Márton Makrai

computational linguist

+36 20 773 8807
makrai.marton@ttk.hu
http://hlt.bme.hu/en/makrai
makrai

Innumerable roads lead to "knowledge"
and we try to explore many of them. (Nicolas Findler)

Education, Courses, and Summer Schools

2021–2024 **PhD in Theoretical Linguistics**, *ELTE BTK Doctoral School of Linguistics*, Supervisor: András Kornai.
Title of the dissertation: *Symbolic and Distributed Word Representations*

2017–2021 **IRDTA DeepLearn**, *Winter and Summer Schools*, `https://irdta.eu/deeplearn/`, Summers of 2017–2019 and 2021.

2015 **Deep Learning for Machine Translation**, *Winter School*, `http://dl4mt.computing.dcu.ie`.

2012 **Natural Language Processing**, *Online Course*, Stanford University, D. Jurafsky, C. Manning.

2012 **Machine Learning**, *Online Course*, Stanford University, A. Ng.

2004–2010 **MSc in Mathematics**, *Budapest University of Technology and Economics, Faculty of Natural Sciences*, Major: Algebra, with a focus on formal languages and graph theory.
The thesis, titled *Subregular Categorial Grammars* and supervised by András Kornai, defined categorial grammars that are learnable in the formal linguistic sense *(identification in the limit)*.

2000-2004 **Secondary School**, *Budai Ciszterci Szent Imre Gimnázium*.

Employment

2015– **HUN-REN TTK Institute of Cognitive Neuroscience and Psychology**, *Environmental Adaptation and Space Research Group*, Research Fellow.
CoALa: Thematic and emotional analysis of crew members' work diaries in isolated groups operating under extreme conditions;
AstroSpeech: Monitoring astronauts' mental states and emotions with full privacy

2024– **HUN-REN Research Centre for Linguistics**, *Institute of Lexicology, Lexical Knowledge Representation Research Group*, External collaborator, The Hungarian Constructicon.
Constructions include all linguistic signs i.e. form-meaning pairs, which we acquire (memorize), whether it be a suffix, a syntactic structure (one containing a specific word or a general pattern), or even a conventional discourse/rhetorical turn of phrase.

July 2022 – March 2024 **ELKH/HUN-REN TTK Institute of Cognitive Neuroscience and Psychology**, *Speech and Language Perception Research Group*, External collaborator, CommGame.
Predicting surprise at specific points in Hungarian conversations using deep language models.

Jan 2022 – Aug 2023 **University of Szeged, Faculty of Science and Informatics**, Research fellow, Súlylesők.
Investigating the stability of the pre-training and fine-tuning of neural language models; predicting performance based on spectral properties of weight matrices.

2021 **BME Department of Telecommunications and Media Informatics**, External collaborator, Summarization for Hungarian.
https://huggingface.co/BME-TMIT/foszt2oszt/

2014–2020 **MTA Research Institute for Linguistics**, *Department of Language Technology and Applied Linguistics*, Research fellow.

- efnilex 2014–2015: Computational lexicography and machine translation for European national languages, dictionary generation using vector-based language models;
- cliqz 2014–2015: Web search;
- Young Researcher 2015–2018: Thematic roles, word embeddings, lexical relations, word sense clustering;
- siren 2015–2017: Editing reports on the digital vitality of languages:
- emtsv 2019: A modular Hungarian NLP pipeline;
- HILBert 2020: Preparation for pre-training Hungarian deep language models.

2010–2013 **MTA SZTAKI Institute for Computer Science and Control**, *Language Technology Research Group*, Software Developer.
Investigating the formal linguistic learnability of human language; General Natural Language Understanding, `4lang` semantic representation, and conceptual dictionary.

Awards

2025 **Best Reviewer Award (shared)**.
at the Hungarian Conference on Computational Linguistics (MSZNY)

2023 **Outstanding Reviewer Award**.
at the *Association for Computational Linguistics* conference

2014 **Best Young Researcher Award at MSZNY, shared**.
Deep Cases in the `4lang` Conceptual Dictionary

2013 **Best Young Researcher Award at the MSZNY, shared**.
The `4lang` Conceptual Dictionary

Teaching

Spring 2022 **Computational Linguistics**, *MTA-ELTE Theoretical Linguistics Group*, Introduction to NLP.
`https://hlt.bme.hu/hu/course/nlp22`

Fall 2018 **Computational Lexical Semantics 2**, *MTA-ELTE Theoretical Linguistics Group and PPKE ITK*.
`https://hlt.bme.hu/hu/course/lexsem18`

Spring 2018 **Vector Space Models**, *MTA-ELTE Theoretical Linguistics Group and PPKE ITK*.
`https://hlt.bme.hu/hu/course/vsm18`

Fall 2017 **Computational Lexical Semantics – Symbolic Representations**, *MTA-ELTE Theoretical Linguistics Group and PPKE ITK*.
`https://hlt.bme.hu/hu/course/lexsem17`

Fall 2015 **Semantic Representation**, *MTA-ELTE Theoretical Linguistics Group*.
`http://hlt.bme.hu/hu/course/reprs`

Spring 2015 **Digital Linguistics (in English)**, *MTA-ELTE Theoretical Linguistics Group*.
`http://hlt.bme.hu/en/course/digling`

Fall 2013 and Fall 2014 **Efficient Methods in Linguistic Description**, *MTA-ELTE Theoretical Linguistics Group*, Finite-State Automata and Regular Expressions for Linguists.
`https://budling.nytud.hu/~makrai/oktatas/modszerek/`

Languages

English Fluent in conversation and technical language

German Intermediate (Type C certificate)

Hungarian Native

Computer Skills

Python, scikit-learn, huggingface transformers, git, jupyter, LaTeX

Interests

- psycholinguistics
- amateur chamber music

Selected publications

Google Scholar

2015 **Competence in lexical semantics**. *A Kornai, J Acs,* M Makrai, *DM Nemeskey, K Pajkossy, G Recski*. *SEM

2019 **One format to rule them all–The emtsv pipeline for Hungarian**. *B Indig, B Sass, E Simon, I Mittelholcz, N Vadász,* M Makrai. Linguistic Annotation Workshop

2018 **300-sparsans at SemEval-2018 Task 9: Hypernymy as interaction of sparse attributes**. *G Berend,* M Makrai, *P Földiák*. Proceedings of the 12th International Workshop on Semantic Evaluation, 928-934

2021 **HILBERT, magyar nyelvű BERT-large modell tanítása felhő környezetben**. *Á Feldmann, R Hajdu, B Indig, B Sass,* M Makrai, *I Mittelholcz, D Halász, ...* Szegedi Tudományegyetem TTIK, Informatikai Intézet

2015 **Comparison of distributed language models on medium-resourced languages**. *M Makrai*. XI. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2015), 22–33

2013 **The `4lang` concept dictionary [A `4lang` fogalmi szótár]**. *A Kornai,* M Makrai. Szegedi Tudományegyetem Informatikai Tanszékcsoport

2016 **Evaluating multi-sense embeddings for semantic resolution monolingually and in word translation**. *G Borbély,* M Makrai, *DM Nemeskey, A Kornai*. RepEval

2016 **Filtering Wiktionary Triangles by Linear Mapping between Distributed Word Models**. *M Makrai*. LREC

2013 **Spreading activation in language understanding**. *DM Nemeskey, GA Recski,* M Makrai, *A Zséder, A Kornai*. National Academy of Sciences of Armenia