

Komplex vizsga olvasmánylista – Virág Nándor

Számítógépes nyelvészet

- Cheng, Xuyou – Michael Schlichtkrull – Guy Emerson 2023. Are Embedded Potatoes Still Vegetables? On the Limitations of WordNet Embeddings for Lexical Semantics. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. Singapore: Association for Computational Linguistics. 8763–8775.
- Dalal, Mita K – Mukesh A Zaveri 2011. Automatic text classification: a technical review. *International journal of computer applications* **28/2** Citeseer:37–40.
- Dhungana, Udaya Raj – Subarna Shakya 2015. Hypernymy in WordNet, Its Role in WSD, and Its Limitations. *2015 7th International Conference on Computational Intelligence, Communication Systems and Networks*. Riga, Latvia: IEEE. 15–19.
- Drávucz Fanni – Szabó Martina Katalin – Vincze Veronika 2017. Szentiment- és emóciósztárak eredményességének mérése emóció- és szentimentkorpuszokon. In Vincze Veronika (szerk.) *XIII. Magyar számítógépes nyelvészeti konferencia*. Szeged: Szegedi tudományegyetem, TTIK, Informatikai intézet. 228–239.
- Dunn, Jonathan 2022. *Natural language processing for corpus linguistics*. Cambridge, United Kingdom: Cambridge University Press.
- Frank, Eibe – Remco R. Bouckaert 2006. Naive Bayes for Text Classification with Unbalanced Classes. In Johannes Fürnkranz – Tobias Scheffer – Myra Spiliopoulou (szerk.) *Knowledge Discovery in Databases: PKDD 2006*. Vol. 4213 (Lecture Notes in Computer Science) Berlin, Heidelberg: Springer Berlin Heidelberg. 503–510. doi:[10.1007/11871637_49](https://doi.org/10.1007/11871637_49).
- Indig Balázs – Sass Bálint – Simon Eszter – Mittelholcz Iván – Kundráth Péter – Vadász Noémi 2019. emtsv – Egy formátum mind felett. In Berend Gábor – Gosztolya Gábor – Vincze Veronika (szerk.) *XV. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem TTIK, Informatikai Intézet. 235–247.
- Jakubiček, Miloš – Adam Kilgarriř – Diana McCarthy – Pavel Rychlý 2010. Fast Syntactic Searching in Very Large Corpora for Many Languages. In Ryo Ootoguro – Kiyoshi Ishikawa – Hiroshi Umemoto – Kei Yoshimoto – Yasunari Harada (szerk.) *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation*. Tohoku University, Sendai, Japan: Institute of Digital Enhancement of Cognitive Processing, Waseda University. 741–747.
- Jurafsky, Daniel – James Martin 2008. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Vol. 2.
- Jurafsky, Daniel – James H. Martin 2025. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*. 3. Online kézirat.
- Kim, Sang-Bum – Kyoung-Soo Han – Hae-Chang Rim – Sung-Hyon Myaeng 2006. Some Effective Techniques for Naive Bayes Text Classification. *Knowledge and Data Engineering, IEEE Transactions on* **18**:1457–1466. doi:[10.1109/TKDE.2006.180](https://doi.org/10.1109/TKDE.2006.180).
- McEnery, Tony – Andrew Hardie 2013. The History of Corpus Linguistics. In Keith Allan (szerk.) *The Oxford Handbook of the History of Linguistics*. 1. kiad. Oxford University Press. 727–745. doi:[10.1093/oxfordhb/9780199585847.013.0034](https://doi.org/10.1093/oxfordhb/9780199585847.013.0034).

- Miháltz, Márton – Bálint Sass – Balázs Indig 2013. What Do We Drink? Automatically Extending Hungarian WordNet With Selectional Preference Relations. Trento. 105–109.
- Miller, George A. 1995. WordNet: a lexical database for English. *Communications of the ACM* **38/11**:39–41. doi:[10.1145/219717.219748](https://doi.org/10.1145/219717.219748).
- Nemeskey, Dávid Márk 2021. Introducing huBERT. In Gábor Berend – Gábor Gosztolya – Veronika Vincze (szerk.) *XVII. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem, Informatikai Intézet. 3–14.
- Nielsen, Frank 2016. Hierarchical Clustering. In Frank Nielsen *Introduction to HPC with MPI for Data Science*. (Undergraduate Topics in Computer Science) Cham: Springer International Publishing. 195–211. doi:[10.1007/978-3-319-21903-5_8](https://doi.org/10.1007/978-3-319-21903-5_8).
- Resnik, Philip 1995. Using Information Content to Evaluate Semantic Similarity in a Taxonomy. arXiv. doi:[10.48550/ARXIV.CMP-LG/9511007](https://doi.org/10.48550/ARXIV.CMP-LG/9511007).
- Romano, Maurizio – Giulia Contu – Francesco Mola – Claudio Conversano 2024. Threshold-based Naïve Bayes classifier. *Advances in Data Analysis and Classification* **18/2**:325–361. doi:[10.1007/s11634-023-00536-8](https://doi.org/10.1007/s11634-023-00536-8).
- Szécsényi Tibor 2019. Argumentumszerkezet-variánsok korpusz alapú meghatározása. In Berend Gábor – Gosztolya Gábor – Vincze Veronika (szerk.) *XV. Magyar számítógépes nyelvészeti konferencia*. Szeged: Szegedi tudományegyetem, TTIK, Informatikai intézet. 315–329.
- Szécsényi Tibor – Virág Nándor 2022. Az ige helyhatározói bővítményeinek megkülönböztetése és az argumentumszerkezeti variánsok korpusz alapú szétválasztása. In Berend Gábor – Gosztolya Gábor – Vincze Veronika (szerk.) *XVIII. Magyar számítógépes nyelvészeti konferencia*. Szeged: JATEPress. 535–547.
- Taboada, Maite 2016. Sentiment Analysis: An Overview from Linguistics. *Annual Review of Linguistics* **2/1**:325–347. doi:[10.1146/annurev-linguistics-011415-040518](https://doi.org/10.1146/annurev-linguistics-011415-040518).
- Tilk, Ottokar – Vera Demberg – Asad Sayeed – Dietrich Klakow – Stefan Thater 2016. Event participant modelling with neural networks. In Jian Su – Kevin Duh – Xavier Carreras (szerk.) *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, Texas: Association for Computational Linguistics. 171–182.
- Van de Cruys, Tim 2014. A Neural Network Approach to Selectional Preference Acquisition. In Alessandro Moschitti – Bo Pang – Walter Daelemans (szerk.) *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics. 26–35.
- Várad, Tamás – Eszter Simon – Bálint Sass – Iván Mittelholcz – Attila Novák – Balázs Indig – Richárd Farkas – Veronika Vincze 2018. E-magyar – A Digital Language Processing System. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association (ELRA).
- Veziroğlu, Merve – Erkan Veziroğlu – İhsan Ömür Bucak 2024. Performance Comparison between Naive Bayes and Machine Learning Algorithms for News Classification. In İhsan Ömür Bucak (szerk.) *Bayesian Inference - Recent Trends*. IntechOpen. doi:[10.5772/intechopen.1002778](https://doi.org/10.5772/intechopen.1002778).
- Vincze, Veronika – György Szarvas – Janos Csirik 2008. Why are wordnets important? *European Computing Conference. (ECC 08), WSEAS*. Malta. 316–322.

- Warstadt, Alex – Amanpreet Singh – Samuel R. Bowman 2019. Neural Network Acceptability Judgments. *Transactions of the Association for Computational Linguistics* 7:625–641. doi:[10.1162/tacl_a_00290](https://doi.org/10.1162/tacl_a_00290).
- Ye, Patrick 2004. Selectional Preference Based Verb Sense Disambiguation Using WordNet. *Proceedings of the Australasian Language Technology Workshop 2004*. 155–162.
- Zsibrita, János – Richárd Farkas – Veronika Vincze 2013. magyarlanc: A toolkit for morphological and dependency parsing of Hungarian. *International Conference on Recent Advances in Natural Language Processing*. Shoumen, Bulgária: INCOMA Ltd. 763–771.

Korpusznyelvészet

- Aijmer, Karin 2018. Corpus pragmatics: From form to function. In Andreas H. Jucker – Klaus P. Schneider – Wolfram Bublitz (szerk.) *Methods in Pragmatics*. De Gruyter. 555–586. doi:[10.1515/9783110424928-022](https://doi.org/10.1515/9783110424928-022).
- Archer, Dawn – Johnatan Culpeper – Matthew Davies 2009. Pragmatic annotation. In Anke Lüdeling – Merja Kytö (szerk.) *Corpus Linguistics: An International Handbook*. Berlin; New York: De Gruyter. 613–642.
- Biber, Douglas 2015. Corpus-Based and Corpus-Driven Analyses of Language Variation and Use. In Bernd Heine – Heiko Narrog (szerk.) *The Oxford Handbook of Linguistic Analysis*. Oxford University Press. 193–224. doi:[10.1093/oxfordhb/9780199677078.001.0001](https://doi.org/10.1093/oxfordhb/9780199677078.001.0001).
- Biber, Douglas – Randi Reppen (szerk.) 2015. *The Cambridge handbook of English corpus linguistics*. (Cambridge handbooks in language and linguistics) Cambridge: Cambridge University Press.
- Bibok, Károly – Martina Katalin Szabó 2025. A corpus analysis of negative emotive words in Hungarian (with special focus on their discursive functions). *ARGUMENTUM* 21:1–13. doi:[10.34103/ARGUMENTUM/2025/1](https://doi.org/10.34103/ARGUMENTUM/2025/1).
- Collins, Luke 2019. *Corpus linguistics for online communication: a guide for research*. (Routledge corpus linguistics guides) London ; New York: Routledge/Taylor & Francis Group.
- Dunn, Jonathan 2022. *Natural language processing for corpus linguistics*. Cambridge, United Kingdom: Cambridge University Press.
- Egbert, Jesse – Tove Larsson – Douglas Biber 2020. *Doing Linguistics with a Corpus: Methodological Considerations for the Everyday User*. (Elements in Corpus Linguistics) Cambridge University Press.
- Gyulai, Livia 2021. Az igekötők legjellemzőbb argumentumszerkezet-változtató hatásainak korpuszalapú vizsgálata. *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből 2021* MTA Nyelvtudományi Intézet:176–198. doi:[10.18135/ALKNYELVDOK.2021.15.8](https://doi.org/10.18135/ALKNYELVDOK.2021.15.8).
- McEnery, Tony – Andrew Hardie 2012. *Corpus linguistics: method, theory and practice*. (Cambridge Textbooks in Linguistics) Cambridge New York: Cambridge University Press.
- Milà-Garcia, Alba 2018. Pragmatic Annotation for a Multi-Layered Analysis of Speech Acts: A Methodological Proposal. *Corpus Pragmatics* 2/3:265–287. doi:[10.1007/s41701-018-0037-z](https://doi.org/10.1007/s41701-018-0037-z).
- Oravecz, Csaba – Tamás Váradi – Bálint Sass 2014. The Hungarian Gigaword Corpus. In Nicoletta Calzolari – Khalid Choukri – Thierry Declerck – Hrafn Loftsson – Bente Maegaard – Joseph Mariani – Asuncion Moreno – Jan Odijk – Stelios Piperidis (szerk.) *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik, Iceland: European Language Resources Association. 1719–1723.
- Sass Bálint 2017. Keresés korpuszban: a kibővített Magyar történeti szövegtár új keresőfelülete. In Forgács Tamás – Németh Miklós – Sinkovics Balázs (szerk.) *Nyelvtörténeti kutatások újabb eredményei*. Szeged: Szegedi tudományegyetem, Magyar nyelvészeti tanszék. 267–277.
- Sass Bálint 2018. Mazsola - mindenkinek. In Vincze Veronika (szerk.) *XIV. Magyar számítógépes nyelvészeti konferencia*. Szeged: Szegedi tudományegyetem, TTIK, Informatikai intézet. 16–24.

- Szabó, Martina Katalin – Veronika Vincze – Katalin Ilona Simkó – Viktor Varga – Viktor Hangya 2016. A Hungarian Sentiment Corpus Manually Annotated at Aspect Level. In Nicoletta Calzolari (szerk.) *LREC 2016, Tenth International Conference on Language Resources and Evaluation*. Paris: European Language Resources Association. 2873–2878.
- Szécsényi Tibor – Nagy C. Katalin – Németh T. Enikő 2024. Felszólításannotálás a MedCollect egészségügyi álhírkorpuszban. *XX. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: SZTE TTIK Informazikai Intézet. 159–170.
- Vincze, Veronika – Dóra Szauter – Attila Almási – György Móra – Zoltán Alexin – János Csirik 2010. Hungarian Dependency Treebank. In Nicoletta Calzolari – Khalid Choukri – Bente Maegaard – Joseph Mariani – Jan Odijk – Stelios Piperidis – Mike Rosner – Daniel Tapias (szerk.) *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Valletta, Malta: European Language Resources Association (ELRA).