

Psych156A/ Ling150

Spring 2010

Review Questions: Word Segmentation

(1) Terms/concepts to know: syllables, transitional probability, transition probability minimum, word segmentation, psychological plausibility, CHILDES database, precision, recall, algebraic learning

(2) What was Saffran, Aslin, and Newport (1996)'s belief about the relation between transitional probability and word boundaries? That is, when did they believe transitional probability between syllables was likely to be higher and when did they believe it was likely to be lower?

(3) What is the difference between a transitional probability minimum and a low transitional probability? It is possible to have one without the other – explain how.

(4) Why was it necessary for Saffran et al. (1996) to test children on real words vs. part-words from the artificial language? That is, why was the second experiment necessary to make their point about children's ability to track transitional probabilities for word segmentation?

(5) Why doesn't a computational model of a language acquisition process (such as word segmentation) need to include every detail of a child's experience?

(6) What does it mean for a learning algorithm to be psychologically plausible? Why would it be important for a computational model to use a psychologically plausible learning algorithm?

(7) Where did Gambell & Yang (2006) get the input for their model from? Is this a good source of realistic input that a child might hear? Why or why not?

(8) Gambell & Yang found that tracking transitional probabilities failed to reliably segment child-directed speech data. What property of the data set caused the model to fail in Gambell & Yang's (2006) study? Why did this cause the transitional probability model to fail?

(9) Here are some words from the imaginary Guin language:

pengo          zu          pencrom          pentanor          az

(a) Where would an algebraic learner put word boundaries in the syllable sequence below? Your answer should indicate a boundary by identifying the syllable sequence the boundary would come between (ex: between "pen" and "go"). There may be more than one boundary.

Sequence: pen go az la to pen crom mer tem pen ta nor

Here are the same Guin words, along with their stress patterns:

(Note: pén = the “pen” syllable has stress)

péngo      zú      péncrom      péntanor      áz

(b) Where would an algebraic learner with knowledge of the Unique Stress Constraint put word boundaries in the syllable sequence below? Your answer should indicate a boundary by identifying the syllable sequence the boundary would come between (ex: between “pen” and “go”). There may be more than one boundary.

Sequence: pén go áz lá tó pén crom mér tem pén ta nor

(10) Does using statistical learning by itself always yield poor performance for word segmentation? Cite evidence to support your answer.