

Ling151/Psych156A

Winter 2018

Review Questions: Syntactic Categorization

(1) Terms/concepts to know: grammatical category, syntactic category, noun, verb, adjective, preposition, semantic bootstrapping hypothesis, frequent frame, corpora, bigram, hit, miss, false alarm, correct rejection, precision, recall, closed-class categories, open-class categories, productivity, lexical overlap, Zipfian distribution, observed overlap, expected overlap

(2) One way to think about a word's syntactic category (like noun or verb) is that it is really just a description of the way that word can be used in the language. Given the following contexts, say whether you think the novel word in each example (indicated by CAPITALS) is a noun (like "goblin"), a verb (like "sing"), an adjective (like "hot"), an adverb (like "dreamily"), or a preposition (like "in"). Be sure to briefly explain why you think so. (You might find it helpful to substitute words you know in place of the novel words, and see which ones fit best.)

- (a) That's a very BOFT bog.
- (b) Is Sir Didymus REKking the Bog of Eternal Stench?
- (c) Ludo shouldn't have STROOPed his paw in the bog.
- (c) Sarah was bitten by a fairy PRING the wall.
- (d) They had to beware the BREER's tricks.
- (e) Hoggle couldn't believe how FREEMILY he was able to do it.

(3) Why can't the semantic bootstrapping hypothesis be used to categorize all the words of a language? (Hint: What mapping rules does it assume? Can you think of an example where that rule doesn't apply?)

(4) Here's an example utterance from the imaginary Guin language.

"felgo bofty mu az berg mu merk berg felgo zu mu var berg az porto mu freggo berg felgo seech mu set berg draz pino trem felgo trem mu peri berg lootem bleeter."

- (a) What is the most frequent frame in this utterance?
- (b) Which words would this frame cluster together?
- (c) What is the second most frequent frame in this utterance?
- (d) What words would this second most frequent frame cluster together?

(5) What does it mean if a syntactic categorization model has a high precision score, but a low recall score? Is this model likely to correctly put together words belonging to the same grammatical category? Is this model likely to correctly group all the words of one category together? What would it mean if the opposite were true, and the model has a high recall score but a low precision score? (Hint: Imagine a scenario where the model groups all words into a single category.) Is this model likely to put together words

belonging to the same grammatical category? Is this model likely to correctly group all the words of one category together?

(6) What experimental evidence do we have that very young children are sensitive to frequent frames in the input?

(7) Is it better for the categorization unit to be a frame or to be a sequential unit like a bigram? (Hint: What's the difference between using bigrams alone vs. being aware that a trigram can be decomposed into pieces that look like bigrams?)

(8) Is it better for the units that make up the frames to be words or to be more abstract units (for example, clusters of words that make up a category)?

(9) Do frequent frames always operate best over words? (Hint: Think about languages like Turkish and German.) How well do frequent frames (using any unit) operate cross-linguistically?

(10) What considerations about children's memory did the model used by Wang & Mintz (2008) make? (Think about what they were trying to improve on from Mintz (2003)'s study.) How did the model by Wang & Mintz (2008) perform, compared to the study done by Mintz (2003)?

(11) Why is it hard to observe directly whether children have abstract categories like Verb? (For example, why can't we tell if a child who says "don't worry" recognizes these as categories Negation and Verb, and not just as an amalgam of don't+worry?)

(12) What's the different between a productive rule, a semi-productive rule, and a non-productive rule for generating two word combinations? (Hint: Think about what each claims about a child's category knowledge.)

(13) Even if a child knows the category Noun, why wouldn't we expect that child to produce all possible nouns in combination with all the determiners (like "a" and "the") that the child knows? That is, if a child knows the nouns "penguin", "kitty", and "puppy", should we expect the child to spontaneously produce all of the following in natural communication settings: "a penguin", "the penguin", "a kitty", "the kitty", "a puppy", and "the puppy"? Why or why not?

(14) What does it mean for linguistic output to have a Zipfian distribution? Why does this make it hard for us to assess everything a child knows? (Hint: Think about what this means for figuring out lexical overlap when determining knowledge about syntactic categories.)

(15) Is observed or expected overlap calculated directly from children's productions? What does it mean if the observed overlap and the expected overlap match?

(16) Based on the results from Bates et al. (in prep) and Yang (2010, 2011), do very young children seem to have categories for open-class categories? What about closed-class categories? Does it matter which type of phrase (noun or verb phrase) these items appear in?