

<b>DOKTORAND:</b>	Gisle Ytrestøl
<b>GRAD:</b>	Philosophiae doctor
<b>FAKULTET:</b>	Det matematisk-naturvitenskapelige fakultet
<b>INSTITUTT:</b>	Institutt for informatikk
<b>FAGOMRÅDE:</b>	Språkteknologi
<b>VEILEDERE:</b>	Stephan Oepen, Jan Tore Lønning, Joakim Nivre
<b>DISPUTASDATO:</b>	16. august 2012
<b>AVHANDLINGENS TITTEL:</b>	<i>Transition-Based Parsing for Large-Scale Head-Driven Phrase Structure Grammars</i>

Ei datamaskinell handsaming av tekst vert nytta i ei rekkje samanhengar i dag. For dei fleste vil bruken av stavekontroll og web-søk vere dei mest openbare bruksområda for språkteknologi i kvardagen. Denne avhandlinga presenterer ein strategi som gjer det mogleg å prosessere store mengder tekst på svært kort tid, samstundes som ein produserer ein svært detaljert, semantisk rik analyse av teksten.

Ein semantisk rikare analyse kan ha ei rekkje føremoner, blant anna innan automatisk omsetjing frå eit språk til eit anna, fordi ein semantisk rik analyse gjerne er betre i stand til å abstrahere over det intenderte innhaldet i setninga, og er mindre låst til ordvalet som er nytta.

I avhandlinga undersøker vi bruken av overgangsbasert prosessering av språk, der ein les inn eitt og eitt ord, og suksessivt byggjer opp ein semantisk analyse av setninga basert på orda som er innlesne, og på ei hypotese om kva ord som vil kome seinare i setninga. Denne strategien skil seg frå ei tradisjonell prosessering av innhaldet i ei setning, der ein ser på heile setninga under eitt. Ein kan seie at ei ord-for-ord-prosessering også ligg nærare den menneskelege, kognitive handsaminga av språk, då vi som menneske ikkje ventar til samtalepartneren er ferdig med ei heil setning før vi tek til å analysere kva setninga tyder.

Overgangsbasert prosessering har synt seg å vere ein svært tidseffektiv strategi for ei rekkje grammatikkar. Denne avhandlinga evaluerer denne strategien for English Resource Grammar (ERG), som er ein semantisk detaljert grammatikk for engelsk. Den overgangsbaserte prosesseringsalgoritmen som vert presentert i denne avhandlinga er over 100 gonger snøggare enn den per i dag mest nytta prosesseringsstrategien for ERG. Avhandlinga demonstrerer at ein kan oppnå både høg effektivitet og presisjon også for svært detaljerte, semantiske grammatikkar, noko som kan gjere det attraktivt å nytte slike grammatikkar også for bruksområde som set store krav til effektivitet.

