

UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i: INF2820 Datalingvistikk

Eksamensdag: 10. juni 2013

Tid for eksamen: 0900-1300

Oppgavesettet er på 4 side(r)

Vedlegg: 0

Tillatte hjelpemidler: ingen

*Kontroller at oppgavesettet er komplett
før du begynner å besvare spørsmålene.*

- Du kan svare på norsk, dansk, svensk eller engelsk.
- Du skal besvare alle spørsmålene. Vekten på de ulike spørsmålene er indikert.
- Du bør lese gjennom hele settet slik at du kan stille spørsmål til faglærerne når de kommer i tilfelle noe er uklart.
- Hvis du føler noen forutsetninger mangler, lag dine egne og redegjør for dem.!

Oppgave 1

La G1 være grammatikken med hovedsymbol S og følgende regler:

1. S → NP VP
2. VP → IV
3. VP → TV NP
4. NP → DET N
5. N → N R
6. R → RP S_NP
7. S_NP → NP_NP VP
8. S_NP → NP VP_NP
9. VP_NP → TV NP_NP
10. NP_NP →
11. IV → 'smiled' | 'slept' | 'danced'
12. TV → 'chased' | 'saw' | 'owned'
13. N → 'cat' | 'dog' | 'child'
14. DET → 'a' | 'the' | 'some'
15. RP → 'that'

S_NP er her en regulær ikke-terminal med samme status som f.eks. TV, og det er NP_NP og VP_NP også.

Spørsmål 1.1 (10%)

Tegn trærne (eller treet) grammatikken tilordner til

- a) the cat that the child owned smiled

Spørsmål 1.2 (5%)

Hva vil det si at et språk er regulært?

Spørsmål 1.3 (10%)

Er språket L(G1), dvs. språket generert av G1, et regulært språk? Begrunn svaret!

Oppgave 2 (10%)

Noen av reglene i grammatikk G1 likner på andre regler, f.eks. er regel (9) parallell til regel (3) og reglene (7) og (8) er parallelle til regel (1). En kan derfor redusere antall regler i grammatikken ved å bruke trekk (eng.: "features"). Konstruer en trekkbasert grammatikk (unifikasjonsgrammatikk) for språket L(G1) som har færre regler enn G1.

Oppgave 3

Spørsmål 3.1 (5%)

Noen parsere har problemer med grammatikk G1. Hvilke parsere? Hvilken egenskap ved G1 er det som lager et problem for slike parsere, og hvorfor er det et problem?

Spørsmål 3.2 (10%)

Vis hvordan en chart-parser vil anerkjenne setning (b) ut i fra grammatikk G1.

b) the cat smiled

Spørsmål 3.3 (10%)

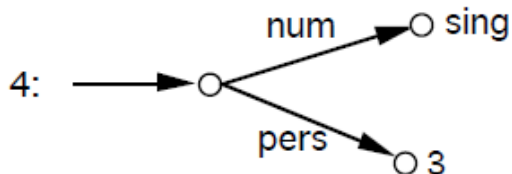
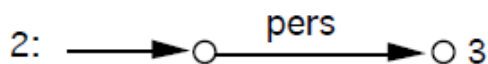
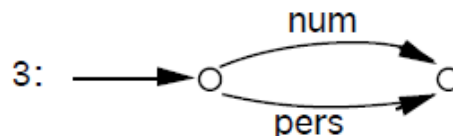
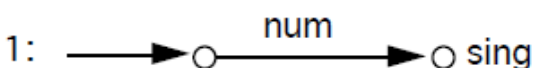
Finn en grammatikk G2 på Chomsky Normal Form (CNF) for språket $L(G1)$. Vis stegene i omformingen av G1 til CNF.

Spørsmål 3.4 (10%)

Vis hvordan en CKY-parser vil anerkjenne setning (a) fra oppgave 1 ut i fra grammatikk G2.

Oppgave 4 (10%)

Betrakt følgende fire trekstruktur. Hvilke par av dem (A,B) er slik at A subsummerer B? Hvilke par av strukturer er unifiserbare? Tegn resultatet av unifikasjonen for de parene som er unifiserbare. Vis også attributt-verdimatriserepresentasjonene for de resulterende unifiserte strukturerne.



Oppgave 5

Vi skal se på semantikk. Vi skal forenkler litt. Vi vil ikke se på kvantorer ("quantifiers"). De eneste NP-er vi vil se på er egennavn ("proper names"). Vi ønsker at (c) representeres som (d):

- c) Fido walks
- d) $walk(f)$

Følgende regler vil gi det ønskede resultat:

S[SEM = <?vp(?subj) >]	-> NP[SEM=?subj] VP[SEM=?vp]
NP[SEM=?np]	-> PropN[SEM=?np]
VP[SEM=?v]	-> IV[SEM=?v]
PropN[SEM=<f>]	-> 'Fi do'
IV[SEM=<\x. wal k(x) >]	-> 'wal ks'

Vi ønsker også at setning (e) får representasjonen (f) og at setning (g) får representasjonen (h).

e) Fido chases Socks

f) *chase(f, s)*

g) Mary gives John Fido

h) *give(m, j, f)*

Spørsmål 5.1 (10%)

Utstyr følgende regler med trekk ("features") som gir det ønskede resultatet, dvs. fyll inn for prikkene.

VP[SEM=...]	-> TV[SEM=...]	NP[SEM=...]	
VP[SEM=...]	-> DTV[SEM=...]	NP[SEM=...]	NP[SEM=...]
TV[SEM=...]	-> 'chases'		
DTV[SEM=...]	-> 'gives'		

Spørsmål 5.2 (10%)

Vi vil også inkludere negasjon. For eksempel ønsker vi at setning (i) får representasjonen (j).

i) Socks does not chase Fido

j) - *chase(s, f)*

Utvid og modifier grammatikken til å gi denne representasjonen. Forsikr deg om at du får riktige grammatiske begrensninger. Grammatikken skal ikke generere:

k) Socks does not chases Fido

l) Socks chase Fido

SLUTT