

# Frontiers in Artificial Intelligence and Applications

FAIA covers all aspects of theoretical and applied artificial intelligence research in the form of monographs, doctoral dissertations, textbooks, handbooks and proceedings volumes. The FAIA series contains several sub-series, including "Information Modelling and Knowledge Bases" and "Knowledge-Based Intelligent Engineering Systems". It also includes the biennial ECAI, the European Conference on Artificial Intelligence, proceedings volumes, and other ECCAI – the European Coordinating Committee on Artificial Intelligence – sponsored publications. An editorial panel of internationally well-known scholars is appointed to provide a high quality selection.

Series Editors:

J. Breuker, N. Guarino, J.N. Kok, J. Liu, R. López de Mántaras,  
R. Mizoguchi, M. Musen, S.K. Pal and N. Zhong

Volume 219

*Recently published in this series*

- Vol. 218. C. Soares and R. Ghani (Eds.), *Data Mining for Business Applications*
- Vol. 217. H. Fujita (Ed.), *New Trends in Software Methodologies, Tools and Techniques – Proceedings of the 9th SoMeT\_10*
- Vol. 216. P. Baroni, F. Cerutti, M. Giacomin and G.R. Simari (Eds.), *Computational Models of Argument – Proceedings of COMMA 2010*
- Vol. 215. H. Coelho, R. Studer and M. Wooldridge (Eds.), *ECAI 2010 – 19th European Conference on Artificial Intelligence*
- Vol. 214. I.-O. Stathopoulou and G.A. Tsihrintzis, *Visual Affect Recognition*
- Vol. 213. L. Obrst, T. Janssen and W. Ceusters (Eds.), *Ontologies and Semantic Technologies for Intelligence*
- Vol. 212. A. Respicio et al. (Eds.), *Bridging the Socio-Technical Gap in Decision Support Systems – Challenges for the Next Decade*
- Vol. 211. J.I. da Silva Filho, G. Lambert-Torres and J.M. Abe, *Uncertainty Treatment Using Paraconsistent Logic – Introducing Paraconsistent Artificial Neural Networks*
- Vol. 210. O. Kutz et al. (Eds.), *Modular Ontologies – Proceedings of the Fourth International Workshop (WoMO 2010)*
- Vol. 209. A. Galton and R. Mizoguchi (Eds.), *Formal Ontology in Information Systems – Proceedings of the Sixth International Conference (FOIS 2010)*
- Vol. 208. G.L. Pozzato, *Conditional and Preferential Logics: Proof Methods and Theorem Proving*
- Vol. 207. A. Bifet, *Adaptive Stream Mining: Pattern Learning and Mining from Evolving Data Streams*
- Vol. 206. T. Welzer Družovec et al. (Eds.), *Information Modelling and Knowledge Bases XXI*

ISSN 0922-6389 (print)  
ISSN 1879-8314 (online)

# Human Language Technologies The Baltic Perspective

Proceedings of the Fourth International Conference  
Baltic HLT 2010

Edited by

**Inguna Skadiņa**

*Institute of Mathematics and Computer Science, University of Latvia  
Tilde, Latvia*

and

**Andrejs Vasiljevs**

*Tilde, Latvia*

**IOS**  
Press

Amsterdam • Berlin • Tokyo • Washington, DC

**Spoken and Written Dialog**

- Internet Commentators as Dialogue Participants: Coherence Achieved Through Membership Categorization 83  
*Tiit Hemmoste, Olga Gerassimenko, Riina Kasterpalu, Mare Koit, Kirsi Laanesoo, Anni Oja, Andriela Rääbis and Krista Strandson*
- Uncertainty in Spoken Dialogue Management 91  
*Kristiina Jokinen*
- Human-Computer Interaction in Estonian: Collection and Analysis of Simulated Dialogues 99  
*Siiri Pärkson*
- A Framework for Asynchronous Dialogue Systems 107  
*Margus Treumuth*

**Machine Translation**

- SMT of Latvian, Lithuanian and Estonian Languages: A Comparative Study 117  
*Maxim Khalilov, Lauma Pretkalniņa, Natalja Kuvaldina and Veronika Pereseina*
- Improving SMT for Baltic Languages with Factored Models 125  
*Raivis Skadiņš, Kārlis Goba and Valters Šics*
- LetsMT! – Online Platform for Sharing Training Data and Building User Tailored Machine Translation 133  
*Andrejs Vasiljevs, Tatiana Gornostay and Raivis Skadins*

**Written Corpora and Linguistic Resources**

- The Estonian Reference Corpus: Its Composition and Morphology-Aware User Interface 143  
*Heiki-Jaan Kaalep, Kadri Muischnek, Kristel Uihoaed and Kaarel Veskis*
- Adaptive Automatic Mark-Up Tool for Legacy Dictionaries 147  
*Lauma Pretkalniņa and Ilze Millere*
- Corpus of Contemporary Lithuanian Language – The Standardised Way 154  
*Erika Rimkutė, Jolanta Kovalevskaitė, Vida Melninkaitė, Andrius Utkā and Daiva Vitkutė-Adžgauskienė*
- A Collection of Comparable Corpora for Under-Resourced Languages 161  
*Inguna Skadiņa, Ahmet Aker, Voula Giouli, Dan Tufis, Robert Gaizauskas, Madara Mieriņa and Nikos Mastropavlos*
- The Database of Estonian Word Families: A Language Technology Resource 169  
*Ülle Viks, Silvi Vare and Heete Sahkai*
- Digitization of Historical Texts at the National Library of Latvia 177  
*Arturs Zogla and Jurgis Skilters*

**Semantics**

- Verbalizing Ontologies in Controlled Baltic Languages 187  
*Normunds Grūzītis, Gunta Nešpore and Baiba Saulīte*
- Enriching Estonian WordNet with Derivations and Semantic Relations 195  
*Neeme Kahusk, Kadri Kerner and Kadri Vider*
- Main Trends in Semantic-Research of Estonian Language Technology 201  
*Haldur Ōim, Heili Orav, Kadri Kerner and Neeme Kahusk*
- Semantic Analysis of Sentences: The Estonian Experience 208  
*Haldur Ōim, Heili Orav, Neeme Kahusk and Piia Taremaa*

**Methods and Tools for Language Processing**

- An Ensemble of Classifiers Methodology for Stemming in Inflectional Languages: Using the Example of Latvian 217  
*Steffen Eger and Ineta Sējāne*
- Using Syllables as Indexing Terms in Full-Text Information Retrieval 225  
*Kimmo Kettunen, Paul McNamee and Feza Baskaya*
- Comparison of the SemTi-Kamols and Tesnière's Dependency Grammars 233  
*Gunta Nešpore, Baiba Saulīte, Guntis Bārzdīņš and Normunds Grūzītis*
- Cloud Computing for the Humanities: Two Approaches for Language Technology 241  
*Graham Wilcock*
- Subject Index 249
- Author Index 251

Future work will include spoken WOZ experiment series and analysing of written and spoken WOZ dialogues. The useful source of information could be comparison of WOZ dialogues with human-human phone calls. Due to comparison of different kinds of dialogue it is possible to identify universal and specific patterns of communication, and therefore detect the patterns which are suitable for the real DS.

### Acknowledgement

This paper is supported by the European Regional Development Fund through the Estonian Center of Excellence in Computer Science (EXCS) and the Estonian Science Foundation (grant 7503).

### References

- [1] A. Batliner, K. Fischer, R. Huber, J. Spilker, E. Nöth, How to find trouble in communication, *Speech Communication* 40 (2003), 117-143.
- [2] A. Bellucci, P. Bottoni, S. Levialdi, WOEB: Rapid Setting of Wizard of Oz Experiments and Reuse for Deployed Applications, Dipartimento di Informatica, Università Sapienza di Roma, Italy, 2009.
- [3] N. Dahlbäck, A. Jönsson, L. Ahrenberg, Wizard of Oz Studies – Why and How, *Knowledge-Based Systems Vol. 6, No. 4* (1993), 258-266.
- [4] K. Mäkelä, E.-P. Salonen, M. Turunen, J. Hakulinen, R. Raisamo, Conducting a Wizard of Oz Experiment on a Ubiquitous Computing System Doorman, *Proceedings of the International Workshop on Information Presentation and Natural Multimodal Dialogue*, Computer-Human Interaction Unit, Department of Computer and Information Sciences, University of Tampere, Finland, 2001.
- [5] M. Kullasaar, Eestikeelse dialoogikorpuse arendamine „võlur Ozi“ tehnikaga, Magistritöö, Tartu Ülikool, Arvutiteaduse instituut, Käsikiri, 2001.
- [6] J. Nielsen, *Usability Engineering*, San Fransisco, published by Morgan Kaufmann, 1993.
- [7] K. Jokinen, *Constructive Dialogue Modelling Speech Interaction and Rational Agents*, UK, John Wiley & Sons Ltd, 2009.
- [8] I. Hutchby, R. Wooffitt, *Conversation analysis: principles, practices, and applications*, Cambridge (UK): Polity Press, Malden (Mass.): Blackwell, 2006.
- [9] A.J. Liddicoat, *An Introduction to Conversation Analysis*, London, New York: Continuum, 2007.
- [10] T. Hennoste, A. Rääbis, *Dialoogiaktid eesti infodialogides: tüpoloogia ja analüüs*, Tartu Ülikooli Arvutiteaduse Instituut, Tartu: Tartu Ülikooli Kirjastus, 2004.
- [11] J. Sidnell, *Conversation analysis: an introduction*, Oxford (UK): John Wiley & Sons Ltd, 2010.
- [12] M.-L. Sorjonen, *Korjajäsemys. Keskusteluanalüüsi perusteet* (1997), edit. L. Tainio, Vastapaino, Tampere, Finland, 111 – 137.

# A Framework for Asynchronous Dialogue Systems

MARGUS TREUMUTH  
 University of Tartu, Estonia

**Abstract.** This paper presents a framework for asynchronous dialogue systems. It is used in developing text-based dialogue systems. The framework features web-based asynchronous turn management and AI-assisted live agent chat. Some other features are also briefly covered in this paper (a language independent solution for spell checking the user input, a language independent solution for the word-order problem, semantic resolution and a web-based interface for WOZ data collection). The exploitation of the asynchronous communication pattern has improved the communication style of the user which has resulted in decreased number or single word utterances. Higher word count per utterance is important when performing shallow language analysis without complete semantic understanding. The framework is currently tailored for Estonian language, yet most of its features and modules are language independent.

**Keywords.** asynchronous turn management, phrase pattern search, word order, text-based dialogue systems, semantic resolution

## 1. Introduction and Related Work

This paper presents a framework for asynchronous dialogue systems – the ADS framework. The ADS framework, developed by the author, is a collection of integrated modules that can be used in building text-based natural language dialogue systems. The framework features web-based asynchronous turn management and AI-assisted live agent chat. Some other features are also briefly covered in this paper (a language independent solution for spell checking the user input, a language independent solution for the word-order problem, semantic resolution and a web-based interface for WOZ data collection). The dialogue systems built on the ADS framework use an open prompt (non-restrictive) approach [1]. The framework is currently tailored for Estonian language, yet most of its features and modules are language independent.

There are several projects that are similar to the ADS framework. I studied some outstanding frameworks, such as Olympus/RavenClaw [2] [3], Semantra [4] [5] and CLSU Toolkit [6]. I also searched for demonstrations of web-based dialogue systems (DS) with asynchronous turn management, yet I was unable to find any evidence of such systems.

I was not able to use these frameworks because the Olympus/RavenClaw does not support web based asynchronous communication pattern. Semantra is an NLI engine and not really suitable for building DS. In addition, Semantra is a commercial tool and not freely available for experimenting. The modules of the CSLU Toolkit are not language independent. Also, the CSLU Toolkit is not easily portable to the web.