# Publications by Beáta Megyesi

## 2024

Fornés, A., Chen, J., Torras, P., Badal, C., Megyesi, B., Waldispühl, M., Kopal, N., Lasry, G. (2024). ICDAR 2024 Competition on Handwriting Recognition of Historical Ciphers. In: Barney Smith, E.H., Liwicki, M., Peng, L. (eds) *Document Analysis and Recognition - ICDAR 2024*. Lecture Notes in Computer Science, vol 14809. Springer, Cham. https://doi.org/10.1007/978-3-031-70552-6_20

Goio, G., Torras, P., Fornés, A., and Megyesi, B. (2024) Exploring the Alignment of Transcriptions to Images of Encrypted Manuscripts. In *Proceedings of the 7th International Conference on Historical Cryptology (HistoCrypt 2024.)* Doi: https://doi.org/10.58009/aere-perennius0096

Héder, M., Fornés, A., Kopal, N., Szigeti, F., and Megyesi, B. (2024) Supporting Historical Cryptology: The Decrypt Pipeline. In *Proceedings of the 7th International Conference on Historical Cryptology (HistoCrypt 2024.)* Doi: https://doi.org/10.58009/aere-perennius0100

Láng, B. and Megyesi, B. (2024) An STS analysis of a digital humanities collaboration: trading zones, boundary objects, and interactional expertise in the DECRYPT project. *Humanities and Social Sciences Communication* **11**, 618 (2024). https://doi.org/10.1057/s41599-024-03135-w

Megyesi, B., Fornés, A., Kopal, N., Láng, B., Waldispühl, M., Mikhalev, V. and Esslinger, B. (2024) Historical Cryptology. In Esslinger, B. (ed), Chapter 3 in *Learning and Experiencing Cryptography with CrypTool and SageMath* , Artech, 2024.

Megyesi, B., Tudor, C., Láng, B., Lehofer, A., Kopal, N., de Leeuw, K., and Waldispühl, M. (2024) Keys with Nomenclatures in the Early Modern Europe. Journal of *Cryptologia* 48:2 pp. 97-139. Taylor and Francis. Doi: 10.1080/01611194.2022.2113185

Megyesi, B., Láng, B., Kopal, N., Mikhalev, V., Tudor, C., and Waldispühl, M. (2024) A Typology for Cipher Key Instructions in Early Modern Times. In *Proceedings of the 7th International Conference on Historical Cryptology (HistoCrypt 2024.)* Doi: https://doi.org/10.58009/aere-perennius0107

Waldispühl, M. and Megyesi, B. (eds.) (2024) Proceedings of the 7th International Conference on Historical Cryptology (HistoCrypt 2024) NEALT Proceedings Series 53. D-Space at Tartu University Library Doi: https://doi.org/10.58009/aere-perennius0084

## 2023

Ilinykh, N., Morger, F., Dannélls, D., Dobnik, S., and Megyesi, B., and Nivre, J. (editors) Proceedings of the Second Workshop on Resources and Representations for Under-Resourced Languages and Domains. NoDaLiDa 2023. Association for Computational Linguistics.

Megyesi, B., Sikora, J., Fornmark, F., Waldispühl, M., Kopal, N., and Mikhalev, V. (2023) Historical Language Models in Cryptanalysis: Case Studies on English and German. In the Proceedings of the 6th International Conference on Historical Cryptology. HistoCrypt 2023. pp. 120-129.

Mikhalev, V., Kopal, N., Esslinger, B., Waldispühl, M. Láng,B., and Megyesi B. (2023) What is the Code for the Code? – Historical Cryptology Terminology. In the Proceedings of the 6th International Conference on Historical Cryptology. HistoCrypt 2023. pp. 130-138.

Wu, Y., Nouri, J., Megyesi, B., Henriksson, A., Duneld, M., Li, X. (2023). Towards Data-Effective Educational Question Generation with Prompt-Based Learning. In: Arai, K. (eds) Intelligent Computing. SAI 2023. Lecture Notes in Networks and Systems, vol 711. Springer, Cham. https://doi.org/10.1007/978-3-031-37717-4_11

## 2022

Dahlke, C. and Megyesi, B. (eds) (2022) Proceedings of the 5th International Conference on Historical Cryptology. NEALT Proceedings Series 49. HistoCrypt 2022. Published by Linköping Electronic Press.

Gambardella, M.E., Megyesi, B., and Pettersson, E. (2022) Identifying Cleartext in Historical Ciphers. In *Proceedings of the Workshop on Language Technologies for Historical and Ancient Languages*. LT4HALA 2022. pp. 1-10.

Héder, M. and Megyesi, B. (2022) The DECODE Database of Historical Ciphers and Keys: Version 2. In *Proceedings of the 5th International Conference on Historical Cryptology*. HistoCrypt 2022.

Magnifico, G., Megyesi, B., Souibgui, M.A., Chen, J., and Fornés, A. (2022) Lost in Transcription of Graphic Signs in Ciphers. In *Proceedings of the 5th International Conference on Historical Cryptology*. HistoCrypt 2022.

Megyesi, B., Tudor, C., Kopal, N., Láng, B., Lehofer, A., and Waldispühl, M. (2022) What Was Encoded in Historical Cipher Keys in the Early Modern Era?. In *Proceedings of the 5th International Conference on Historical Cryptology*. HistoCrypt 2022.

Souibgui, M.A., Fornés, A., Kessentini, Y., and Megyesi, B. (2022) Few Shots Are All You Need: A Progressive Learning Approach for Low Resource Handwritten Text Recognition. Journal of *Pattern Recognition Letters*. Elsevier.

## 2021

Chen, j., Souibgui, M.A., Fornés, A., and Megyesi, B. (2021) Unsupervised Alphabet Matching in Historical Encrypted Manuscript Images. In *Proceedings of the 4th International Conference on Historical Cryptology*. HistoCrypt 2021.

Megyesi, B. and Láng, B. (2021) Revealing Secrets from the Past: Studying Historical Ciphers. Abstract to Digital History in Sweden, Dec. 9-10, 2021, Umeå University, Sweden.

Megyesi, B., Rudebeck, L., and Volodina, E. (2021) SweLL Pseudonymization Guidelines GU-ISS Research Reports from the Dept. of Swedish. GU-ISS 2021-02.

Megyesi, B. and Tudor, C. (2021) Transcription of Historical Ciphers and Keys. Guidelines, version 2.0. Dept. of Linguistics and Philology, Uppsala University, Sweden.

Megyesi, B., Tudor, C., Láng, B, and Lehofer, A. (2021) Key Design in the Early Modern Era in Europe. In *Proceedings of the 4th International Conference on Historical Cryptology*. HistoCrypt 2021.

Volodina, E. and Megyesi, B. (2021) SweLL Transcription Guidelines, L2 Essays GU-ISS Research Reports from the Dept. of Swedish. GU-ISS 2021-01.

## 2020

Chen, J., Souibgui, M. A., Fornés, A., and Megyesi, B. (2020). A Web-Based Interactive Transcription Tool for Encrypted Manuscripts. In *Proceedings of the 3rd International Conference on Historical Cryptology*. HistoCrypt 2020. pp. 52-59. Linköping Electronic Press.

Lasry, G., Megyesi, B., and Kopal, N. (2020). Deciphering Papal Ciphers from the 16th to the 18th Century. *Cryptologia.* Taylor and Francis. pp. 1-62. DOI: https://doi.org/10.1080/01611194.2020.1755915

Megyesi, B., Esslinger, B., Fornés, A., Kopal, N., Láng, B., Lasry, G., de Leeuw, K., Pettersson, E., Wacker A., and Waldispühl, M. (2020) Decryption of historical manuscripts: the DECRYPT project. *Cryptologia.* Taylor and Francis. pp. 1-15. DOI: 10.1080/01611194.2020.1716410

Megyesi, B. (2020) Transcription of Historical Ciphers and Keys: Guidelines. Version February 10, 2020. Department of Linguistics and Philology, Uppsala University, Sweden.

Megyesi, B. (2020) Transcription of Historical Ciphers and Keys. In *Proceedings of the 3rd International Conference on Historical Cryptology*. HistoCrypt 2020. pp. 106-115. Linköping Electronic Press.

Megyesi, B. (2020) (Editor) Proceedings of the 3rd International Conference on Historical Cryptology. NEALT Proceedings Series 44. HistoCrypt 2020. Published by Linköping Electronic Press.

Tudor, C., Megyesi, B., and Láng, B. (2020) Automatic Key Structure Extraction. In *Proceedings of the 3rd International Conference on Historical Cryptology*. HistoCrypt 2020. pp. 146-152. Linköping Electronic Press.

Volodina, E., Mohammed, Y.A., Derbring, S., Matsson, A., and Megyesi, B. (2020) Towards Privacy by Design in Learner Corpora Research: A Case of On-the-fly Pseudonymization of Swedish Learner Essays. In *Proceedings of the 28th International Conference on Computational Linguistics*. COLING 2020. pp.357-369.

## 2019

Ahrenberg, L. and Megyesi, B. eds. (2019) Proceedings of the Workshop on NLP and Pseudonymisation. September 30, 2019. NEALT Proceedings Series 166, Linköping Electronic Press and ACL anthology.

Baró A., Chen, J., Fornés, A., and Megyesi, B. (2019) Towards a Generic Unsupervised Method for Transcription of Encoded Manuscripts. In *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage (DATeCH2019)*, May 2019 Brussels, Belgium.

Megyesi, B., Palmér, A., and Näsman, J. (2019) SWEGRAM: Annotering och analys av svenska texter. Institutionen för lingvistik och filologi och Institutionen för nordiska språk, Uppsala universitet.

Megyesi, B., Blomqvist, N., and Pettersson, E. (2019) The DECODE Database: Collection of Historical Ciphers and Keys. In *Proceedings of the 2nd International Conference on Historical Cryptology*. HistoCrypt 2019, June 23-25, 2019, Mons, Belgium. NEALT Proceedings Series 37, Linköping Electronic Press.

Megyesi, B. and Volodina, E. (2019) Pseudonymization of Language Learner Data. Abstract. Workshop om pseudonymisering av textdata, 22 mars 2019, Kungliga biblioteket, Stockholm.

Pettersson, E. and Megyesi, B. (2019) Matching Keys and Encrypted Manuscripts. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics (NoDaLiDa'19)*. NoDaLiDa 2019, September 30-October 3, 2019, Turku, Finland. NEALT Proceedings Series 167, Linköping Electronic Press.

Volodina, E., Granstedt, L., Matsson, A., Megyesi, B., Pilán, I., Prentice, J. Rosén, D., Rudebeck, L., Schenström, C., Sundberg G., and Wirén, M. (2019) The SweLL Language Learner Corpus: From Design to Annotation. *Northern European Journal of Language Technology (NEJLT)*.

Yin, X., Aldarrab, N., Megyesi, B., and Knight, K. (2019) Decipherment of Historical Manuscript Images. In *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*. DOI:10.1109/ICDAR.2019.00022

**2018**

Megyesi, B. (2018) (Editor) Proceedings of the 1st International Conference on Historical Cryptology. NEALT Proceedings Series 34. HistoCrypt 2018, June 18-20, 2018, Uppsala, Sweden. Published by Linköping Electronic Press.

Megyesi, B., Granstedt, L., Johansson, S., Prentice, J., Rosén, D., Schenström, C-J., Sundberg, G., Wirén, M., Volodina, E. (2018) Learner Corpus Anonymization in the Age of GDPR: Insights from the Creation of a Learner Corpus of Swedish. In *Proceedings of the 7th NLP4CALL*, SLTC workshop, Stockholm, Sweden.

Pettersson, E. and Megyesi, B. (2018) The HistCorp Collection of Historical Corpora and Resources. In *Proceedings of the Third Conference on Digital Humanities in the Nordic Countries*, March 2018, Helsinki, Finland.

Volodina, E., Granstedt, L., Megyesi, B., Prentice, J., Rosén, D., Schenström, C-J., Sundberg, G., and Wirén, M. (2018) Annotation of learner corpora: first SweLL insights. Abstract. In *Abstracts of SLTC 2018*, Stockholm, Sweden

**2017**

Borin, L., Tahmasebi, N., Volodina, E., Ekman, S., Jordan, C., Viklund, J., Megyesi, B., Näsman, J., Palmér, A., Wirén, M., Björkenstam, N. K., Grigonyté, G., Gustafson Capková, S., and Kosiński, T. (2017) Swe-Clarin: Language Resources and Technology for Digital Humanities. In *Extended Papers of the International Symposium on Digital Humanities* , Nov. 7-8, 2016, Växjö, Sweden.

Fornés A., Megyesi, B., and Mas, J. (2017) Transcription of Encoded Manuscripts with Image Processing Techniques. In *Proceedings of Digital Humanities.* Montreal, Canada, August 8-11, 2017.

Näsman, J., Megyesi, B., and Palmér, A. (2017) SWEGRAM - A WebBased Tool for Automatic Annotation and Analysis of Swedish Texts. In *Proceedings of 21st Nordic Conference on Computational Linguistics*, Nodalida 2017.

Stymne, S., Pettersson, E., Megyesi, B., and Palmér, A. (2017) Annotating Errors in Student Texts: First Experiences and Experiments. In *Proceedings of Joint 6th NLP4CALL and 2nd NLP4LA* Nodalida workshop, May 22, 2017, Gothenburg.

**2016**

Megyesi, B., Näsman, J., and Palmér, A. (2016) The Uppsala Corpus of Student Writings - Corpus Creation, Annotation, and Analysis. In *Proceedings of Language Resources and Evaluation*, LREC 2016.

Volodina, E., Megyesi, B., Wirén, M., Granstedt, L., Prentice, J., Reichenberg, M., and Sundberg, G. (2016) A Friend in Need? Research agenda for electronic Second Language infrastructure. Abstract. *Proceedings of SLTC 2016*, Umeå, Sweden

**2015**

Csató, É. Á. & Kaşıkara, H. & Megyesi, B. & Nivre, J. (2015) Parallel corpora and Universal Dependency for Turkic. Turkic Languages 19, 259–273.

Megyesi, B. (2015) (Editor) Proceedings of the 20th Nordic Conference of Computational Linguistics. NEALT Proceedings Series 23. NODALIDA 2015, May 11-13, 2015, Vilnius, Lithuania. Published by ACL anthology and Linkoping Electronic Press.

Pettersson, E., Megyesi, B., and Nivre, J. (2015) Ranking Relevant Verb Phrases Extracted from Historical Text. In *Proceedings of the 9th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 39-47, Beijing, China, July 30, 2015.

**2014**

Antomonov, F. and Megyesi, B. (2014) Automatic Morphosyntactic Analaysis of Clinical Text. Extended abstract. Swedish Language Technology Conference, SLTC 2014.

Pettersson, E., Megyesi, B., Nivre J. (2014) A Multilingual Evaluation of Three Spelling Normalization Methods for Historical Text. In *EACL 2014 Workshop on Language Technology for Cultural Heritage, Social Sciences and Humanities*, European Association for Computational Linguistics, LaTeCH 2014, EACL 2014.

Pettersson, E., Megyesi, B., Nivre J. (2014) Verb Phrase Extraction in a Historical Context. In *The First Swedish National SWE-CLARIN Workshop*, Swedish Language Technology Conference, SLTC 2014.

Seraji, M., Jahani, C., Megyesi, B., and Nivre, J. (2014) A Persian Treebank with Stanford Typed Dependencies. In *Proceedings of Language Resources and Evaluation*, LREC 2014.

Smith, K., Megyesi, B., Velupillai, S., and Kvist, M. (2014) Professional Language in Swedish Clinical Text: Linguistic Characterization and Comparative Studies. Nordic Journal of Linguistics. Volume 37, issue 02, pp. 297-323. Cambridge University Press.

Tengstrand, L., Megyesi, B., Henriksson, A., Duneld, M., Kvist, M. (2014) EACL - Expansion of Abbreviations in CLinical Text. In EACL 2014 *Workshop on Predicting and Improving Text Readability for Target Reader Populations*, PITR 2014, EACL 2014.

**2013**

Pettersson, E., Megyesi, B., Nivre J. (2013) Normalisation of Historical Text Using Context-Sensitive Weighted Levenshtein Distance and Compound Splitting. In Proceedings of *19th Nordic Conference on Computational Linguistics (Nodalida)* 2013.

Pettersson, E., Megyesi, B., and Tiedemann J. (2013) An SMT Approach to Automatic Annotation of Historical Text. *Workshop on Computational Historical Linguistics, Nodalida* 2013.

**2012**

Pettersson, E., Megyesi, B., and Nivre, J. (2012) Rule-Based Normalisation of Historical Text - a Diachronic Study. Workshop on *Language Technology for Historical Text(s)*, KONVENS 2012, The 11th Conference on Natural Language Processing, Vienna, Austria.

Pettersson, E., Megyesi, B., and Nivre, J. (2012) Parsing the Past - Identification of Verb Constructions in Historical Text. Workshop on *Language Technology for Cultural Heritage, Social Sciences, and Humanities*, European Association for Computational Linguistics, Avignon, France.

Seraji, M., Megyesi, B., and Nivre, J. (2012) Dependency Parsers for Persian. Workshop on *Asian Language Resources*, COLING 2012, 24th International Conference on Computational Linguistics, Mumbai, India.

Seraji, M., Megyesi, B., and Nivre, J. (2012) A Basic Language Resource Kit for Persian. In *Proceedings of Language Resources and Evaluation (LREC 2012)*, May 2012.

Seraji, M., Megyesi, B., and Nivre, J. (2012) Bootstrapping a Persian Dependency Treebank. Special Issue of the Journal of Linguistic Issues in Language Technology (LiLT).

**2011**

Knight, K., Megyesi, B., and Schaefer, Ch. (2011) The Secrets of the Copiale Cipher. Journal for Research into Freemasonry and Fraternalism. Vol. 2, No 2. See also the Copiale page.

Knight, K., Megyesi, B., and Schaefer, Ch. (2011) The Copiale Cipher. Part of invited talk at *ACL Workshop on Building and Using Comparable Corpora (BUCC)*, 2011. See also the Copiale page.

**2010**

Csato, E., Kilimci, S., and Megyesi, B. (2010) Using Parallel Corpora in Data-Driven Teaching of Turkish in Sweden. In Proceedings of the *International Educational Technology Conference (IETC)*, April 2010.

Megyesi, B., Dahlqvist, B., Csato, E., and Nivre, J. (2010) The English-Swedish-Turkish Parallel Treebank. In *Proceedings of Language Resources and Evaluation (LREC 2010)*, May 2010.

**2009**

Andreasson, M., Borin, L., Forsberg, M., Merkel, M., Eriksson, A., Beskow, J., Carlson, R., Edlund, J., Elenius, K, Hellmer, K., House, D., Forsbom, E., Megyesi, B., Strömqvist, S. (2009) Swedish CLARIN Activities. *CLARIN Workshop, Nodalida 2009* May 2009.

Megyesi, B. (2009) The Open Source Tagger HunPoS for Swedish. In *Proceedings of the 17th Nordic Conference on Computational Linguistics (NODALIDA)*.

**2008**

Elenius, K., Forsbom, E. and Megyesi, B. B. (2008) Language Resources and Tools for Swedish: A Survey. In *Proceedings of Language Resources and Evaluation (LREC08)*, May 2008.

Elenius,K., Forsbom, E., and Megyesi, B. (2008) Survey on Swedish Language Resources. Report, February 2008. Dept. of Speech, Music and Hearing, KTH and Dept. of Linguistics and Philology, Uppsala University

Megyesi, B., Csato Johanson, E., Dahlqvist, B., Gustafson-Capkova, S., Nivre, J., Pettersson, E., and Sågvall Hein, A. (2008) Supporting Research Environment for Swedish and Turkish. Report, November 2008. Dept. of Linguistics and Philology, Uppsala University

Megyesi, B., Dahlqvist, B., Pettersson, E., Gustafson-Capková, S., and Nivre, J. (2008) Supporting Research Environment for Less Explored Languages: A Case Study of Swedish and Turkish. In *Resourceful Language Technology: Festschrift in Honor of Anna Sågvall Hein*. Uppsala University, Faculty of Languages, Department of Linguistics and Philology

Megyesi, B., Dahlqvist, B., Pettersson, E., and Nivre, J. (2008) Swedish-Turkish Parallel Treebank. In *Proceedings of Language Resources and Evaluation (LREC08)*, May 2008.

Nivre, J., Dahllöf, M., Megyesi, B. (eds.) (2008) *Resourceful Language Technology: Festschrift in Honor of Anna Sågvall Hein*. Uppsala University, Faculty of Languages, Department of Linguistics and Philology.

Nivre, J., Megyesi, B., Gustafson-Capková, S., Salomonsson, F., and Dahlqvist, B. (2008) Cultivating a Swedish Treebank. In *Resourceful Language Technology: Festschrift in Honor of Anna Sågvall Hein*. Uppsala University, Faculty of Languages, Department of Linguistics and Philology

Saxena, A., Megyesi, B., Csato Johanson, E., Dahlqvist, B. (2008) Using Parallel Corpora in Teaching and Research: The Swedish-Hindi-English and Swedish-Turkish-English Parallel Corpora. In *Proceedings of Swedish Linguistic Conference (SLC 2008)*.

**2007**

Alemu Argaw, A., Hulth, A., and Megyesi, B. (2007) General-Purpose Text Categorization Applied to the Medical Domain. Department of Computer and Systems Sciences, Stockholm University, Research Report 2007-016

Dahlqvist, B. and Megyesi, B. (2007) Changing the tokenization in Talbanken to SUC2.0. Working report. Department of Linguistics and Philology, Uppsala University.

Hall, J., Nilsson, J., Nivre, J., Eryigit, G., Megyesi, B., Nilsson, M. and Saers, M. (2007) Single Malt or Blended? A Study in Multilingual Parser Optimization. In *Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL 2007*, 933-939

Megyesi, B. and Dahlqvist, B. (2007) The Swedish-Turkish Parallel Corpus and Tools for its Creation. In *Proceedings of NoDaLida 2007*. May 24-26 2007, Tartu, Estonia.

Nivre, J. and Megyesi, B. (2007) Bootstrapping a Swedish Treebank Using Cross-Corpus Harmonization and Annotation Projection. In *Proceedings of Treebanks and Linguistic Theories*, December 7-8 2007, Bergen, Norway.

**2006**

Hulth, A. and Megyesi, B. (2006) A Study on Automatically Extracted Keywords in Text Categorization. In *Proceedings of International Conference of Association for Computational Linguistics, ACL 2006*. July 17-21, 2006. Sydney, Australia.

Megyesi, B. Sågvall Hein, A., and Csato Johanson, E. (2006) Building a Swedish-Turkish Parallel Corpus. In *Proceedings of Language Resources and Evaluation Conference*. May 22-28, 2006. Genoa, Italy.

**2005**

Megyesi, B. (2005) Vart tar språkteknologer vägen? En miniundersökning av f.d. språkteknologstudenters arbeten och syn på sin utbildning. Department of Linguistics and Philology, Uppsala University

Wastholm, P., Kusma, A., and Megyesi, B. (2005) Using Linguistic Data for Genre Classification. In Proceedings of SAIS-SSLS, 12th -14th April 2005, Mälardalen University, Västerås.

**2004**

Parental leave

**2003**

Heldner, M. and Megyesi, B. (2003) Exploring the Prosody-Syntax Interface in Conversations. In *Proceeding of the 15th International Congress of Phonetic Sciences (ICPhS)*, 2-9 August 2003, Barcelona, Spain.

Heldner, M. and Megyesi, B. (2003) The Acoustic and Morpho-Syntactic Context of Prosodic Boundaries in Dialogs. In *Proceedings of Fonetik 2003*. Umeå, Sweden.

**2002**

Carlson, R., Granström, B., Heldner, M., House, D., Megyesi, B., Strangert, E., Swerts, M. (2002) Boundaries and groupings - the structuring of speech in different communicative situations: a description of the GROG project. In *Proceedings of Fonetik 2002*, TMH-QPSR Vol 44, pp. 65-69.

Gustafson-Capkova, S. and Megyesi, B. (2002) Silence and Discourse Context in Read Speech and Dialogues in Swedish. In *Proceedings of the Speech Prosody 2002 conference*, 11-13 April 2002. Bernard Bel and Isabelle Marlien (eds.), pages: 363-366, Aix-en-Provence: Laboratoire Parole et Language. ISBN 2-9518233-0-4.

Megyesi, B. (2002) Shallow Parsing with PoS Taggers and Linguistic Features. Journal of Machine Learning Research: Special Issue on Shallow Parsing, JMLR (2): 639-668. MIT Press.

Megyesi, B. (2002) *Data-Driven Syntactic Analysis - Methods and Applications for Swedish*. Ph.D.Thesis. Department of Speech, Music and Hearing, KTH, Stockholm, Sweden.

Megyesi, B. and Carlson, R. (2002) Data-Driven Methods for Building a Swedish Treebank. Paper presented at the *Swedish Treebank Symposium*, 28-29 November 2002, Växjö University, Sweden.

Megyesi, B. and Gustafson-Capkova, S. (2002) Production and Perception of Pauses and their Linguistic Context in Read and Spontaneous Speech in Swedish. In *Proceedings of ICSLP'2002 - 7th International Conference on Spoken Language Processing*, Denver, USA.

**2001**

Gustafson-Capkova, S. and Megyesi, B. (2001) A Comparative Study of Pauses in Dialogues and Read Speech. In *Proceedings of Eurospeech 2001*, Volume 2, pp. 931-935, Aalborg, Danmark, September 3-7, 2001.

Megyesi, B. (2001) Data-Driven Methods for PoS tagging and Chunking of Swedish. Presented at *NoDaLiDa2001*, May 21-22 2001, Uppsala, Sweden.

Megyesi, B. (2001) Phrasal Parsing by Using Data-Driven PoS Taggers. In Proceedings of the *Conference on Recent Advances in Natural Language Processing*, Euro Conference RANLP-2001. pp. 166-173, 5-7 September 2001, Tzigov Chark, Bulgaria.

Megyesi, B. (2001) Comparing Data-Driven Learning Algorithms for PoS Tagging of Swedish. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2001)*. pp. 151-158, Carnegie Mellon University, Pittsburgh, PA, USA, June 3 and 4 2001.

Megyesi, B. and Gustafson-Capkova, S. (2001) Pausing in Dialogues and Read Speech: Speaker's Production and Listeners Interpretation. In Proceedings of the *Workshop on Prosody in Speech Recognition and Understanding.* pp. 107-113, October 22-24, 2001, New Jersey, USA.

## 2000

Berthelsen, H. and Megyesi, B. (2000) Ensemble of Classifiers for Noise Detection in PoS Tagged Corpora. In *Proceedings of the Third International Workshop on TEXT, SPEECH and DIALOGUE*, Brno, Czech Republic, September 13-16, 2000. Springer-Verlag in LNCS/LNAI series, pp. 27-32.

Megyesi, B. and Rydin, S. (2000) Towards a Finite-State Parser for Swedish. In *Proceedings of NoDaLiDa 99*, December 9-10, 1999, Trondheim, Norway, pp. 115-123.

## 1999

Megyesi, B. (1999) Improving Brill's PoS Tagger for an Agglutinative Language.
In *Proceedings of the Joint Sigdat Conference on Empirical Methods in Natural Language Processing and Very Large Corpora* (EMNLP/VLC '99), University of Maryland, USA, June 21-22, 1999, pp. 275-284.

Megyesi, B. (1999)  Brill's PoS Tagger with Extended Lexical Templates for Hungarian.
In *Proceedings of the Workshop (W01) on Machine Learning in Human Language Technology*, ACAI'99, Chania, Crete, Greece July 5 - July 16, 1999, pp. 22-28.

## Earlier

Megyesi, B. (1999) Minoriteter i Ungern. Department of Linguistics, Stockholm University

Megyesi, B and Rydin, S. (1999) Varmt och kallt: ett typologiskt ministudie. Department of Linguistics, Stockholm University

Megyesi, B. (1998) Brill's Rule-Based Part of Speech Tagger for Hungarian. D-level thesis (Master's thesis) in Computational Linguistics, Spring 1998. Computational Linguistics, Department of Linguistics, Stockholm University, Sweden.

Megyesi, B. (1998) Transformation-Based Learning: A Short Description of Brill's PoS-Tagger. Department of Linguistics, Stockholm University

Megyesi, B. (1998) A Short Descriptive Grammar for Hungarian. Department of Linguistics, Stockholm University

Megyesi, B. (1996) Implementering av partikelverb för projektet 'Datorstödd inlärning av grammatik och språkteori'. C-level thesis (Bachelor's thesis) in Computational Linguistics, Autumn 1996. Institutionen för lingvistik, Stockholms Universitet, Sverige.