Language Technology in Estonia

What is language technology and computational linguistics?

Language technology is a branch of informatics that deals with the processing of natural language. **Computational linguistics** is a hybrid speciality that combines linguistics and computer science. Both language technology and computational linguistics deal with automatic processing of natural language; however, computational linguistics has a more theoretical and language technology a more applied focus.

THEORY	THEORY	APPLICATION
LINGUISTICS natural language		
	COMPUTATIONAL LINGUISTICS Computational processing of natural language	→ LANGUAGE TECHNOLOGY its methods and means
INFORMATICS Computational processing of information		

Source: Ülle Viks *Keeleteaduse ja -tehnoloogia vahekorrast* – presentation at the conference "Applied Linguistics in Estonia" 11–12 April 2002


Haldur Õim


Mare Koit


Ülle Viks

Which are the centres of computational linguistics and language technology in Estonia?

There are research groups of computational linguistics and language technology at the University of Tartu. The Institute of the Estonian Language focuses on speech synthesis and computational lexicography. The Institute of Cybernetics at the Tallinn University of Technology has a Laboratory of Phonetics and Speech Technology.

It can be studied at the University of Tartu within the framework of the curriculum of Estonian and Finno-Ugric linguistics at the Faculty of Philosophy and language technology a part of the curriculum of information technology at the Faculty of Mathematics and Computer Science. Courses in speech technology are taught by the Tallinn University of Technology.

What are Estonian computational linguists and language technologists engaged in?

A few examples

Everybody who writes texts on a computer is familiar with a spell-checker. A spell-checker is based on an automatic morphological analyser – a computer program that knows the lemma of each text word (token) and the inflected form used in the text. The spell-checker will underline in red all those text words that are unfamiliar to the morphological analyser. Automatic morphological analysis or its reverse process – morphological synthesis – is required also for the search function. For example, the search *suveaeg* 'summer time' will also cover sentences containing the expression *suveajale üleminek* 'switch to the summer time'.

In addition to the spell-checker, many people would be happy to use a grammar checker that can, for example, spot comma errors in the text. In order to create this kind of a writing assistant, one would need a syntactic analyser that can divide a sentence into clauses and detect the syntactic structure of clauses.

Lexicographers appreciate a web-based working environment that integrates various devices of language technology (language software and language resources) in order to make dictionary compilation and editing more simple and efficient.

Dialog systems, that is, user interfaces that enable the use of a natural language, will have a great future. For example, a computer could answer an information helpline instead of a human. In order to create a phone-based dialog system, at first speech recognition is required. The latter will convert human speech into a text; then one needs a program that will 'translate' the question by a human into a database search and then 'translate' the search again into a reply that is understood by a human. For this purpose, one has to know how people communicate by means of the phone – how they ask questions and answer to them. Finally, a speech synthesizer is needed to forward the reply to the person who had asked the question.

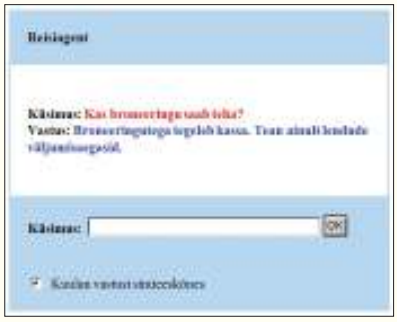
Web addresses:

Research group of computational linguistics http://www.cl.ut.ee; research group of language technology http://www.cs.ut.ee/~koit/KT/; the Institute of the Estonian Language: http://www.eki.ee;
Laboratory of Phonetics and Speech Technology at the Institute of Cybernetics of the Tallinn University of Technology: http://www.ioc.ee/;
Dialog systems: http://www.dialoogid.ee
Text-speech synthesis http://www.eki.ee/keeletehnoloogia/projektid/syntees/;
http://www.phon.ioc.ee/ -> projects -> text-speech synthesis for the blind
Electronic dictionaries can be consulted at http://www.keeleveeb.ee;
http://www.keelevara.ee
Dictionary compilation software EELex at the Institute of the Estonian Language
http://exsa.eki.ee/
Automatic summarizer http://math.ut.ee/~kaili/estsum/estsumframe.cgi


A sample of an online article from *Postimees* 'Most of the collected rubbish to be recycled' to be reviewed by the automated summarizer. The user can choose the amount of the summary: either 10, 20, 30, 40 or 50% of the article.


The output of the automated summarizer: 30% summary of the article 'Most of the collected rubbish to be recycled'

The dialogue system *Theatre Agent* offers information about performances in Estonian theatres.



Dialoogisüsteem Reisiagent annab infot Tallinna lennujaamast lähtuvate lennuliinide kohta. Suudab vastata ka


Ka Keeleveebi sõnastikuportaalil on ingliskeelne

The poster was prepared by the Mother Tongue Society, the Estonian Literary Museum, and the Estonian Ministry of Education and Research. Text by Külli Habicht. Design by Margus Nõmm, Multimedia Centre of the University of Tartu, 2008.