Exercise 2

This exercise counts 25% toward your final grade. There are two possible exercises to chose from (2A or 2B). Pick the one that best suits your interests. Exercise 2A can be further developed in the final report if you wish to.

Exercise 2A: Examining polysemy in common nouns

Some nouns accept propositional complement clauses (1), and can also modified with propositional predicates (2). For example, in English, for 'message':

1. Alex received the message that Billie will attend the party.
2. The message was true/factual/false/misleading.

This suggests that 'message' can denote propositional or informational contents of some kind. However, some such nouns are polysemous in that they have multiple, interrelated senses. For instance, 'message' can denote an eventuality:

3. A three-minute message was broadcast yesterday.

'Eventuality' here means an event (e.g., run to the shop), a process (e.g., run), or a state (be Catalan).

Your task is to investigate informational-eventuality polysemy in a natural language, L, of your choice. You should establish:

1. What nouns in L can denote informational contents. (Are there contexts in addition to those mentioned in the background that might help you establish this?)
2. What nouns in L can denote eventualities. (Think about what modifiers, verbs, PPs etc. might help you to establish this.)
3. What the intersection is between these classes.
4. Whether one reading is more or less dominant over another. (Propose a reasonable way to quantify this.)

The report should be a PDF document; no longer than 2 pages. It should include the sections described below. The 2-page limit does not include references (optional); supplementary material (optional); and the list of contributions (only if working in groups).

Introduction (5%)

Briefly explain the task in your own words.

Material and methods (25%)

Describe your approach to investigate informational-eventuality polysemy in a natural language, as well as the data you will be using (where did you get it? how did you process it?)

Results (40%)

Your main results. Remember to address the 4 points mentioned above. Why (not)? Discuss limitations of your data and methods.

Code (30%: 20% replicability/10% clarity)

Make your code publicly available (hosted on, e.g., OSF or github). Remember to extensively comment it. Mention the dependencies that need to be fulfilled to run the code.

List of contributions (unlimited space / only if working in groups)

Who did what in your group. You can use the CRediT system or a variant thereof.

Exercise 2B: Evaluation of spaCy's dependency parser

Evaluate spaCy's dependency parser in two languages. You are free to decide how you go about this but you need to minimally discuss (i) why the benchmark that you used (e.g. Universal Dependency Treebank 2.x) is approriate and (ii) why the measures that you used to evaluate the parser are appropriate.

The report should be a PDF document; no longer than 2 pages. It should include the sections described below. The 2-page limit does not include references (optional); supplementary material (optional); and the list of contributions (only if working in groups).

Introduction (5%)

Give a (very short) introduction to dependency parsing.

Material and methods (25%)

Describe the data you will be using (where did you get it? how did you process it?) and how you will evaluate it (what metrics are you using and why?)

Results (40%)

Your main results. Discuss limitations of your data and methods.

Code (30%: 20% replicability/10% clarity)

Make your code publicly available (hosted on, e.g., OSF or github). Remember to extensively comment it. Mention the dependencies that need to be fulfilled to run the code.

List of contributions (unlimited space / only if working in groups)

Who did what in your group. You can use the CRediT system or a variant thereof.