Assignment 2: Lexical semantics (non-graded part)

Instructions

- In this part, we will release code already at practice time in class (the code is incomplete; your task is to complete it).
- Otherwise, the same applies as in the ungraded part of Assignment 1: none of the ungraded part needs to be handed in, etc.

A. WordNet synsets and lemmas. ***Note***: you can also get acquainted with WordNet through the interface available here: http://wordnetweb.princeton.edu/perl/webwn (English data only).

1. Download the files from Assignment 2 to a folder and create a new project from this folder in Spyder (or your IDE of choice).
2. Open the script wordnet_practice.py and follow the code. Note: In the case that something needs to be downloaded to run it, follow the instructions in the error message; or call nltk.download() and use the interface that appears.
3. Take the word "study". How many wordnet **synsets** does the word have? How many of those are for the word as a *noun*?
4. For every *noun* synset of "study", print the synset's name, definition, and examples. Would you agree that all synsets capture distinct **senses** of the word? Why (not)?
5. Now do the same for the *verb* synsets.
6. For the same synsets, also print the words each synset contains, i.e., the names of its English **lemmas**. Do you agree that the words for each synset are valid **synonyms**? Why (not)?
7. Print the list of **languages** which WordNet currently offers.
8. Going back to the word "study", in addition to printing the names of its English lemmas, also print the names of its **Catalan lemmas**. What does this tell you about the structure of WordNet in how it represents different languages?
9. The wordnet function *all_lemma_names()* has a *lang* argument, but the function *all_synsets()* does not. Why is that?
10. For all languages in WordNet, print how many lemmas the language has in WordNet. (Even when efficiently coded this can take a minute to run.) Which languages are the best represented?

B. Lexical relations in WordNet

11. We provide code to reproduce (approximately) Figure 18.5 from Jurafsky and Martin.
12. Do code that prints the hypernyms for each of the synsets of a word of choice.
13. Do code that prints the hyponyms for each of the verb synsets of a different word of choice.
14. Do code that extracts the lexical relations for the words in Figure 18.3 (e.g., that *meal.n.01* is a hypernym of *breakfast.n.01*).