

# Language technology education and research in Lithuania

## Graduate education arrangements in general

Lithuania has one model of higher education

4 years of bachelor + 2 years of master + 4 years of doctoral studies

### Doctoral studies

Doctoral students have

- to pass at least three exams 20 credits (e.g. 5 at Vytautas Magnus University),
- to teach a certain amount of hours (e.g. 240 academic hours at VMU)
- to produce at least two papers in the peer reviewed magazines,
- to produce and to defend orally a doctoral dissertation in the form of a monograph with a printed abstract (in a foreign language if the dissertation is in Lithuanian and vice versa if the dissertation is in a foreign language)

Admission to the post-graduate studies is based on previous publications of the applicant and interview or entrance exams. The progress of doctoral students is reviewed by the departments in charge every term, inefficient students are expelled from the state supported doctoral studies.

Doctoral dissertations are defended in front of the board, unique for each dissertation. The board consists of 5 people, research supervisor included. Two opponents are appointed for each dissertation.

Dissertations can be defended by previous doctoral students after 4 years of doctoral studies but also by people from outside universities who present their dissertations and pass all the necessary exams.

The right to organize doctoral studies in specific fields is given to the universities and research institutions (separately or jointly) by the Government of Lithuania after the recommendation of the Ministry of Research and Education.

The field of Language Technology is not singled out as such thus HLT dissertations are defended either under Computer Science or Philology (Linguistics).

### Master's studies

There is one specialised master program on Computational Linguistics in Lithuania at Vytautas Magnus University. Its scope is 80 AP credits: 28 AP for 7 obligatory courses, 20 AP for 5 elective courses and 32 for 3 research projects and master's thesis. Obligatory courses included HLT, Corpus Linguistics, Machine Aided Human Translation, NLP, Computational Linguistics, Statistical Linguistics and Informational Text Technologies. The program on Computational Linguistics is affiliated to the Department of Lithuanian Language in the Faculty of Humanities.

## Sites relevant for language technologies

### **Academic institutions**

#### **Research institutes**

Institute of Mathematics and Informatics  
Recognition Processes Department  
[http://www.science.mii.lt/mii\\_an/frame\\_an/pradz\\_an.htm](http://www.science.mii.lt/mii_an/frame_an/pradz_an.htm)  
Institute of the Lithuanian Language  
<http://www.lki.lt/English/index.html>

#### **Universities**

Vilnius University <http://www.vu.lt/english/>  
Faculty of Mathematics and Informatics  
Department of Computer science II  
<http://www.mif.vu.lt/katedros/cs2/katedhp.htm>  
Faculty of Philology Laboratory of Experimental Phonetics  
[www.eunet.lt/girdenis](http://www.eunet.lt/girdenis)

Kaunas University of Technology <http://www.ktu.lt/en/>  
Speech Signals Research Laboratory  
<http://www.ktu.lt/en/index1.html>

Vytautas Magnus University <http://www.vdu.lt/>  
Centre of Computational Linguistics  
<http://donelaitis.vdu.lt/indexa.html>

### **Governmental institutions**

Seimas (Parlament) of the Republic of Lithuania  
[http://www3.lrs.lt/pls/inter/w4\\_home.int\\_ang](http://www3.lrs.lt/pls/inter/w4_home.int_ang)  
Commision of the development of Information Society  
[http://www3.lrs.lt/pls/inter/w3\\_viewer.ViewTheme?p\\_int\\_tv\\_id=838&p\\_kalb\\_id=2&p\\_org=0](http://www3.lrs.lt/pls/inter/w3_viewer.ViewTheme?p_int_tv_id=838&p_kalb_id=2&p_org=0)  
The State Commission of the Lithuanian Language  
[http://www3.lrs.lt/pls/inter/w3\\_viewer.ViewTheme?p\\_int\\_tv\\_id=530&p\\_kalb\\_id=2&p\\_org=0](http://www3.lrs.lt/pls/inter/w3_viewer.ViewTheme?p_int_tv_id=530&p_kalb_id=2&p_org=0)  
Department of Information technologies  
<http://www.lrs.lt/>  
Information Society Development Committee  
Under the Government of Republic of Lithuania  
[http://www.ivpk.lt/main\\_en.php](http://www.ivpk.lt/main_en.php)

### **Private companies**

InfoBalt <http://www.infobalt.lt/english/>  
Fotonija <http://www.fotonija.lt/>  
VTeX <http://www.led.lt/index.html>  
Tilde <http://www.tilde.lv/English/>

Association *InfoBalt* <http://www.infobalt.lt/english/>

Lithuanian Computer Society <http://www.liks.lt/>

## Activities and products

### **Institute of Mathematics and Informatics Recognition Processes Department**

Signal processing, random processes recognition and computer-aided multimedia:  
Transfer of processes (speech, music signals) and graphics (texts, notes, pictures) to computer-aided multimedia. Digitizing, compressing and remastering of noise-distorted processes and graphics. Applications include speaker identification, speech, music signals and multimedia presentation of speech, music signals.

A tool for isolated Lithuanian word recognition system for 10 digits and 2 control words.

### **Institute of the Lithuanian Language**

Digitalization of explanatory dictionaries of the Lithuanian language  
Compilation and tagging of the corpus of the old writings (16-18th century)

### **Vilnius University Faculty of Mathematics and Informatics Department of Computer Science II**

image and signal processing, speech synthesis from text.

### **Faculty of Philology Laboratory of Experimental Phonetics**

Speech sound analysis (a tool called *Kalbame 4.4*), speech synthesis (a tool called *Aistis*)

### **Kaunas University of Technology Speech Signals Research Laboratory**

Phoneme discrimination using regularized discriminant analysis; speech recognition based on advanced local discriminants in the continuous density hidden Markov model; noisy speech detection, endpointing and enhancement; Lithuanian speech databases, speaker recognition by analyzing features of phonemes, also investigated recognition of phonemes using discriminative analysis and phonetic segmentation of speech signals.

Speech recognition applications: services for citizens, tools for disabled people, voice driven telephone services, forensic speech examination.

### **Vytautas Magnus University Centre of Computational Linguistics**

Compilation of the corpus of 120 million running words of the present day Lithuanian (on internet);  
lemmatized translation of the Holy Bible (on internet);  
morphological tagging of a small corpus (ca 1 million words);  
parallel corpora of a) Czech-Lithuania (fiction);  
b) English-Lithuanian, German-Lithuanian (EU documents, ca 1 million words each);  
compilation of the dictionary of collocations;  
a tool of morphological analysis and synthesis *Lemuoklis*;  
development of a parser;  
special tools for corpus administration and query systems;  
the tool *Collector* designed for the compilation of a corpus from Internet newspapers.

Universal annotated speech corpus of standard Lithuanian (4 speakers, 750 isolated words, 1 hour duration);  
the tool for automatic segmentation of music "Solo Explorer", based on the new method of logical recognition;  
HMM speech recognition system based on MFCCs; speech synthesis.

Research areas: morphological ambiguity resolution, statistical vs. linguistic collocations, semi-automatic term identification and extraction, automatic summarisation and genre detection, non-linear speech processing, statistical language modelling (probabilities of words and word forms).

## **Seimas (Parlament) of the Republic of Lithuania**

### **Commision of the Development of Information Society**

Initiation of legal issues concerning information society

### **The State Commission of the Lithuanian Language**

Research and development program *Lithuanian Language and the Information Society 2000-2006*

### **Department of Information technologies**

Eurovoc and EU documents translation, search system for Lithuanian legislation

## **Information Society Development Committee Under the Government of Republic of Lithuania**

Participation in the process of shaping state policy for the development of information technologies and telecommunications and coordinate its implementation, coordination of planning, creation and development of the ITT infrastructure complying with the European Union standards

### **Association *InfoBalt***

Representation and defence of the interests of the IST industry in Lithuania and abroad

### **Private Company *Fotonija***

Electronic dictionaries of international words *Interleksis*, *TŽŽ*; English-Lithuanian dictionary *Alkonas*, spellchecker *Juodos avys*

### **Lithuanian Dutch Join Venture *VTeX***

English-Lithuanian dictionaries *WinLed 4* and *WinLed 4*

### **Private Company *Tilde* (in Ryga, Latvia with bureau in Vilnius)**

Localisation and translation services, multilanguage support software, spellchecker.