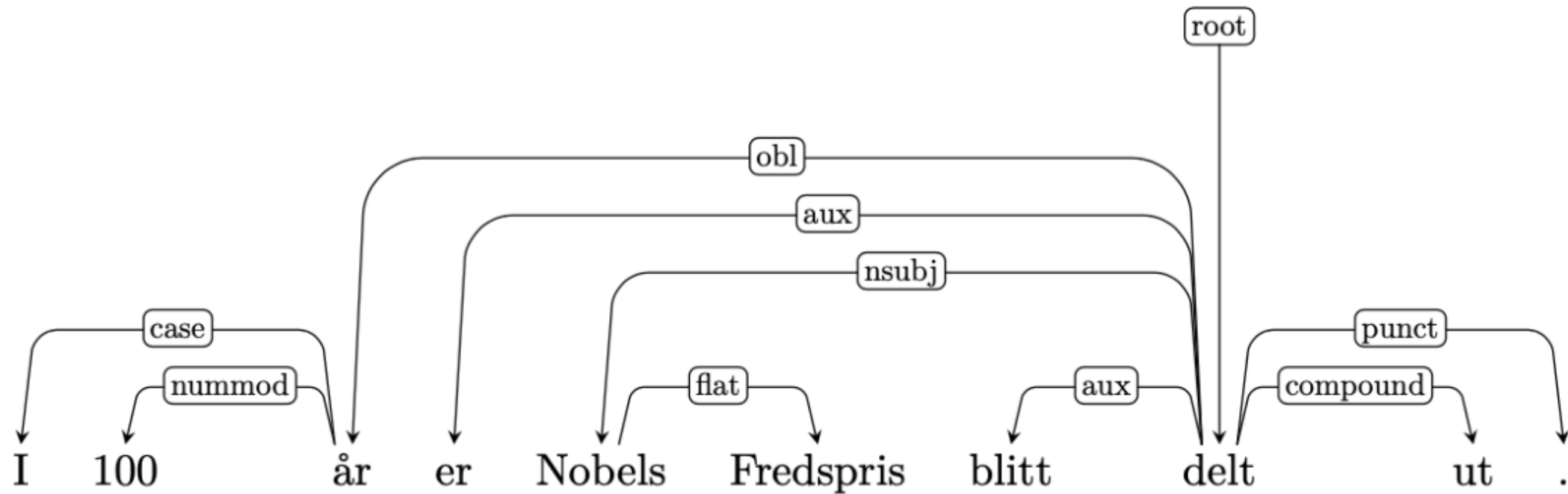# Gjennomgang av oblig 2a

Dependenstre

Oppgave 2 – Trene modeller

```
$ spacy train -p parser nb <model-mappe> <train-set-json> <dev-set-json>
```

Oppgave 3– Evaluering (lese inn)

```
MODEL="models/model-best"
UD_TRAIN="no_bokmaal-ud-train.conllu"
UD_DEV="no_bokmaal-ud-dev.conllu"
NN_DEV="no_nynorsk-ud-dev.conllu"
NNLIA_DEV="no_nynorsklia-ud-dev.conllu"
```

```
# Load conllu files
conllu_dev = ConlluDoc.from_file(UD_DEV)
conllu_nn_dev = ConlluDoc.from_file(NN_DEV)
conllu_nnlia_dev = ConlluDoc.from_file(NNLIA_DEV)


# Load spacy model for Norwegian
nb = spacy.load(MODEL)
```

Oppgave 3– Evaluering (konvertere til spCy-objekter med og uten annotasjon)

```python
# Convert conllu data to spacy format
dev_docs_labeled = conllu_dev.to_spacy(nb)
dev_docs_unlabeled = conllu_dev.to_spacy(nb, keep_labels=False)


# Load nynorsk
dev_docs_nn_labeled = conllu_nn_dev.to_spacy(nb)
dev_docs_nn_unlabeled = conllu_nn_dev.to_spacy(nb, keep_labels=False)


# Load NynorskLIA
dev_docs_nnlia_labeled = conllu_nnlia_dev.to_spacy(nb)
dev_docs_nnlia_unlabeled = conllu_nnlia_dev.to_spacy(nb, keep_labels=False)
```

Oppgave 3– Evaluering (parsing)

```python
def parse(model, docs):
    for doc in docs:
        model.parser(doc)
```

```python
# Parse bokmål
parse(nb, dev_docs_unlabeled)

# Parse nynorsk
parse(nb, dev_docs_nn_unlabeled)

# Parse nynorskLIA
parse(nb, dev_docs_nnlia_unlabeled)
```

Oppgave 3– Evaluering (Attachment score)

```python
def attachment_score(true, pred):
    las = 0
    uas = 0
    total = 0

    for s1, s2 in zip(true, pred):
        for true_token,pred_token in zip(s1, s2):
            if true_token.head.i == pred_token.head.i:
                uas += 1
# Addition: spaCy introduces its own root label (ROOT) which does not match gold, hence lowercasing fixes this
                if true_token.dep_.lower() == pred_token.dep_.lower():
                    las += 1

            total += 1

    return uas/total, las/total
```

Oppgave 3– Evaluering (Attachment score)

```
# Evaluate
uas, las = attachment_score(dev_docs_labeled, dev_docs_unlabeled)
nnuas, nnlas = attachment_score(dev_docs_nn_labeled, dev_docs_nn_unlabeled)
nnliauas, nnlialas = attachment_score(dev_docs_nnlia_labeled, dev_docs_nnlia_unlabeled)
```