Generality of substitution — but, proceed with care: We introduced the idea of using distributions to characterize categories, and in doing so we focused mostly on syntactic distributions where we highlighted the tool of substitution. In the video, we included a fixed set of contexts for you all to use when trying to do substitutions, but note that you can use this tool more generally, given that you take proper care. For example, consider all the words which can come after *the* and before *couch*.

- the {big, comfy, maroon, ...} couch ADJ − GOOD

- the {very*, easily*, completely*, ...} couch ADV − NO GOOD

- the {in*, until*, of*, ...} couch PREP − NO GOOD

- ... ... − NO GOOD

This would lead you to conclude that *big, comfy, maroon*, etc. are of the same category, which is true; however, now consider the following data.

(a) the velvet jumpsuit

(b) the suede couch

Are *suede* and *velvet* adjectives? Look at the following data.

(c) the velvet / the soft velvet

(d) the suede / the fresh suede

(e) ?the very velvet jumpsuit

(f) ?the very suede couch

(g) *the more velvet jumpsuit / *the velveter jumpsuit

(h) *the more suede couch / *the suedest couch

What does this suggest about the category of *suede* and *velvet* and the structure of *suede couch* and *velvet jumpsuit*?

Extending the list of categories: In the second video for today (Monday), we saw that we could use syntactic contexts to create a rough partition of the lexicon into groups of words which exhibit similar behavior (you could think of them as equivalence classes), and we called these groups of words **categories**. They mapped on decently to our previous notions of the major groups of words; now, let's consider a less intuitive one.

(i) Consider the words *that, whether, if, for.* Call them **complementizers** and we'll denote the category with a C from here on. Find (a) context(s) in which only these words can appear.

(ii) Are there other categories you can argue for?

Subcategorization — a finer partition within each category: Although the categories we have arrived at so far can account for a lot of the global trends in English data, they by no means highlight the more complex structure within categories. Just as we did with phonology when we introduced features to carve out subsets (natural classes) of phonemes, we can look at finer partitions within each major category of words. We've mostly been using syntactic evidence because it has provided us with clearer conclusions, but please feel free to use morphological evidence if it gets the job done for any of the following.

<u>Nouns</u>: Take the words *goat, milk, sand, tamale, oxygen, water, oil, laptop* and *whiteboard* and separate them into two groups. What data did you use to separate them?

<u>Verbs</u>: Verbs have well-defined subcategories, based on their **argument structure**. A verb's argument structure is the syntactic (and, more generally, semantic) arguments it can take — how many and the type of each one. Try to fill out the table below. (Read $V_{\{NP, \ NP/PP\}}$ as a verb which takes a noun phrase argument and then another argument, which is a noun phrase or prepositional phrase. A verb that does this is *give*. I understand we haven't gone over phrases yet, but assume each word-level category we've learned about can also have a phrasal counterpart. For simplicity, you can consider something like 'the happy dog' as a noun phrase (NP); this will change on Wednesday where we will call it a determiner phrase (DP)).

Subcategory	Example
$V_{\{\varnothing\}}$	
$V_{\{NP\}}$	
	ask
	spare
$V_{\{NP, \ PP\}}$	
$V_{\{NP, \ NP/PP\}}$	*give, send*
$V_{\{NP, \ NP/PP/CP\}}$	

2

Adverbs: There are at least two types of adverbs — let's look into this. Consider the adverbs *completely, certainly, probably* and *luckily.* Put them into two groups, using the distributions they have with respect to the sentence

The crew can rely on my support.

Hint: first figure out which positions an adverb can take in this sentence and then see which adverbs can go where.