

UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i: INF2820 Datalingvistik
Eksamensdag: 8. juni 2011
Tid for eksamen: 14.30-18.30
Oppgavesettet er på ... side(r) 4
Vedlegg: Ingen
Tillatte hjelpemidler: Ingen

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

Der du synes det mangler forutsetninger,
gjør dine egne forutsetninger og redegjør for dem!

1 Kontekstfri grammatik og parsing (45%)

- (a) **(Teller 10%)** I engelsk skrives sammensatte substantiv oftest som to ord, som i “mountain bike”. Siden det første av de to substantivene kan modifieres, som i “natural language processing” er det naturlig å betrakte dette som et syntaktisk fenomen. Følgende lille eksempelgrammatikk, kall den G1, er ment å fange dette fenomenet.

- i) $S \rightarrow NP VP$
- ii) $NP \rightarrow DET N$
- iii) $VP \rightarrow TV NP$
- iv) $N \rightarrow N N$
- v) $N \rightarrow A N$
- vi) $NP \rightarrow Mary|John$
- vii) $DET \rightarrow a|the|her|his$
- viii) $TV \rightarrow bought|loves|misses$
- ix) $N \rightarrow bike|jersey|mountain|sleeve|brake|$
- x) $A \rightarrow long|hydraulic|knitted|expensive|steep$

Tegn opp alle trærne grammatikken tilskriver setningen

(1) *Mary bought a mountain bike hydraulic brake.*

- (b) **(Teller 5%)** Noen parsere vil ha problemer med denne grammatikken. Hvilke parsere? Forklar kort hva problemet består i.
- (c) **(Teller 10%)** En del konstruksjoner kan bli svært flertydige med denne grammatikken. For eksempel vil setning (2) få 7 forskjellige analyser. Du skal ikke vise dette.

(2) *John bought a mountain bike long sleeve jersey.*

Denne flertydigheten taler for å bruke en tabellparser når vi skal avgjøre om setninger aksepteres av grammatikk G1. Gjør et fullstendig CKY-parse av (2). Dette kan du gjøre ved å lage en tabell og fylle ut rutene med symboler. Nummerer symbolene etter hvert som de legges inn i tabellen slik at det er mulig å se hvordan du har gått frem.

- (d) **(Teller 10%)** Vis hvordan en chart-parser for denne grammatikken aksepterer setning (3).

(3) *John misses his expensive bike.*

- (e) **(Teller 10%)** La G2 være grammatikken vi får ved å fjerne regel (iv) $N \rightarrow N N$. Er språket beskrevet av G2 regulært? Er språket beskrevet av hele grammatikken G1 regulært? Begrunn svarene kort.

2 Typehierarkier og multippel arv (20%)

- (a) (**Teller 5%**) Det følgende er en beskrivelse av et typehierarki (tatt fra en av våre grammatikker) i TDL (Type Description Language). Tegn grafen som svarer til dette hierarkiet, dvs. tegn et diagram over super- og subtyperelasjoner som strekker seg nedover fra typen ***top***. For denne oppgaven ser vi bort fra alle deklarasjoner av veldefinerte trekk ('appropriate features') på typene.

```
pos := *top*.
agr-pos := pos.
det := agr-pos.
modable := agr-pos.
noun := modable.
verb := modable.
premodifier := pos.
postmodifier := pos.
prep := postmodifier.
adj := premodifier.
adv := premodifier & postmodifier.
```

- (b) (**Teller 5%**) Er det noen par av typer som ikke kan unificeres i dette hierarkiet, dvs. er det noen par av typer som ikke har en største nedre grense ('greatest lower bound')? I så fall, gi et eksempel eller to på slike par.
- (c) (**Teller 10%**) Beskriv med to eller tre setninger tolkningen av den vertikale dimensjonen av typehierarkiet, dvs. hva betyr det for en type å være subtype av en annen type? I lys av dette, beskriv to av de følgende fire termene med en setning hver: *spesifisitet* ('specificity'), *arv* ('inheritance'), *monotonitet* ('monotonicity'), eller *subsumpsjon* ('subsumption').

3 Unifikasjonsbasert grammatikk (15%)

- (a) (**Teller 10%**) Husk at alle de unifikasjonsbaserte grammatikkene for engelsk som vi utviklet gjennom semesteret (med TDL formalismen) inneholdt reglene kalt *specifier-head* og *head-complement*. Vis de (tilnærmete) trekkstrukturerepresentasjonene for disse reglene. Forklar de tre viktigste trekkene som er involvert, de relevante føringene ("constraints"), og overføring av informasjon mellom døtrene og mora i hver regel. For hver regel, gi eksempler på to fraser som er bygget med hjelp av denne regelen.
- (b) (**Teller 5%**) Nevn minst to fenomen i naturlige spørk som er (i beste fall) ubehagelige å beskrive i en ren kontekstfri grammatikk (CFG). Gi noen få eksempler, antyd det relevante spennet av variasjon i hvert tilfelle, og oppsummer i et par setninger hvorfor disse er en utfordring for en ren CFG-analyse. Med utgangspunkt i disse fenomenene, hva er hovedgevinsten ved i stedet å bruke en unifikasjonsbasert grammatikk (med toppen to setninger)?

4 Formell morfologi og syntaks (20%)

- (a) **(Teller 10%)** I typehierarkiene våre lagde vi et skarpt skille mellom objekter av type **lexeme** og av type **word**. Videre delte vi morfologiske regler inn i avledningsregler ('derivational rules') og bøyingsregler ('inflectional rules'). Karakteriser med et par setninger hovedskillet mellom avledning og bøyning. Gi ett eksempel på hver type av morfologiske regler (for engelsk) og forklar kort hva slags input (dvs. argument) hver regel aksepterer, og hva slags output den produserer.
- (b) **(Teller 5%)** Er det flertydighet i analysen av den morfologiske strukturen til den engelske formen *barking*? Skisser alle mulige morfologiske analyser av denne formen med et uformelt språk (dvs. bruk passende forkortelse for syntaktiske og morfologiske kategorier).
- (c) **(Teller 5%)** Skisser kontituenttrær for hver av de mulige lesningene av følgende setning:

(4) *The fierce cat chased the dog near the aardvark.*

Bruk forkortete merkelapper som 'Det', 'N', 'NP', 'VP', osv. på nodene. Merk i tillegg hver gren med om konstituenten den dominerer opptrer som en *head*, *specifier*, *complement*, eller *modifier*.

SLUTT