

**THE 11<sup>TH</sup> NORDIC CONFERENCE ON  
COMPUTATIONAL LINGUISTICS**

Copenhagen, 28-29 January 1998

**NODALIDA '98  
PROCEEDINGS**

Center for Sprogteknologi  
and  
Department of General and Applied Linguistics (IAAS)  
University of Copenhagen  
Njalsgade 80  
DK-2300 Copenhagen S, Denmark

ISBN 87-90708-01-6

©NODALIDA '98 and the authors  
c/o Center for Sprogteknologi, Njalsgade 80, DK-2300 Copenhagen S

Thanks go to all members of the programme committee for their unvaluable help reading the papers for this conference

Copenhagen, March 1998

Bente Maegaard.

# CONTENT

Welcome <i>John Kuhlmann Madsen</i> .....	1
Introduction <i>Bente Maegaard</i> .....	3
LMT at Tivoli Gardens <i>Arendse Bernth, Michael McCord</i> .....	4
Drug Terminology. A multilingual term database. The AVENTINUS project <i>Christian Sjøgreen</i> .....	13
AVENTINUS, GATE and Swedish Lingware <i>Dimitrios Kokkinakis</i> .....	22
Norwegian Computational Lexicon ( <i>NorKompLeks</i> ) <i>Torbjørn Nordgård</i> .....	34
Structural Lexical Heuristics in the Automatic Analysis of Portuguese <i>Eckhard Bick</i> .....	44
An HPSG Marking Analysis of Danish Determiners and Clausal Adverbials <i>Costanza Navarretta, Anne Neville</i> .....	57
A Chart-Based Framework for Grammar Checking. Initial Studies <i>Anna Sågvall Hein</i> .....	68
CP-UDOG: An Algorithm for the Disambiguation of Compound Participles in Danish <i>Jens Ahlmann Hansen, Poul Søren Kjærsgaard</i> .....	81
Evaluation of the Syntactic Parsing Performed by the ENGCG Parser <i>Klas Prytz</i> .....	87
CATCH: A Program for Developing World Wide Web CALL Material <i>Erik F. Tjong Kim Sang</i> .....	94
Assembling a Balanced Corpus from the Internet <i>Johan Dewe, Jussi Karlgren, Ivan Bretan</i> .....	100
Peeking Into the Danish Living Room. Internet access to a large speech corpus <i>Peter Juel Henriksen</i> .....	109

Extraction of Translation Equivalents from Parallel Corpora <i>Jörg Tiedemann</i> .....	120
The Construction of a Tagged Danish Corpus <i>Thomas Bilgram, Britt Keson</i> .....	129
Linguistics isn't always the answer: Word comparison in computational linguistics <i>Lars Borin</i> .....	140
Logic for Part-of-Speech Tagging and Shallow Parsing <i>Torbjörn Lager</i> .....	152
A statistical and structural approach to extracting collocations likely to be of relevance in relation to an LSP sub-domain text <i>Bjarne Blom</i> .....	160
A Reasonably Language Independent, Heuristic Algorithm for the Marking of Names in Running Texts <i>Benny Brodda</i> .....	168
Some Results Regarding Tree Homomorphic Feature Structure Grammar and the Empty String <i>Tore Burheim</i> .....	176
Teaching and learning computational linguistics in an international setting <i>Koenraad de Smedt</i> .....	186

# Welcome

**John Kuhlmann Madsen**

Dean of the Faculty of Humanities, University of Copenhagen

Ladies and gentlemen, dear colleagues,

I have been told that this is the second time you have honoured our Faculty with a Nodalida meeting. The first one took place in the now – computationally speaking, if you allow me – almost prehistoric year of 1979.

Permit me to take some minutes of your precious time to talk about language studies in rather general terms.

The seventies were the heyday of massive student interest in formal linguistics. In later years we discovered a rather disturbing movement away from the formal aspects of language study, more or less corresponding to the exodus from the maths and physics departments. A couple of years ago we made a small survey of the modern language departments, focusing on the themes of the minor theses, (which are compulsory for obtaining the MA degree). It was shocking for me to learn that in two consecutive academic years there was no minor thesis on theoretical aspects of grammar, phonology, or semantics, in English, French, German, or Spanish. The closest we got to linguistics in language studies were the essays on sociological aspects of some speech community in some distant regional areas, and the like. We can detect the same tendency when applicants make their first choice: first in line we have psychology, general literature, film and media studies, history of art, philosophy, theatre studies, and history. I am sure you get the overall picture: the main interest lies within hermeneutics and aesthetics. According to fellow deans from other arts and humanities faculties the results of our survey are not exceptional. We are facing a similar trend in the nineties. The way I see it, we are talking about a trend that must be halted; not to save our language and linguistics departments, but simply because the small Nordic countries cannot do without top-level language education. We like to think that we speak foreign languages better than the foreigners do. I must confess that I am not so sure of the correctness of this opinion. On the other hand, you need not be a formal linguist to speak a foreign language beautifully, but, if only a tiny minority feels attracted by the study of grammar, I fear teachers will soon experience some difficulty correcting their pupils.

I feel that in promoting linguistic language studies we are moving upstream. I am confident that it will be worth our while. However, we must be realistic in our efforts. Some years ago our Minister of Education tried to force students in specific directions. He succeeded only in causing severe damage to many subject areas, especially within the humanities and, ironically enough, in their language departments. We still suffer from his peculiar ideology. The way I see it, the study of language must regain its attraction on its own terms. The jobs are there, the teachers are – still – there, but how do we attract students other than those who have been *au pairs*? The usual trick is to sit back and leave it to the secondary school system. The interest must be created in secondary school, but why leave it all to others. Why do we not collaborate with the language

teachers and their principals to at least gain some understanding of the situation? For our part that is exactly what we are doing now. In mid February headmasters from the eastern part of the country will meet at our Faculty to start a discussion on how we can get to understand each other's needs. Hopefully, that meeting will lead to further meetings on specific subjects, including language and linguistics.

Last year we noticed an increased interest in linguistics among our applicants. Hoping for a turn of the tide I look forward to welcoming many future linguists – computational or not – at our Faculty. I am confident that your work – here and elsewhere – will be inspiring both for your fellow specialists and for the study of language in general.

It's my pleasure to bid you welcome to two days of mutual inspiration.

Thank you

# Introduction

**Bente Maegaard**

Programme chair

The Nordic conference of Computational Linguistics (Nordiske Datalingvistikdage, NODALIDA) is a biannual event, directed to the researchers, students and industrialists in the field of computational linguistics as well as to all interested parties. Most of the participants are from the Nordic countries.

Papers were invited from all areas of computational linguistics, including but not restricted to: morphological, syntactic, semantic and textual analysis and generation, speech processing, machine translation, computational lexicography and processing of monolingual or multilingual text corpora.

The programme featured many of these with a specific focus on corpus related work, i.e. the creation of corpora, and their use for different purposes. Research and development concerning computational lexica was discussed in a couple of papers. Another feature worth mentioning is that several papers discussed the use of the Internet in connection with computational linguistic applications: making corpora available on the Internet, making educational software packages available on the Internet etc.

I wish to thank warmly the programme committee who with their expertise and their readiness to provide feedback to the authors and to me very fast, made it possible to provide a very interesting programme. The programme committee consisted of Kimmo Koskenniemi, Helsinki, Peter Molbæk Hansen, Copenhagen, Torbjørn Nordgård, Trondheim, and Anna Sågvall Hein, Uppsala.

A panel discussion on *The Nordic languages in the Information Society - a responsibility for computational linguistics and computational linguists?* showed that the interest from public funding agencies in supporting computational linguistics has been and is different in the various Nordic countries. The Nordic Council discussed the language issue last May, and it was suggested that a coordinated approach be made to the Nordic Council concerning the protection and reinforcement of the Nordic languages which are 'less used' on a world basis.