

## Problem Set 2 – Syntax

**[due Nov 19]**

### Sample text

(Source: <https://www.nytimes.com/2019/11/11/science/black-fashion-physics-animals.html>)

“The human eye is extraordinarily sensitive to light,” Dr. Woods said. Throw a few dozen photons its way, a few dozen quantum-sized packets of light, and the eye can readily track them.

Dr. Woods pulled a laser pointer from his pocket. “This pointer,” he said, “puts out 100 trillion photons per second.” He switched on the laser and began slowly sweeping its bright beam across the surface of the tray.

### 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 set<sup>1</sup>. 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 Universal Dependencies 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 Workbench<sup>2</sup>. 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.

**Note:** You can consider names as single words for this exercise.

---

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

<sup>2</sup> <https://cqpweb.lancs.ac.uk/>