Publications of Beáta Megyesi

Journal Articles

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

2. 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).

3. Knight, K., **Megyesi, B.**, and Schaefer, Ch. (2012) The Secrets of the Copiale Cipher. Journal for Research into Freemasonry and Fraternalism. 2(2)

4. 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.

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

6. 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).

7. **Megyesi, B.**, Esslinger, B., Fornés, A., Kopal, N., Láng, B., Lasry, G., de Leeuw, K., Pettersson, E., Wacker A., and Waldispühl, M. (2020) Decryp- tion of historical manuscripts: the DECRYPT project. Cryptologia. DOI: doi.org/10.1080/01611194.2020.1716410.

8. Lasry, G., **Megyesi, B.**, and Kopal, N. (2020). Deciphering Papal Ciphers from the 16th to the 18th Century. Cryptologia. DOI: doi.org/10.1080/01611194.2020.1755915

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

10. **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. Cryptologia 48:2. pp. 97-139. Taylor and Francis. DOI: doi.org/10.1080/01611194.2022.2113185

11. 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

Book Chapters

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

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

3. **Megyesi, B.**, Fornés, A., Kopal, N., Láng, B., Waldispühl, M., Mikhalev, V., and Esslinger, B. (2024) *Historical Cryptology.* Chapter 3 from the book "Learning and Experiencing Cryptography with CrypTool and SageMath", Artech House.

Edited Volumes

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

2. **Megyesi, B.** (editor) (2015) *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 Linköping Electronic Press.

3. **Megyesi, B.** (editor) (2018) *Proceedings of the 1st International Confe- rence on Historical Cryptology HistoCrypt 2018.* NEALT Proceedings Se- ries 34. HistoCrypt 2018, June 18-20, 2018, Uppsala, Sweden. Published by Linköping Electronic Press.

4. Ahrenberg, L. and **Megyesi, B.** (editors) (2019) *Proceedings of the Works- hop on NLP and Pseudonymisation.* September 30, 2019. NEALT Procee- dings Series 166, Linköping Electronic Press and ACL anthology.

5. **Megyesi, B.** (2020) (editor) *Proceedings of the 3rd International Confe- rence on Historical Cryptology.* NEALT Proceedings Series 44. HistoCrypt 2020. Published by Linköping Electronic Press.

6. Dahlke, C. and **Megyesi, B.** (editors) Proceedings of the 5th Internatio- nal Conference on Historical Cryptology. NEALT Proceedings Series 49. HistoCrypt 2022. Published by Linköping Electronic Press.

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

8. Waldispühl, M. and **Megyesi, B**. (eds.) (2024) Proceedings of the 7th International Conference on Historical Cryptology (HistoCrypt 2024) NEALT Proceedings Series 53.

Dissertations

1. **Megyesi, B.** (2002) *Data-Driven Syntactic Analysis — Methods and Applications for Swedish.* Doctoral Dissertation, Department of Speech, Mu- sic and Hearing, Kungliga Tekniska Högskolan, Stockholm, Sweden.

Refereed Conference Papers

1. **Megyesi, B.** (1999) Brill's PoS Tagger with Extended Lexical Templates for Hungarian. *Workshop (W01) on Machine Learning in Human Langu- age Technology, ACAI'99*, Chania, Crete, Greece July 5 - July 16, 1999.

2. **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),* pp. 275–284, University of Maryland, USA, June 21–22, 1999.

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

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

5. **Megyesi, B.** (2001) Data-Driven Methods for PoS tagging and Chunking of Swedish. *In Proceedings of NoDaLiDa 2001,* Uppsala, Sweden, May 21–22, 2001.

6. Gustafson-Čapková, 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, Denmark, September 3–7, 2001.

7. **Megyesi, B.** and Gustafson-Čapková, S. (2001) Pausing in Dialogues and Read Speech: Speakers' Production and Listeners' Interpretation. *In Proceedings of the Workshop on Prosody in Speech Recognition and Un- derstanding,* pp. 107–113, New Jersey, USA, October 22–24, 2001.

8. **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 3–4, 2001.

9. **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, Tzigov Chark, Bulgaria, September 5–7, 2001

10. Gustafson-Čapková, S. and **Megyesi, B.** (2002) Silence and Discourse Context in Read Speech and Dialogues in Swedish. *In Proceedings of the Speech Prosody 2002 conference,* Bernard Bel and Isabelle Marlien (eds.), pp. 363–366, Aix-en-Provence, France, April 11–13, 2002.

11. Carlson, R., Granström, B., Heldner, M., House, D., **Megyesi, B.**, Strang- ert, E. and 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 Volume 44, pp. 65– 69, Stockholm, Sweden, May 29–31, 2002.

12. **Megyesi, B.** and Gustafson-Čapková, S. (2002) Production and Percep- tion of Pauses and their Linguistic Context in Read and Spontaneous Speech in Swedish. *In Proceedings of ICSLP'2002 - 7th International Con- ference on Spoken Language Processing,* Denver, USA, September 16–20, 2002.

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

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

15. Heldner, M. and **Megyesi, B.** (2003) *The Acoustic and Morpho-Syntactic Context of Prosodic Boundaries in Dialogs.* In Proceeding of Fonetik 2003, 2-3 June 2003, Umeå, Sweden.

16. Wastholm, P., Kusma, A., and **Megyesi, B.** (2005) Using Linguistic Data for Genre Classification. In *Proceedings of Swedish Artificial Intelligence and Learning Systems SAIS-SSLS 2005.*

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

18. Hulth, A. and **Megyesi, B.** (2006) A Study on Automatically Extracted Keywords in Text Categorization. In *Proceedings of Association for Computational Linguistics*, ACL 2006.

19. **Megyesi, B.** and Dahlqvist, B. (2007) The Swedish-Turkish Parallel Cor- pus and Tools for its Creation. In *Proceedings of NoDaLida 2007*.

20. 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*, pp. 933-939.

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

22. **Megyesi, B.**, Dahlqvist, B. Pettersson, E. and Nivre, J. (2008) Swedish- Turkish Parallel Treebank. In *Proceedings of Language Resources and Eva- luation Conference*, LREC 2008.

23. Elenius, K. Forsbom, E. and **Megyesi, B.** (2008) Swedish Language Re- sources. In *Proceedings of Language Resources and Evaluation Conference*, LREC 2008.

24. 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.

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

26. Andreasson, M., Borin, L., Forsberg, M., Merkel, M., Eriksson, A., Be- skow, J., Carlson, R., Edlund, J., Elenius, K, Hellmer, K., House, D., Forsbom, E., **Megyesi, B.**, Strömqvist, S. (2009) *Swedish CLARIN Ac- tivities. CLARIN Workshop*, Nodalida 2009.

27. **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.

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

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

30. Pettersson, E., **Megyesi, B.**, and Nivre, J. (2012) Parsing the Past - Identification of Verb Constructions in Historical Text. In *Proceedings of the Workshop on Language Technology for Cultural Heritage, Social Sci- ences, and Humanities*, European Association for Computational Lingu- istics, EACL 2012.

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

32. Pettersson, E., **Megyesi, B.**, and Nivre, J. (2012) Rule-Based Normalisa- tion of Historical Text - a Diachronic Study. In Proceedings of the *Works- hop on Language Technology for Historical Text(s)*, 11th Conference on Natural Language Processing, KONVENS 2012, Vienna, Austria.

33. Seraji, M., **Megyesi, B.**, and Nivre, J. (2012) Dependency Parsers for Per- sian. In the Proceedings of the *Workshop on Asian Language Resources*, 24th International Conference on Computational Linguistics, COLING 2012.

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

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

36. 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, La- TeCH 2014, EACL 2014.

37. Seraji,M., Jahani, C., **Megyesi, B.**, and Nivre, J. (2014) A Persian Tree- bank with Stanford Typed Dependencies. In *Proceedings of Language Re- sources and Evaluation*, LREC 2014.

38. Tengstrand, L., **Megyesi, B.**, Henriksson, A., Duneld, M., and 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.

39. Antomonov, F. and **Megyesi, B.** (2014) Automatic Morphosyntactic Analysis of Clinical Text. Extended abstract. *Swedish Language Techno- logy Conference*, SLTC 2014.

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

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

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

43. Borin, L., Tahmasebi, N., Volodina, E., Ekman, S., Jordan, C., Viklund, J., **Megyesi, B.**, Näsman, J., Palmér, A., Wirén, M., Björkenstam, N.

44. K., Grigonyté, G., Gustafson Capková, S., and Kosiński, T. (2017) Swe- Clarin: Language Resources and Technology for Digital Humanities. In *Proceedings of International Digital Humanities Symposium* ,7-8 Nov 2016, Växjö, Sweden.

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

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

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

48. Pettersson, E. and **Megyesi, B.** (2018) The HistCorp Collection of Histo- rical Corpora and Resources. In *Proceedings of the Third Conference on Digital Humanities in the Nordic Countries*, March 2018, Helsinki, Fin- land.

49. **Megyesi, B.**, Granstedt, L., Johansson, S., Prentice, J., Rosén, D., Schen- ström, C-J., Sundberg, G., Wirén, M., and 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.

50. 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.

51. **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.

52. Pettersson, E. and **Megyesi, B.**. Matching Keys and Encrypted Ma- nuscripts. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, NoDaLiDa 2019.

53. Yin, X., Aldarrab, N., Knight k., and **Megyesi, B.**, Decipherment of Historical Manuscript Images. In *Proceedings of the 15th International Conference on Document Analysis and Recognition*, September 20–25, 2019, Sydney, Australia.

54. Chen, J., Souibgui, M. A., Fornés, A., and **Megyesi, B.** (2020). A Web- Based Interactive Transcription Tool for Encrypted Manuscripts. In *Pro- ceedings of the 3rd International Conference on Historical Cryptology (Hi- stoCrypt 2020)*. NEALT Proceedings Series 44, Linköping Electronic Press.

55. **Megyesi, B.** (2020) Transcription of Historical Ciphers and Keys. In *Proceedings of the 3rd International Conference on Historical Cryptology. (HistoCrypt 2020)*. NEALT Proceedings Series 44, Linköping Electronic Press.

56. Tudor, C., **Megyesi, B.**, and Láng, B. (2020) Automatic Key Structure Extraction. In *Proceedings of the 3rd International Conference on Histo- rical Cryptology. (HistoCrypt 2020)*. NEALT Proceedings Series 44, Lin- köping Electronic Press.

57. Volodina, E., Mohammed, Y.A., Derbring, S., Matsson, A., and **Megy- esi, 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 Lin- guistics. COLING 2020*. pp. 357-369.

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

59. **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 HistCryptology. HistoCrypt 2021*.

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

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

62. 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*.

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

64. 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.

65. 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.

66. **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.

67. 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/10.1007/978-3-031-37717-411

68. 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

69. 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

70. **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

71. 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

Reports, other papers

1. **Megyesi, B.** (1996) *Implementering av partikelverb för projektet Dator- stödd inlärning av grammatik och språkteori.* C-level thesis (Bachelor's thesis) in Computational Linguistics, Autumn 1996. Institutionen för ling- vistik, Stockholm University, Sverige.

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

3. **Megyesi, B.** (1998) *Brill's Transformation-Based Tagger.* Department of Linguistics, Stockholm University, Sweden.

4. **Megyesi, B.** (1998) *A Short Descriptive Grammar for Hungarian.* Dept. of Linguistics, Stockholm University, Sweden.

5. **Megyesi, B.** (2005) *Vart tar språkteknologer vägen? En miniundersök- ning av f.d.*

språkteknologstudenters arbeten och syn på sin utbildning. Department of Linguistics and Philology, Uppsala University, Sweden.

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

7. Dahlqvist, B. and **Megyesi, B.** (2007) *Changing the tokenization in Tal- banken to SUC2.0.* Working report. Department of Linguistics and Philo- logy, Uppsala University, Sweden.

8. Elenius, K. Forsbom, E. and **Megyesi, B.** (2008) *Survey on Swedish Lan- guage Resources.* Report, February 2008. Dept. of Speech, Music and He- aring, KTH and Dept. of Linguistics and Philology, Uppsala University, Sweden.

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

10. Volodina, E., **Megyesi, B.**, Wirén, M., Granstedt, L., Prentice, J., Reichen- berg, M., and Sundberg, G. (2016) *A Friend in Need? Research agenda for electronic Second Language infrastructure.* Abstract. Swedish Langu- age Technology Conference (SLTC) 2016, Umeå, Sweden.

11. Volodina, E., Granstedt, L., Johansson, S., **Megyesi, B.**, Prentice, J., Pilán, I., Rosén, D., Schenström, C-J., Sundberg, G., and Wirén, M. (2018) *Annotation of Learner Corpora–in (SweLL) practice.* Abstract. Swedish Language Technology Conference (SLTC) 2018, Stockholm Uni- versity, Sweden.

12. **Megyesi, B.**, Palmér, A., and Näsman, J. (2019) *SWEGRAM: Annote- ring och analys av svenska texter.* Institutionen för lingvistik och filologi och Institutionen för nordiska språk, Uppsala universitet, Sweden.

13. **Megyesi, B.** and Volodina, E. (2019) *Pseudonymization of Language Le- arner Data.* Abstract. Workshop om pseudonymisering av textdata, March 22, 2019. Kungliga biblioteket, Stockholm, Sweden.

14. **Megyesi, B.** (2020) *Transcription of Historical Ciphers and Keys: Guide- lines.* Version February 10, 2020. Department of Linguistics and Philology, Uppsala University, Sweden.

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

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

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

18. **Megyesi, B.**, Rudebeck, L., and Volodina, E. (2021) SweLL Pseudony- mization