

## Problem Set 3 – Syntax 2

[due Dec 3]

### Exercise 1: Syntactic dependencies and FrameNet semantic frames

<https://framenet.icsi.berkeley.edu/fndrupal/luIndex>

<http://nlp.stanford.edu:8080/parser/index.jsp>

Sample text:

(1) At 6.30pm, a black SUV hit a lamp post at Sesame Street in Brooklyn. The driver had not seen it because of the glaring sun. He stepped out of his vehicle and inspected the damage. Luckily, just one headlight broke.

a. For each sentence in the text, construct an ideal *FrameNet* representation. (Look up the verb frame, assign the right core and peripheral frame elements.

b. Run each sentence through the Stanford parser.

c. Consider the syntactic dependency representations produced by Stanford parser. Are there any obvious mistakes? For each verb, specify the role linking between the syntactic dependencies and the *FrameNet* roles from your ideal representation.

d. Now try the following sentence:

(2) A hearing is scheduled on the issue today.

Give the Stanford parse and discuss whether or not it accurately reflects the structure of the sentence. What is an unproblematic rephrase of the sentence? Any further observations?

### Exercise 2: Subcategorization

a. Give a CFG that can generate the following sentences (using standard POS tags and the standard phrase types S, VP, NP, PP):

(3) Kim pets the dog.

(4) The cat snores.

(5) Kim puts the cat in the basket.

b. Show that your grammar overgenerates (i.e., generates ungrammatical sentences). [Hint: Use the title of the exercise to guess what I'm aiming for]

c. Adapt your grammar so that these kinds of ungrammatical sentences are not generated any more.

d. Briefly discuss whether this is a good way to treat such properties of the grammar and what effect it will have on the resulting CFGs.

### Exercise 3: Language Profile

Write one paragraph about the basic syntactic properties of your language. What is the basic word order and what role does word order play in this language? Provide a couple of example sentences.

Which NLP resources exist for your language? Are there POS taggers or parsers you can find online? Any corpora with linguistic annotations?