

NODALIDA '93

*Proceedings of
'9:e Nordiska Datalogistikdagarna'
Stockholm 3–5 June 1993*

Robert Eklund, editor

Stockholm 1994

This volume was sponsored by:
Humanistisk-Samhällsvetenskapliga Forskningsrådet (HSFR)
Skriptor International AB
Department of linguistics, Stockholm University

Typeset and layout:
Robert Eklund

Published by:
Department of Linguistics, Computational Linguistics
Stockholm University, S-106 91 Stockholm, Sweden

Printed by:
Akademitryck AB, Edsbruk 1994

ISBN 91-7153-262-5

Contents

<i>Preface</i>	7
Lars Ahrenberg <i>Topological frames in sign-based grammars</i>	9
Jan Anward <i>Pieces for a Global Puzzle</i>	19
Björn Beskow <i>System Architecture and Control in the Multra System</i>	41
Benny Brodda <i>Automatic Tagging of Turns in the London-Lund Corpus with Respect to Type of Turn</i>	51
Douglas Cutting <i>Porting a Stochastic Part-of-Speech Tagger to Swedish</i>	65
Martin Eineborg & Björn Gambäck <i>Tagging Experiments Using Neural Networks</i>	71
Robert Eklund <i>A Probabilistic Word Class Tagging Module Based On Surface Pattern Matching</i>	83
Björn Gambäck <i>On Implementing Swedish Tense and Aspect</i>	97
Steffen Leo Hansen <i>Reasoning with a Domain Model</i>	111

Peter Ingels <i>Robust Parsing with Charts and Relaxation</i>	123
Per Anker Jensen, Bodil Nistrup Madsen, Annie Stahél & Carl Vikner <i>From Semantic Representations to SQL Queries</i>	133
Jussi Karlgren, Björn Gambäck & Christer Samuelsson <i>Clustering Sentences — Making Sense of Synonymous Sentences</i>	143
Arne Larson & Magnus Merkel <i>Semiotics at Work: Technical Communication and Translation in a Multilingual Corporate Environment</i>	155
Joakim Nivre <i>Pragmatics Through Context Management</i>	165
Torbjørn Nordgård <i>On GB Parsing and Semantic Interpretation</i>	175
Ole Norling-Christensen <i>Methods and Tools for Corpus Lexicography</i>	187
Claus Povlsen <i>Natural language processing in dialogue systems with spoken input</i>	197
Björn Rauch <i>Automatisk igenkänning av nominalfraser i löpande text</i>	207
Atle Ro <i>Interlanguage and Set Theory</i>	217
Christer Samuelsson <i>Morphological Tagging Based Entirely on Bayesian Inference</i> ...	225
Peter Seipel <i>Vad jag i min verksamhet som rättsinformatiker och jurist skulle vilja att datorlingvistiken bidrog med</i>	239
Annie Stahél & Helle Wegener <i>Domain Modeling and Knowledge Structures</i>	253

Anna Sågvall Hein <i>Preferences and Linguistic Choices in the Multra Machine Translation System.....</i>	267
Torben Thrane <i>Constituency and Semantic Interpretation.....</i>	277
Marta Thunes <i>Machine Translation Strategies: A Comparison of F-Structure Transfer and Semantically Based Interlingua.....</i>	291
Atro Voutilainen <i>A Noun Phrase Parser of English.....</i>	301
Margareta Westman <i>Vad jag i min verksamhet som språkvårdare skulle vilja att datorlingvistiken bidrog med.....</i>	311
Jordan Zlatev <i>From English to PFO: A Formal Semantic Parser.....</i>	317
List of Participants.....	331
Conference Program.....	335

Preface

Vi datorlingvister har egentligen alltid vetat om det men nu (äntligen) har omvärlden också så sakta börjat komma till insikt om det, nämligen att datorlingvistiken är ett nyckelområde för det som i dag benämns informationsteknologi, IT. För bara ett år sedan var det väl bara ett litet fåtal specialister som hade hört talas om IT, men nu plötsligt har intresset för området formligen exploderat. Den svenska regeringen har nyligen lagt fram ett visonärt program om storsatsning på IT i en nära framtid – "elektroniska highways" är slagordet – och tidningarna är fyllda med fantastiska visioner om hur informationsteknologin skall omdana vår värld och lägga grunden för vårt framtida välstånd. Med hjälp av de elektroniska nätverken kommer vi ha tillgång till all världens vetenskapliga, tekniska och kulturella information bokstavligen i våra fingerspetsar. Med multimediatekniken kan vi få bilder och ljud från alla jordens hörn direkt i våra vardagsrum, et cetera, et cetera. En fantastisk ny värld finns bara runt hörnet.

Men – det finns alltid ett men – vägen dithän är inte så där alldeles lätt. Rent tekniskt är det väl kanske inte så väldigt långt dit (men med tanke på att universitetsnäten fortfarande efter 20 år bara klarar sjubits-ASCII är åtminstone inte jag reservationslöst imponerad av den fart med vilket den tekniska utvecklingen skrider fram inom det här området), men för att hitta all den information som vi vet finns "där ute" så måste vi ha ofantligt mycket intelligentare användargränssnitt än vad vi i dag har – Gopher i sin nuvarande form är inte framtidens NLI. Vi måste ha söksystem som förstår att söka efter den information som användaren frågar efter och inte bara söka efter de informationskällor som råkar innehålla de ord som frågaren använt i sin sökfråga. Söksystemen måste också kunna finna den relevanta informationen oavsett vilket språk den finns representerad i och informationen skall kunna presenteras för användaren på ett språk som denne förstår. Utöver allt detta så finns den ytterligare dimensionen att systemen i tillämpliga fall skall kunna förstå och ta emot muntliga instruktioner och likaså i tillämpliga fall leverera sina svar i form av talat output.

För alla tillämpningar som nämnts ovan – frågebesvarande system, informationssökning, automatisk översättning samt taligenkänning och talgenerering – så finns det i dag ett nästan obegränsat behov av färdiga teknologier, det vet alla. Vi datorlingvister vet också att färdiga lösningar rätt och slätt inte föreligger, utan att det krävs ett långt och mödosamt arbete för att få något så när hyggliga lösningar. De nämnda tillämpningsområdena har alla en komponent av datorlingvistisk karaktär i sig, och det är inte en tillfällighet att för vart och ett av dem så finns det presentationer med klara implikationer för just det området i denna volym.

Även om de nämnda tillämpningarna är utomordentligt viktiga så får vi naturligtvis inte sälja vår själ. Vi har också ett inomvetenskapligt uppdrag. Men som jag ser det så behöver det inte finnas en motsättning mellan det samhälleliga behovet och detta uppdrag. Man kan fråga sig vad målet egentligen är för datorlingvistiken – om man nu kan tala om ett mål. En central frågeställning måste i alla fall vara att åstadkomma en semantisk modell som på ett något så när entydigt och fullständigt sätt möjliggör en automatisk härledning av betydelseinnehållet i första hand i en godtycklig mening ("sentence") och på sikt också också i en sekvens av sammanhängande meningar, alltså i en text. Inom parentes sagt, så är detta också en central frågeställning för den teoretiska lingvistikens.

Lyckas vi med det uppdraget, ja då har vi också början till lösningar för de nämnda utomvetenskapliga problemställningarna.

Benny Brodda