PROCEEDINGS of the 16th Nordic Conference of Computational Linguistics

NODALIDA-2007

Editors: Joakim Nivre, Heiki-Jaan Kaalep, Kadri Muischnek and Mare Koit



Editors: Joakim Nivre, Heiki-Jaan Kaalep, Kadri Muischnek and Mare Koit J. Liivi 2, 50409 Tartu, Estonia

Published by the University of Tartu University of Tartu, Ülikooli 18, 50090 Tartu, Estonia Tartu, 2007

ISBN 978-9985-4-0514-7 (CD-ROM) ISBN 978-9985-4-0513-0 (online)

Contents

Contents	iii
Preface	ix
Committees	xi
Program Committee	xi
Local Organization Committee	xi
Reviewers	xii
Conference program	xiii
I Invited talks	1
DIANA F. MCCARTHY Evaluating Automatic Approaches for Word Meaning Discovery and Disambiguation using Lexical Substitution	n- 2
WALTER DAELEMANS Text Analysis and Machine Learning for Stylome rics and Stylogenetics	t- 3
II Regular papers	4
TANEL ALUMÄE	
Automatic Compound Word Reconstruction for Speech Recognition Compounding Languages	of 5
Guntis Bārzdiņš, Normunds Grūzītis, Gunta Nešpore and Baie Saulīte	A
Dependency-Based Hybrid Model of Syntactic Analysis for the Lag guages with a Rather Free Word Order	n- 13
ECKHARD BICK AND LARS NYGAARD	
Using Danish as a CG Interlingua: A Wide-Coverage Norwegian-Eng Machine Translation System	<mark>lish</mark> 21

JANNE BONDI JOHANNESSEN, KRISTIN HAGEN, JOEL JAMES PRIESTLEY AND LARS NYGAARD An Advanced Speech Corpus for Norwegian	29
MARKUS BORG Time Extraction from Real-time Generated Football Reports	37
Inger Ekman and Kalervo Järvelin Spoken Document Retrieval in a Highly Inflectional Language	44
EVA FORSBOM Inducing Baseform Models from a Swedish Vocabulary Pool	51
OLGA GERASSIMENKO, MARE KOIT, ANDRIELA RÄÄBIS AND KRISTA STRA	AND-
SON Achieving Goals in Collaboration: Analysis of Estonian Institutional Calls	59
KĀRLIS GOBA AND ANDREJS VASIĻJEVS Development of Text-To-Speech system for Latvian	67
JONAS GRANFELDT AND PIERRE NUGUES Evaluating Stages of Development in Second Language French: A Mach Learning Approach	<mark>ine-</mark> 73
KARIN HARBUSCH AND GERARD KEMPEN Clausal Coordinate Ellipsis in German: The TIGER Treebank as a Source of Evidence	81
Martin Hassel and Jonas Sjöbergh Widening the HolSum Search Scope	89
HANS HJELM Identifying Cross Language Term Equivalents Using Statistical Ma- chine Translation and Distributional Association Measures	97
RICHARD JOHANSSON AND PIERRE NUGUES Extended Constituent-to-Dependency Conversion for English	105
TOOMAS KIRT AND ENE VAINIK Comparison of the Methods of Self-Organizing Maps and Multidimen- sional Scaling in Analysis of Estonian Emotion Concepts	113
HANJING LI, TIEJUN ZHAO, SHENG LI, JIYUAN ZHAO The Extraction of Trajectories from Real Texts Based on Linear Clas- sification	121

 BEATA B. MEGYESI AND BENGT DAHLQVIST The Swedish-Turkish Parallel Corpus and Tools for its Creation NICOLAS MORALES, JOHN H. L. HANSEN, DOROTEO T. TOLEDANO A JAVIER GARRIDO Multivariate Cepstral Feature Compensation on Band-limited Data Robust Speech Recognition VICTORIA ROSÉN AND KOENRAAD DE SMEDT Theoretically Motivated Treebank Coverage TUOMO SAARNI, JUSSI HAKOKARI, TAPIO SALAKOSKI, JOUNI ISOA AND OLLI AALTONEN Utterance-Initial Duration of Finnish Non-Plosive Consonants INGUNA SKADIŅA, ANDREJS VASIĻJEVS, DAIGA DEKSNE, RAIVIS SKA AND LINDA GOLDBERGA Comprehension Assistant for Languages of Baltic States RICHARD SOCHER, CHRIS BIEMANN AND RAINER OSSWALD Combining Contexts in Lexicon Learning for Semantic Parsing ANDERS SØGAARD Polynomial Charts For Totally Unordered Languages
JAVIER GARRIDO Multivariate Cepstral Feature Compensation on Band-limited Data Robust Speech Recognition VICTORIA ROSÉN AND KOENRAAD DE SMEDT Theoretically Motivated Treebank Coverage TUOMO SAARNI, JUSSI HAKOKARI, TAPIO SALAKOSKI, JOUNI ISOA AND OLLI AALTONEN Utterance-Initial Duration of Finnish Non-Plosive Consonants INGUNA SKADIŅA, ANDREJS VASIĻJEVS, DAIGA DEKSNE, RAIVIS SKA AND LINDA GOLDBERGA Comprehension Assistant for Languages of Baltic States RICHARD SOCHER, CHRIS BIEMANN AND RAINER OSSWALD Combining Contexts in Lexicon Learning for Semantic Parsing ANDERS SØGAARD Polynomial Charts For Totally Unordered Languages
Robust Speech Recognition VICTORIA ROSÉN AND KOENRAAD DE SMEDT Theoretically Motivated Treebank Coverage TUOMO SAARNI, JUSSI HAKOKARI, TAPIO SALAKOSKI, JOUNI ISOA AND OLLI AALTONEN Utterance-Initial Duration of Finnish Non-Plosive Consonants INGUNA SKADIŅA, ANDREJS VASIĻJEVS, DAIGA DEKSNE, RAIVIS SKA AND LINDA GOLDBERGA Comprehension Assistant for Languages of Baltic States RICHARD SOCHER, CHRIS BIEMANN AND RAINER OSSWALD Combining Contexts in Lexicon Learning for Semantic Parsing ANDERS SØGAARD Polynomial Charts For Totally Unordered Languages MARTIN VOLK AND FRIDA TIDSTRÖM
Theoretically Motivated Treebank Coverage TUOMO SAARNI, JUSSI HAKOKARI, TAPIO SALAKOSKI, JOUNI ISOA AND OLLI AALTONEN Utterance-Initial Duration of Finnish Non-Plosive Consonants INGUNA SKADIŅA, ANDREJS VASIĻJEVS, DAIGA DEKSNE, RAIVIS SKA AND LINDA GOLDBERGA Comprehension Assistant for Languages of Baltic States RICHARD SOCHER, CHRIS BIEMANN AND RAINER OSSWALD Combining Contexts in Lexicon Learning for Semantic Parsing ANDERS SØGAARD Polynomial Charts For Totally Unordered Languages MARTIN VOLK AND FRIDA TIDSTRÖM
 TUOMO SAARNI, JUSSI HAKOKARI, TAPIO SALAKOSKI, JOUNI ISOA AND OLLI AALTONEN Utterance-Initial Duration of Finnish Non-Plosive Consonants INGUNA SKADIŅA, ANDREJS VASIĻJEVS, DAIGA DEKSNE, RAIVIS SKA AND LINDA GOLDBERGA Comprehension Assistant for Languages of Baltic States RICHARD SOCHER, CHRIS BIEMANN AND RAINER OSSWALD Combining Contexts in Lexicon Learning for Semantic Parsing ANDERS SØGAARD Polynomial Charts For Totally Unordered Languages MARTIN VOLK AND FRIDA TIDSTRÖM
AND OLLI AALTONEN Utterance-Initial Duration of Finnish Non-Plosive Consonants INGUNA SKADIŅA, ANDREJS VASIĻJEVS, DAIGA DEKSNE, RAIVIS SKA AND LINDA GOLDBERGA Comprehension Assistant for Languages of Baltic States RICHARD SOCHER, CHRIS BIEMANN AND RAINER OSSWALD Combining Contexts in Lexicon Learning for Semantic Parsing ANDERS SØGAARD Polynomial Charts For Totally Unordered Languages MARTIN VOLK AND FRIDA TIDSTRÖM
 INGUNA ŠKADIŅA, ANDREJS VASIĻJEVS, DAIGA DEKSNE, RAIVIS ŠKA AND LINDA GOLDBERGA Comprehension Assistant for Languages of Baltic States RICHARD SOCHER, CHRIS BIEMANN AND RAINER OSSWALD Combining Contexts in Lexicon Learning for Semantic Parsing ANDERS SØGAARD Polynomial Charts For Totally Unordered Languages MARTIN VOLK AND FRIDA TIDSTRÖM
AND LINDA GOLDBERGA Comprehension Assistant for Languages of Baltic States RICHARD SOCHER, CHRIS BIEMANN AND RAINER OSSWALD Combining Contexts in Lexicon Learning for Semantic Parsing ANDERS SØGAARD Polynomial Charts For Totally Unordered Languages MARTIN VOLK AND FRIDA TIDSTRÖM
RICHARD SOCHER, CHRIS BIEMANN AND RAINER OSSWALD Combining Contexts in Lexicon Learning for Semantic Parsing ANDERS SØGAARD Polynomial Charts For Totally Unordered Languages MARTIN VOLK AND FRIDA TIDSTRÖM
Combining Contexts in Lexicon Learning for Semantic Parsing ANDERS SØGAARD Polynomial Charts For Totally Unordered Languages MARTIN VOLK AND FRIDA TIDSTRÖM
Polynomial Charts For Totally Unordered Languages Martin Volk and Frida Tidström
Comparing French PP-attachment to English, German and Swedis
PONTUS WÄRNESTÅL, LARS DEGERSTEDT AND ARNE JÖNSSON Interview and Delivery: Dialogue Strategies for Conversational H ommender Systems
III Student papers
BJÖRN ANDRIST AND MARTIN HASSEL Linguistically Fuelled Text Similarity

DAVE COCHRAN Unmediated Data-Oriented Generation	216
KARIN FRIBERG Decomposing Swedish Compounds Using Memory-Based Learning	224
MARIA HOLMQVIST Memory-based Learning of Word Translation	231
FREDRIK JØRGENSEN Clause Boundary Detection in Transcribed Spoken Language	235
FREDRIK JØRGENSEN The Effects of Disfluency Detection in Parsing Spoken Language	240
Anders Nøklestad and Åshild Søfteland Tagging a Norwegian Speech Corpus	245
ANTON RAGNI Initial Experiments with Estonian Speech Recognition	249
MARIANNE SANTAHOLMA Grammar Sharing Techniques for Rule-based Multilingual NLP Sys- tems	253
MARIANNE STARLANDER Using a Wizard of Oz as a Baseline to Determine which System Archi- tecture is the Best for a Spoken Language Translation System	261
MARGUS TREUMUTH A Method for Recognizing Temporal Expressions in Estonian Natural Language Dialogue Systems	265
IV Posters	269
LARS AHRENBERG LinES: An English-Swedish Parallel Treebank	270
DANIEL BOLANOS AND WAYNE H. WARD Posterior Probability Based Confidence Measures Applied to a Chi- dren's Speech Reading Tracking System	274
MARK FISHEL, HEIKI-JAAN KAALEP AND KADRI MUISCHNEK Estonian-English Statistical Machine Translation: the First Results	278

JOHAN HALL, JOAKIM NIVRE AND JENS NILSSON A Hybrid Constituency-Dependency Parser for Swedish	284
ERLA HALLSTEINSDÓTTIR, THOMAS ECKART, CHRIS BIEMANN, UWE QU AND MATTHIAS RICHTER	
Íslenskur Orðasjóður — Building a Large Icelandic Corpus	288
HARALD HAMMARSTRÖM A Survey and Classification of Methods for (Mostly) Unsupervised Lea ing	rn- 292
OLE HARTVIGSEN, ERIK HARBORG, TORE AMBLE AND MAGNE H. JOHNS Marvina — A Norwegian Speech-Centric, Multimodal Visitors' Guide	
PETER JUEL HENRICHSEN A Norwegian Letter-to-Sound Engine with Danish as a Catalyst	305
SIMON KEIZER AND ROSER MORANTE Dialogue Simulation and Context Dynamics for Dialogue Management	310
KIMMO KETTUNEN Managing Keyword Variation with Frequency Based Generation of Word Forms in IR	318
WANWISA KHANARAKSOMBAT AND JONAS SJÖBERGH Developing and Evaluating a Searchable Swedish-Thai Lexicon	324
DIMITRIOS KOKKINAKIS AND ANDERS THURIN Identification of Entity References in Hospital Discharge Letters	329
DIMITRIOS KOKKINAKIS, MARIA TOPOROWSKA GRONOSTAJ, CATALINA HALLETT AND DAVID HARDCASTLE	
Lexical Parameters, Based on Corpus Analysis of English and Swedish Cancer Data, of Relevance for NLG	333
MIKKO LOUNELA Anatomy of an XML-based Text Corpus Server	337
LYA MEISTER Perceptual Assessment of the Degree of Russian Accent	345
MAGNUS MERKEL AND JODY FOO Terminology Extraction and Term Ranking for Standardizing Term Banks	349

JYRKI NIEMI AND KIMMO KOSKENNIEMI Representing Calendar Expressions with Finite-State Transducers that Bracket Periods of Time on a Hierachical Timeline	355
HELEN NIGOL Parsing Manually Detected and Normalized Disfluencies in Spoken Es- tonian	363
LIISI PIITS, MEELIS MIHKLA, TÕNIS NURK AND INDREK KIISSEL Designing a Speech Corpus for Estonian Unit Selection Synthesis	367
INES REHBEIN AND JOSEF VAN GENABITH Evaluating Evaluation Measures	372
JÜRGEN RIEDLER AND SERGIOS KATSIKAS Development of a Modern Greek Broadcast-News Corpus and Speech Recognition System	380
JANNE SAVELA, STINA OJALA, OLLI AALTONEN AND TAPIO SALAKOSKI Role of Different Spectral Attributes in Vowel Categorization: the Case of Udmurt	384
JONAS SJÖBERGH AND KENJI ARAKI Recreating Humorous Split Compound Errors in Swedish by Using Grammaticality	389
HÅKAN SUNDBLAD A Re-examination of Question Classification	394
TARMO TRUU, HALDUR ÕIM AND MARE KOIT Interpretation of Yes/No Questions as Metaphor Recognition	398
CENNY WENNER Rule-based Logical Forms Extraction	402
Author Index	410

Preface

Language technology research in Northern Europe is thriving. One clear sign of this is the fact that the NODALIDA conference (also known as the Nordic Conference of Computational Linguistics) has grown to the point where it is no longer manageable to let the local organization committee do all the work in putting together the program. Hence, the need for a separate program committee.

When trusted with the responsibility of chairing the program committee, the program chair made an informal survey of expectations in the community through an e-mail questionnaire sent out on the NODALI list. (Thanks to everyone who responded.) Four things emerged clearly from this survey: People wanted a high-quality technical track with review of full papers. However, people also wanted an opportunity for students to present their work, and an opportunity to get feed-back on work in progress in the form of posters. Last but not least, people wanted workshops. Hence, we give you regular paper sessions, student paper sessions, a poster session, and a full day of workshops. As a bonus, we also give you two distinguished keynote speakers, Diana McCarthy and Walter Daelemans, the first business meeting of the newly established Northern European Association for Language Technology (NEALT), and a tutorial on the Estonian language, hoping to establish a new NODALIDA tradition of local language tutorials. (Thanks to Koenraad de Smedt for this great idea.)

When issuing the call for workshops, regular papers, student papers, and posters, we were unsure whether there would be enough work going on in Northern Europe and the rest of the world to fill all the categories. The community response surpassed all our expectations. We received 41 regular paper submissions, 24 student paper submissions, 32 poster submissions, and 7 workshop proposals, for a total of 104 submissions, an all time record for NODALIDA. (Thanks to everyone who submitted their work.) Moreover, a fair share of these submissions came from countries outside our region, clearly showing that NODALIDA, while remaining a conference with a strong regional character, is also being recognized in the rest of the world. In the final program, there are 26 regular papers, 12 student papers, 26 posters, and 4 workshops. We want to thank the program committee and all our 69 reviewers for their hard work in putting together the program.

Finally, NODALIDA 2007 is special not only for being the biggest ever in terms of submissions, but also for being the first NODALIDA held in Estonia, in the beautiful city of Tartu, at one of the oldest universities in the region, founded

in 1632. We want to thank the local organization committee for all their hard work to welcome the NODALIDA participants in Tartu. We wish you all an enjoyable NODALIDA 2007!

Joakim Nivre Program Chair

NODALIDA 2007

Mare Koit and Tiit Roosmaa Local Co-Chairs NODALIDA 2007

Committees

Program Committee

Joakim Nivre (chair), Växjö University and Uppsala University Helena Ahonen-Myka, University of Helsinki Daniel Hardt, Copenhagen Business School Kristiina Jokinen, University of Helsinki and University of Tartu Pierre Nugues, Lund University Stephan Oepen, University of Oslo, NTNU Trondheim and Stanford University Patrizia Paggio, University of Copenhagen Torbjørn Svendsen, NTNU Trondheim

Local Organization Committee

Mare Koit (co-chair), University of Tartu Tiit Roosmaa (co-chair), University of Tartu Urve Talvik, University of Tartu Heli Uibo, University of Tartu Kadri Vider, University of Tartu

Reviewers

Lars Ahrenberg Ingunn Amdal Beáta Bandmann Megyesi Francis Bond Janne Bondi Johannessen Lars Borin Matthias Buch-Kromann Rolf Carlson Mathias Creutz Antoine Doucet Laila Dybkjær Helge Dyvik Eva Ejerhed Björn Gambäck Barbara Gawronska Jerneja Zganec Gros Nina Grønnum Petter Haugereid Peter Juel Henrichsen Merle Horne Hannes Högni Vilhjálmsson **Richard Johansson** Magne H. Johnsen Arne Jönsson Viggo Kann Jussi Karlgren Sabine Kirchmeier-Andersen Ola Knutsson Mare Koit Jacques Koreman Kimmo Koskenniemi Mikko Kurimo Leena Kuure Knut Kvale Juha-Pertti Laaksonen

Torbjörn Lager Birger Larsen Krister Lindén Jan Tore Lønning Ramón López-Cózar Delgado Bodil Nistrup Madsen Bente Maegaard Jean-Claude Martin Bilyana Martinovski Michael McTear Tor Andre Myrvoll Costanza Navarretta Anders Nøklestad Torbjørn Nordgård Bjarne Ørsnes Maria Teresa Pazienza **Bolette Pedersen** Jussi Piitulainen Ari Pirkola Aarne Ranta Victoria Rosén Eiríkur Rögnvaldsson Rune Sætre Anders Søgaard Markku Turunen Wim van Dommelen Martin Volk Jürgen Wedekind Stefan Werner Mats Wirén Roman Yangarber Zhang Yi Anssi Yli-Jyrä Fabio Massimo Zanzotto

Conference program NODALIDA-2007 Main Conference

Friday, May 25, 2007

- 9.00–10.30 **Opening Session**
- 9.00- 9.30 Opening
- 9.30–10.30 Invited Talk Evaluating Automatic Approaches for Word Meaning Discovery and Disambiguation using Lexical Substitution Diana F. McCarthy, University of Sussex
- 10.30–11.00 Coffee Break
- 11.00–12.30 Parallel Paper Sessions

	Parsing	Multilingual Resources and Translation	Speech Technology
11.00-	IceParser: An Incre-	Using Danish as a CG	Automatic Compound
11.30	mental Finite-State	Interlingua: A Wide-	Word Reconstruction
	Parser for Icelandic	Coverage Norwegian-	for Speech Recognition
	Hrafn Loftsson and	English Machine Trans-	of Compounding Lan-
	Eiríkur Rögnvaldsson	lation System	guages
	-	Eckhard Bick and Lars	Tanel Alumäe
		Nygaard	
11.30-	Combining Contexts in	Identifying Cross Lan-	Multivariate Cepstral
12.00	Lexicon Learning for	guage Term Equivalents	Feature Compensa-
	Semantic Parsing	Using Statistical Ma-	tion on Band-limited
	Richard Socher, Chris	chine Translation and	Data for Robust Speech
	Biemann and Rainer	Distributional Associa-	Recognition
	Osswald	tion Measures	Nicolas Morales, John
		Hans Hjelm	H. L. Hansen, Doroteo
			T. Toledano and Javier
			Garrido
12.00-	Polynomial Charts for	The Swedish-Turkish	Development of Text-
12.30	Totally Unordered Lan-	Parallel Corpus and	To-Speech System for
	guages	Tools for its Creation	Latvian
	Anders Søgaard	Beata B. Megyesi and	Kārlis Goba and An-
		Bengt Dahlqvist	drejs Vasiljevs

12.30–14.00 Lunch

14.00–15.30 Parallel Paper Sessions

	Treebanks	Information Extraction and Summarization	Phonetics and Speech
14.00-	Theoretically Motivated	Time Extraction from	Utterance-initial Du-
14.30	Treebank Coverage	Real-time Generated	ration of Finnish Non-
	Victoria Rosén and	Football Reports	plosive Consonants
	Koenraad De Smedt	Markus Borg	Tuomo Saarni, Jussi
			Hakokari, Olli Aalto-
			nen, Jouni Isoaho and
			Tapio Salakoski
14.30-	Extended Constituent-	The Extraction of Tra-	An Advanced Speech
15.00	to-Dependency Conver-	jectories from Real	Corpus for Norwegian
	sion for English	Texts Based on Linear	Janne Bondi Johan-
	Richard Johansson and	Classification	nessen, Kristin Hagen,
	Pierre Nugues	Hanjing Li, Tiejun	Joel James Priestley and
		Zhao, Sheng Li, Jiyuan	Lars Nygaard
		Zhao	
15.00-	Clausal Coordinate	Widening the HolSum	Spoken Document
15.30	Ellipsis in German: The	Search Scope	Retrieval in a Highly
	TIGER Treebank as a	Martin Hassel and	Inflectional Language
	Source of Evidence	Jonas Sjöbergh	Inger Ekman and
	Karin Harbusch and		Kalervo Järvelin
	Gerard Kempen		

15.30–16.00 Coffee Break

16.00–17.00 Poster Session 1

LinES: An English-Swedish Parallel Treebank Lars Ahrenberg

Posterior Probability Based Confidence Measures Applied to a Children's Speech Reading Tracking System Daniel Bolanos and Wayne H. Ward

Estonian-English Statistical Machine Translation: the First Results Mark Fishel, Heiki-Jaan Kaalep and Kadri Muischnek

A Hybrid Constituency-Dependency Parser for Swedish Johan Hall, Joakim Nivre and Jens Nilsson

Íslenskur Orðasjóður – Building a Large Icelandic Corpus Erla Hallsteinsdóttir, Thomas Eckart, Chris Biemann, Uwe Quasthoff and Matthias Richter

A Survey and Classification of Methods for (Mostly) Unsupervised Learning

Harald Hammarström

Marvina – A Norwegian Speech-Centric, Multimodal Visitor Guide Ole Hartvigsen, Erik Harborg, Tore Amble and Magne H. Johnsen

A Norwegian Letter-to-Sound Engine with Danish as a Catalyst Peter Juel Henrichsen

Dialogue Simulation and Context Dynamics for Dialogue Management Simon Keizer and Roser Morante

Managing Keyword Variation with Frequency Based Generation of Word Forms in IR

Kimmo Kettunen

Developing and Evaluating a Searchable Swedish-Thai Lexicon Wanwisa Khanaraksombat and Jonas Sjöbergh

Identification of Entity References in Hospital Discharge Letters Dimitrios Kokkinakis and Anders Thurin

Lexical Parameters, Based on Corpus Analysis of English and Swedish Cancer Data, of Relevance for NLG Dimitrios Kokkinakis, Maria Toporowska Gronostaj, Catalina Hallett and David Hardcastle

17.00–18.00 Poster Session 2

Anatomy of an XML-based Text Corpus Server Mikko Lounela

Perceptual Assessment of the Degree of Russian Accent Lya Meister

Terminology Extraction and Term Ranking for Standardizing Term Banks Magnus Merkel and Jody Foo

Representing Calendar Expressions with Finite-State Transducers that Bracket Periods of Time on a Hierachical Timeline Jyrki Niemi and Kimmo Koskenniemi

Parsing Manually Detected and Normalized Disfluencies in Spoken Estonian Helen Nigol

Designing a Speech Corpus for Estonian Unit Selection Synthesis Liisi Piits, Meelis Mihkla, Tõnis Nurk and Indrek Kiissel

Evaluating Evaluation Measures Ines Rehbein and Josef van Genabith

Development of a Modern Greek Broadcast-News Corpus and Speech Recognition System Jürgen Riedler and Sergios Katsikas Role of Different Spectral Attributes in Vowel Categorization: the Case of Udmurt
Janne Savela, Stina Ojala, Olli Aaltonen and Tapio Salakoski
Recreating Humorous Split Compound Errors in Swedish by Using Grammaticality
Jonas Sjöbergh and Kenji Araki
A Re-examination of Question Classification
Håkan Sundblad
Interpretation of Yes/No Questions as Metaphor Recognition
Tarmo Truu, Haldur Õim and Mare Koit
Rule-based Logical Forms Extraction
Cenny Wenner
Conference Dinner

Saturday, May 26, 2007

19.00-

9.00–10.00 **Plenary Session** Invited Talk *Text Analysis and Machine Learning for Stylometrics and Stylogenetics* Walter Daelemans, University of Antwerp

10.00–10.30 **Coffee Break**

10.30–12.30 Parallel Student Sessions

	Spoken Language	Multilingual Resources	Natural Language
	Processing	and Translation	Processing
10.30-	Clause Boundary De-	Memory-Based Learn-	Unmediated Data-
11.00	tection in Transcribed	ing of Word Translation	Oriented Generation
	Spoken Language	Maria Holmqvist	Dave Cochran
	Fredrik Jørgensen		
11.00-	The Effects of Dis-	Using Parallel Corpora	Linguistically Fuelled
11.30	fluency Detection in	to Create a Greek-	Text Similarity
	Parsing Spoken Lan-	English Dictionary with	Björn Andrist and Mar-
	guage	UPLUG	tin Hassel
	Fredrik Jørgensen	Konstantinos Charitakis	

11.30– 12.00	Tagging a Norwegian Speech Corpus Anders Nøklestad and Åshild Søfteland	Using a Wizard of Oz as a Baseline to De- termine Which System Architecture Is the Best for a Spoken Language Translation System Marianne Starlander	Decomposing Swedish Compounds Using Memory-Based Learn- ing Karin Friberg
12.00-12.30	Initial Experiments with Estonian Speech Recognition Anton Ragni	Grammar Sharing Tech- niques for Rule-based Multilingual NLP Sys- tems Marianne Santaholma	A Method for Reso- lution of Temporal Expressions in Estonian Natural Language Dia- logue Systems Margus Treumuth

12.30–14.00 Lunch 14.00–15.30 Parallel Paper Sessions

	Parsing and Transla-	Machine Learning	Dialogue
	tion		
14.00-	Dependency-Based Hy-	Comparison of the	Interview and Delivery:
14.30	brid Model of Syntactic	Self-Organizing Map	Dialogue Strategies for
	Analysis for the Lan-	and Multidimensional	Conversational Recom-
	guages with a Rather	Scaling in Analysis	mender Systems
	Free Word Order	of Estonian Emotion	Pontus Wärnestål, Lars
	Guntis Bārzdiņs, Nor-	Concepts	Degerstedt and Arne
	munds Grūzītis, Gunta	Toomas Kirt and Ene	Jönsson
	Nešpore and Baiba	Vainik	
	Saulīte		
14.30-	Comparing French PP-	Evaluating Stages of	Achieving Goals in
15.00	attachment to English,	Development in Second	Collaboration: Analysis
	German and Swedish	Language French: A	of Estonian Institutional
	Martin Volk and Frida	Machine-Learning Ap-	Calls
	Tidström	proach	Olga Gerassimenko,
		Jonas Granfeldt and	Mare Koit, Andriela
		Pierre Nugues	Rääbis and Krista
			Strandson

15.00-	Comprehension Assis-	Inducing Baseform
15.30	tant for Languages of	Models from a Swedish
	Baltic States	Vocabulary Pool
	Inguna Skadiņa, An-	Eva Forsbom
	drejs Vasiljevs, Daiga	
	Deksne, Raivis Skadiņš	
	and Linda Goldberga	

15.30–16.00 Coffee Break

- 16.00–17.30 Closing Session
- 16.00–17.00 Business Meeting of the Northern European Association for Language Technology (NEALT)
- 17.00–17.30 Closing