

Problem Set 2 – Syntax

[due Nov 20]

Sample text (Source: [https://en.wikipedia.org/wiki/Wired_\(magazine\)](https://en.wikipedia.org/wiki/Wired_(magazine)))

Wired is a full-color monthly American magazine, published in both print and online editions, that reports on how emerging technologies affect culture, the economy and politics. Owned by Condé Nast, it is headquartered in San Francisco, California and has been in publication since its first issue in March/April 1993. Several spin-offs have been launched including: Wired UK, Wired Italia, Wired Japan and Wired Germany.

Exercise 1: PoS tag sets

- a) Search the web for the *Penn Treebank* part-of-speech (PoS) tag set. Then find the Universal Dependencies universal PoS tag set¹. Apply both tag sets to the sample text. (I.e., produce two distinct analyses of the text.) If you encounter any problems or points of ambiguity, note them!
- b) Translate the sample text into your mother tongue. (If your mother tongue is English, translate it to your favorite second language.) Then apply the Google universal PoS tag set to the translation. If you encounter any problems or points of ambiguity, note them!

Exercise 2: PoS distribution in spoken vs written language

Create an account on the CQP Web Workbench². Using the corpus "BNC sampler", get acquainted with the straightforward search scenarios: looking for words, looking for PoS tags. Using the "distribution" menu option on the search results page (upper right corner), determine three PoS tags that have a remarkably different distribution in the spoken versus the written part of the corpus. List each tag and its relative frequency. For each tag, do you have a hypothesis why the distribution is so uneven? (**short** explanation.)

Exercise 3: Constituents and arguments versus adjuncts

Mark all sentence constituents in the sample text, e.g. by parentheses. Then, for each constituent, mark arguments and adjuncts, e.g., by different kinds of underlining.

¹ <http://universaldependencies.org/u/pos/>

² <https://cqpweb.lancs.ac.uk/>