

# EDAN20 - Assignment 6

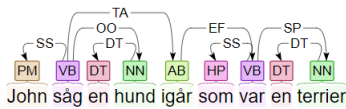
Oliver Cosic

Oktober 2021

## 1 Nonprojective sentence

Took the sentence from the lecture and translated it to swedish. So the sentence becomes: *John såg en hund igår som var en terrier.*

In projective sentences each pair of words which is directly connected is only separated by direct or indirect dependents of Head or Deprel. In the graph below from Langforia we can see that this is not the case for this sentence and that is because of the relation between "såg" and "igår".



## 2 Checking gold standard parsing

So the first sentence is: ROOT Individuell beskattning av arbetsinkomster

Trans	Stack	Queue	Graph
Start	[]	[ROOT, Individuell, beskattning, av, arbetsinkomster]	[]
sh	[ROOT]	[Individuell beskattning av arbetsinkomster]	[]
sh	[Individuell, ROOT]	[beskattning, av, arbetsinkomster]	[]
la	[ROOT]	[beskattning, av, arbetsinkomster]	[Individuell <beskattning]
ra	[beskattning, ROOT]	[av, arbetsinkomster]	[Individuell <beskattning, ROOT>beskattning]
sh	[av, beskattning, ROOT]	[arbetsinkomster]	[Individuell <beskattning, ROOT>beskattning]
la	[beskattning, ROOT]	[arbetsinkomster]	[Individuell <beskattning, ROOT>beskattning, av <arbetsinkomster]
ra	[arbetsinkomster, beskattning, ROOT]	[]	[Individuell <beskattning, ROOT>beskattning, av <arbetsinkomster, arbetsinkomster <beskattning]
END			

### 3 The assignment

So the only thing we actually did ourselves was to extract the features and generate/train the scikit learn models.

The feature was extracted as shown in the figure below:

```
x_dict = []
y_symbols = []

for i, sentence in enumerate(formatted_corpus_train_clean):
    stack, queue, graph = init_config(sentence)
    while queue:
        X_dict.append(extract(2, stack, queue, graph, sentence))
        stack, queue, graph, x=oracle(stack, queue, graph)
        y_symbols.append(x)
    stack, graph = empty_stack(stack, graph)
```

As scikit models we tested both logistic regression and MLP classifier. We got the best result with the logistic regression with n jobs = 16 and verbose = True. The best result obtained was 0.755.