

IN1140 H2017 – Oblig 3b

CFGer og semantikk

Innleveringsfrist: søndag 19.11 kl.23.59

Lever inn svarene dine i Devilry (<https://devilry.ifi.uio.no/>) i filer som angir brukernavnet ditt, slik: `oblig3b_brukernavn.py`. Pass på at filen din kan kjøres som et program; det skal ikke være en REPL-sesjon limt inn i en fil.

En perfekt besvarelse på denne oppgaven er verdt 100 poeng. For å bestå kreves minst 100 poeng (av 200 mulige) for $3(a+b)$. Merk at alle 3 (1, $2(a+b)$, $3(a+b)$) obligene må vurderes som bestått for å kunne ta eksamen.

Viktig informasjon: Innleveringsfrister utsettes kun i forbindelse med sykdom (ved egenmelding eller legeerklæring), og omlevering vil ikke være mulig. For informasjon om utsettelse og behandling av fuskesaker se <http://www.uio.no/studier/emner/matnat/ifi/IN1140/h17/oblig.html>

1 Kontekstfrie grammatikker (20 poeng)

Anta at vi har følgende grammatikk:

```
S -> NP VP
NP -> Norah | Ola | sushi | boka | pinner
NP -> D N
D -> en
N -> bok
PP -> P NP
P -> med
VP -> V NP
VP -> V NP NP
VP -> VP PP
V -> gir
V -> spiser
```

Setningen “Norah spiser sushi med pinner” er strukturelt flertydig.

1. Beskriv de to ulike tolkningene med egne ord.
2. Grammatikken over gir kun én analyse for denne setningen. Hvordan kan du utvide grammatikken slik at den kan representere flertydighet av denne typen?
3. Er grammatikken rekursiv? Hvis ja, på hvilken måte? Hvis nei, gi et eksempel på en rekursiv regel.

2 En grammatikk for norsk (50 poeng)

NLTK inneholder flere forskjellige parsere som tildeler syntaktisk struktur til en setning automatisk, i henhold til en grammatikk. Her skal du bruke RecursiveDescent-parseren som står beskrevet i seksjon 8.3. Du kan formulere grammatikken din direkte som en streng, slik:

```
grammar = nltk.CFG.fromstring("""
S -> NP VP
VP -> V NP | V NP PP
PP -> P NP
V -> "saw" | "ate" | "walked"
NP -> "John" | "Mary" | "Bob" | Det N | Det N PP
Det -> "a" | "an" | "the" | "my"
N -> "man" | "dog" | "cat" | "telescope" | "park"
P -> "in" | "on" | "by" | "with"
""")
```

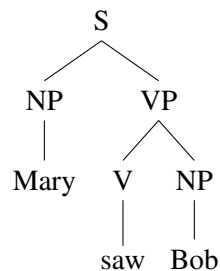
Merk at RecursiveDescent-parseren ikke håndterer venstre-rekursjon, av typen $VP \rightarrow VP PP$, så du må formulere grammatikken uten denne formen for rekursjon. Du kan teste grammatikken på en setning slik:

```
sent = "Mary saw Bob".split()
rd_parser = nltk.RecursiveDescentParser(grammar)
for tree in rd_parser.parse(sent):
    print tree
```

Parseren skriver da ut et tre i klammenotasjon:

```
(S (NP Mary) (VP (V saw) (NP Bob)))
```

Dette tilsvarer dette syntaktiske treet:



1. Du skal nå implementere en kontekstfri grammatikk med denne parseren som analyserer et fragment av norsk slik at setningene under gis riktig analyse:

- (a) (S (NP Per) (VP (V gir) (NP (D en) (N bok)) (PP (P til) (NP Kari))))
- (b) (S (NP Kari) (VP (V gir) (NP Per) (NP boka)))
- (c) (S (NP Ola) (VP (V sover)))
- (d) (S (NP Kari) (VP (V spiser)))
- (e) (S (NP Kari) (VP (V spiser) (NP middag)))
- (f) (S (NP Per) (VP (V finner) (NP boka)))

Vis at setningene i 1-6 gis korrekt analyse ved å parse dem med grammatikken og skrive ut analysen som beskrevet over.

2. Grammatikken slik den er implementert er imidlertid ikke tilfredsstillende, siden den for eksempel vil godta setninger som *Kari sover boka* og *Ola finner*. Verifiser dette ved å skrive ut analysene grammatikken din tildeler disse setningene.
3. Du skal nå skrive en ny og forbedret versjon av grammatikken slik at de grammatiske konstruksjonene i 1-6 tillates, men ugrammatiske konstruksjoner (som *Kari sover boka* og *Ola finner*) er utelukket. Skriv ut analysene den nye og forbedrede grammatikken tildeler de grammatiske setningene 1-6, og vis videre at de ugrammatiske setningene ikke tildeles noen analyse.

3 Manuell annotering av ordbetydning (30 poeng)

I denne oppgaven skal du gjøre en manuell annotering av ordbetydning og kommentere observasjonene dine. Skriv svarene dine som utkommentert tekst i Python-filen din.

Setningene i (1)-(5) under er hentet fra SemCor-korpuset, et korpus som er annotert med ordbetydning, og alle inneholder verbet *leave*.

1. But questions with which committee members taunted bankers appearing as witnesses left little doubt that they will recommend passage of it .
2. The departure of the Giants and the Dodgers to California left New York with only the Yankees .
3. After the coach listed all the boy 's faults , Hartweger said , “ Coach before I leave here , you ’ll get to like me ” .
4. R. H. S. Crossman , M.P. , writing in The Manchester Guardian , states that departures from West Berlin are now running at the rate not of 700 , but of 1700 a week , and applications to leave have risen to 1900 a week .
5. The house has been swept so clean that contemporary man has been left with no means , or at best with wholly inadequate means , for dealing with his experience of spirit .

Slå opp verbet *leave* i WordNet (bruk *Use Word-Net Online*:

<http://wordnetweb.princeton.edu/perl/webwn>). Du skal ikke ta hensyn til betydningene for substantivet *leave*.

For hver av setningene i (1)-(5) skal du velge en betydning (“sense”) fra WordNet for verbet *leave* i setningen og notere valget ditt. På *Use Word-Net online*-siden kan du klikke på Display Options og velge Show Sense Numbers for å få en nummerert oversikt over de forskjellige betydningene. Bruk disse nummerene i svaret ditt.

Videre skal du reflektere rundt arbeidet ditt og besvare følgende spørsmål:

- Hvilke setninger var det vanskelig å annotere og hvorfor?
- Hvilke par (eller grupperinger) av WordNet-betydninger var det vanskelig å skille fra hverandre og hvilke kriterier brukte du for å skille mellom dem?