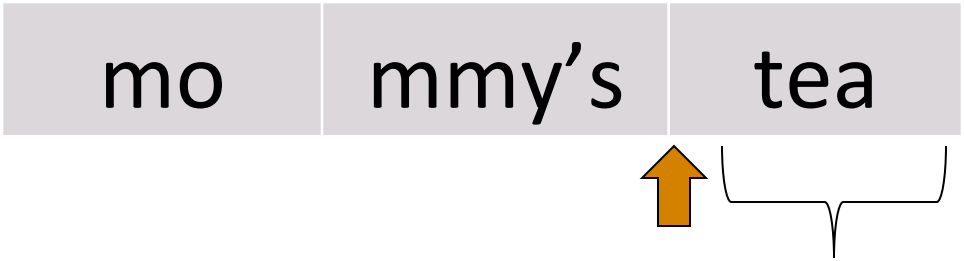
Lexicon:
bigdrum

(Lignos, 2011)

Subtractive Segmentation

- Use words in the lexicon to break up the utterance
- Increase word's score when it is used
- Add new words to lexicon

Lexicon:
mommy's
tea
…



Treat remainder as word.

Either:

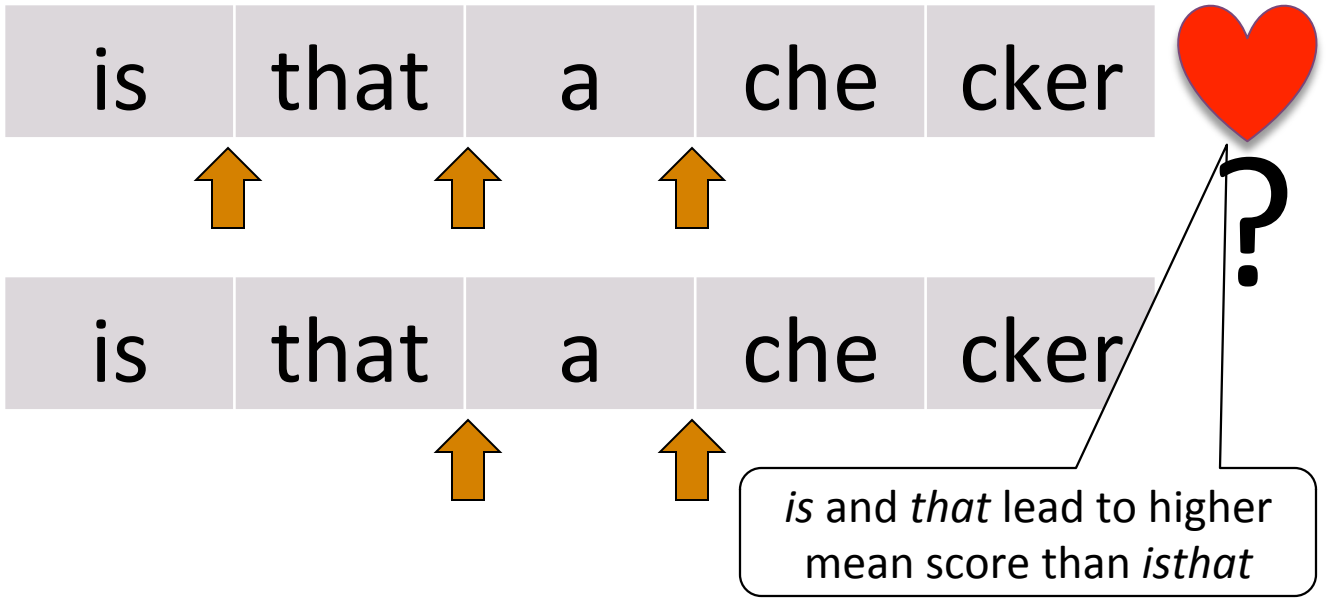
1. Always add to lexicon
2. Add to lexicon only if it touches utterance end (*Trust)*.

(Lignos, 2011)

Multiple hypotheses

- For multiple possible subtractions two options:
  - Greedy approach (Lignos and Yang, 2010)
  - Pursue two hypotheses (*beam search*)
- Two hypotheses allow for *penalization*: reduce score of word that started losing hypothesis

Lexicon:
a
is
isthat
that
…



(Lignos, 2011)

Predictions

- Default assumption of utterance = word → infants will start with oversized units and words in isolation
- Rich-get-richer scoring → As the learner is exposed to more data, learner will tend to use high-frequency elements (Frank, this session)
- Penalization → Use of collocations will decrease with time

Results are true to predictions (Lignos, 2011; Lignos and Yang, to appear) and match developmental patterns (Brown, 1973; Clark, 1977; Peters, 1977, 1983)

III. Moving forward in development

Learning from the lexicon

- The learner's lexicon may be imperfect, but we can still learn from it
- Most relevant to segmentation:
  - Stress pattern
  - Morphology
  - Word length (Lew-Williams and Saffran, 2011)
  - Phonotactics
  - Maybe learn TPs from lexicon instead of utterance?

Stress pattern learning

- Identification of stress pattern in the language
  - Multisyllabic words in the learner's acquired lexicon have stress-initial rate of 70.3%
  - Taking advantage of this bias in learning reduces errors by 37.0%
- How does this bias affect the learner?
  - Hypothesis: learner commits to bias as soon as it is reliable, thus higher initial stress rate → faster adoption of bias
  - Hungarian (stress-initial by rule): word segmentation errors are rare (MacWhinney, 1976; Peters, 1983)
  - English (generally stress-initial, ~80%): children develop reliable bias and apply it readily (*TARis* for *guiTAR is, NAna for baNAna*)
  - French (arguably no word-level stress): Delayed competence compared to other languages (Nazzi et al., 2006)

Morphological learning

- Other modeling work has identified importance of morphology in segmentation (Berg-Kirkpatrick et al., 2010; Johnson, 2008)
  - But no cognitively plausible mechanism for using it
- Segmentation output given to simple MDL-based incremental learner
- Findings:
  - Initial lexicon of high-enough quality to learn frequent suffixes
  - Acquisition order similar to that observed by Brown (1973)

Predicted order	Suffix	Example
1	-ɪŋ	bake-ing
2	-z	happen-s
3	-d	happen-ed
4	-ɚ	bake-er
5	-t	check-ed
6	-ən	broke-en
7	-i:	prett-y

Future work

- Feedback of other levels (morphology, stress, etc.) into segmentation
  - Previous work (Johnson, 2008)
- Understanding role of phonotactic learning
  - Adriaans and Kager, 2010; Gorman, 2012
- Evaluation in (many!) more languages
  - “Correct” segmentation is less trivial

Conclusions

- A simple, language-independent model provides a strong starting point for learning language-specific segmentation strategies
- By using combinations of simple learning mechanisms, we can explore the interplay of learning at various levels of representation
- We've just scratched the surface so far. Ask: how can computational modeling help further understanding of this problem?

Thanks to:
Mitch Marcus
NSF IGERT #50504487

Constantine Lignos
lignos@cis.upenn.edu
http://www.seas.upenn.edu/~lignos
Code/data:
https://github.com/ConstantineLignos/WordSegmentation