

p. 3, 12: "Does the extent of lexical flexibility observed for a language correlate with size of corpus? Does it correlate with the token frequency of the lexeme? (Chapter 2)" -> Not sure I'm getting how this is an interesting question (at least when formulated like that). So, you have *shoes* as a noun in (1a) and *shoe* in (1b); let's assume in the SBCAE, shoe is used 3/4 as a noun and 1/4 as a verb - now why is there an expectation that those proportions change with corpus size and why would that be interesting? Is corpus size here a not-so-well-named proxy for something more along the lines of corpus diversity or register plurality? Then, ok, maybe, but corpus size per se - I don't see what that might do/explain. p. 3, 12: in that combination: Regardless of what you want to do with corpus size and even more so if you mean something other than mere size, I think the token frequency of the lexeme is insufficient. In fact, I myself have shown in the last few years in a variety of papers that corpus dispersion is often more predictive than corpus frequency. One of the best examples: the words enormous and staining are equally long and equally frequent in the Brown corpus (37 times), but the former shows up in 36 parts of the 500 Brown corpus parts, the latter in 1! And this effect is particularly pronounced in middle frequencies of the type you're likely to be interested in. Thus, while Bybee and others don't get that, dispersion or contextual diversity have been shown time and again to be more relevant predictors of things that usage-based linguistics has ascribed to frequency.

p. 4: "Do certain semantic domains tend to exhibit more lexical flexibility than others? Is lexical flexibility sensitive to the animacy hierarchy? Are property concepts more flexible than time-stable items and/or events?" - question 3 seems to be a more precise version of question 2 - do you have specific hypotheses here and, more importantly, do they lead to falsifiable predictions? If yes, make sure you state them (either there or later of course).

p. 4: "Does the current and/or previous choice of grammatical role for a lexeme correlate with choice of lexical category?" -> Are you asking whether, for instance, whether the fact that *shoe* is used in a subject slot increases its probability of being a noun??

p. 4: "Does information status (given vs. new vs. activated) correlate with choice of lexical category?" -> Why would it? (UPDATE: p. 10f. talks a bit about that, ok.)

I realize that you s're saying that these things will be discussed below, but a little more guidance for the reader would be nice so as to not have ppl be annoyed at having to keep all this in memory for later.

p. 5: "This structuralist approach to lexical categories, which came to be known as the distributional method (Harris 1951:5), constituted a major advance in the typological study of parts of speech, and essentially became **the sole method of syntactic analysis in modern linguistics** (Croft 2001:11)." (my emphasis) -> This seems like a reasonable statement to make given your committee members, but it is just as reasonable when you apply for a job where there's 1 functionalist, one generativist, and one literature person on the committee?

