



# Katalin Pajkossy

## ○ Personal data

---

**Name:** Katalin Anna Pajkossy  
**Citizenship:** Hungarian  
**Date of Birth:** 1988.11.09.  
**Mobile:** 06 30/ 248 5294  
**E-mail:** [pajkossyk@gmail.com](mailto:pajkossyk@gmail.com)  
**Github:** <https://github.com/pajkossy/>

## ○ Education

---

**Phd student, Doctoral School of Mathematics and Computer Science, 2014-**  
Budapest University of Technology and Economics, Mathematical Institute  
Topic: Deep learning methods in Natural Language Processing  
Advisor: András Kornai

**Applied Mathematician, M.Sc., 2010-2013**  
Eötvös Loránd Science University, Budapest  
Thesis: Studying feature selection methods in natural language processing  
Advisor: A. Kornai

**Mathematician, B.Sc., 2007-2010**  
Eötvös Loránd Science University, Budapest  
Thesis: Minimization of graphical submodular functions  
Advisor: K. Bérczi

## ○ Work experience

---

### **July 2015- Researcher**

- Projects
  - Computing digital vitality index of languages (for SIL International)
  - Speaker diarization (for hungarian speech archive project)
- Shared task:
  - Semeval-2016 Task 1 - Semantic textual similarity

---

### **2014-July 2015 Hungarian Academy of Sciences, Institute for Linguistics Assistant research fellow**

- Projects:
  - Web search, query normalization for web search (for CLIQZ)
  - Researching digital vitality of languages
- Developing open-source framework for neural network-based sequential tagging (hunvec)

### **2012-2013 Hungarian Academy of Sciences, Institute for Computer Science and Control Software developer**

- Dictionary building
- Corpus processing - data crawling, cleaning, tokenization, parallelization,
- Analyzing NLP algorithms (machine learning, feature selection, computational semantics)

## ○ Skills

---

### **Algorithms**



# Katalin Pajkossy

Machine learning, feature selection, intelligent data structures

## **Natural language processing**

dictionary building, spell-checking, sequential tagging

## **Programming**

Python - numpy, pandas, scikit-learn, ipython  
Linux, bash

## **Spoken languages**

Hungarian Native  
English C1  
German B2

## ○ Publications

---

### **Proceedings of the 10th Edition of the Language Resources and and Evaluation Conference (LREC 2016, to appear)**

The hunvec framework for NN-CRF-based sequential tagging (K.Pajkossy, A.Zséder)

### **Proceedings of \*Sem 2015**

Competence in lexical semantics (A.Kornai, J.Ács, M.Makrai, D.M.Nemeskey, K.Pajkossy, G.Recski)

### **Proceedings of the 6th Workshop on Building and Using Comparable Corpora (BUCC 2013)**

Building basic vocabulary across 40 languages (J. Ács, K.Pajkossy, A.Kornai)