Publications

CLARA Marie Curie Initial Training Network

June 20, 2014

References

[1] Núria Bel, Muntsa Padró, and Silvia Necşulescu. A method towards the fully automatic merging of lexical resources. In Nicoletta Calzolari, Toru Ishida, Stelios Piperidis, and Virach Sornlertlamvanich, editors, *Proceedings of the Workshop on Language Resources, Technology and Services in the Sharing Paradigm*, pages 8–15, Chiang Mai, Thailand, November 2011. Asian Federation of Natural Language Processing. URL `http://www.aclweb.org/anthology-new/W/W11/W11-3302.pdf`.

[2] Serhiy Bykh, Sowmya Vajjala, Julia Krivanek, and Detmar Meurers. Combining shallow and linguistically motivated features in native language identification. In *Proceedings of the 8th Workshop on Innovative Use of NLP for Building Educational Applications (BEA)*, Atlanta, GA, USA, 2013. Association for Computational Linguistics. URL `http://purl.org/dm/papers/Bykh.Vajjala.ea-13.html`.

[3] Koenraad De Smedt, Erhard Hinrichs, Detmar Meurers, Inguna Skadiņa, Bolette Sandford Pedersen, Costanza Navarretta, Núria Bel, Krister Lindén, Markéta Lopatková, Jan Hajič, Gisle Andersen, and Przemysław Lenkiewicz. CLARA: A new generation of researchers in common language resources and their applications. In Nicoletta (conference chair) Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asunción Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 2166–2174, Reykjavik, Iceland, May 2014. European Language Resources Association (ELRA). URL `http://www.lrec-conf.org/proceedings/lrec2014/pdf/410_Paper.pdf`.

[4] Corina Dima. Intui2: A prototype system for question answering over linked data. In Pamela Forner, Roberto Navigli, and Dan Tufis, editors, *CLEF 2013 Evaluation Labs and Workshop, Online Working Notes*, Valencia, Spain, 2013. ISBN 978-88-904810-5-5. URL `http://www.clef-initiative.eu/documents/71612/e2ceddb7-db03-46a2-832f-b9ddfb4ff842`.

[5] Corina Dima and Erhard Hinrichs. A semi-automatic, iterative method for creating a domain-specific treebank. In *Proceedings of the 8th International Conference on Recent*

Advances in Natural Language Processing (RANLP 2011), pages 413–419, Hissar, Bulgaria, September 2011. RANLP 2011 Organising Committee. URL `http://aclweb.org/anthology-new/R/R11/R11-1057.pdf`.

[6] Cheikh M. Bamba Dione. An LFG approach to Wolof cleft constructions. In Miriam Butt and Tracy Holloway King, editors, *Proceedings of the LFG '12 Conference*, LFG Online Proceedings, pages 157–176, Stanford, CA, 2012. CSLI Publications. URL `http://cslipublications.stanford.edu/LFG/17/papers/lfg12dione.pdf`.

[7] Cheikh M. Bamba Dione. A morphological analyzer for Wolof using finite-state techniques. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 894–901, Istanbul, Turkey, May 2012. European Language Resources Association (ELRA). ISBN 978-2-9517408-7-7. URL `http://www.lrec-conf.org/proceedings/lrec2012/pdf/572_Paper.pdf`.

[8] Cheikh M. Bamba Dione. Handling Wolof clitics in LFG. In Christine Meklenborg Salvesen and Hans Petter Helland, editors, *Challenging Clitics*, pages 87–118. John Benjamins Publishing Company, Amsterdam, 2013. URL `http://benjamins.com/#catalog/books/la.206.04dio/details`.

[9] Cheikh M. Bamba Dione. Valency change and complex predicates in Wolof: An LFG account. In Miriam Butt and Tracy Holloway King, editors, *Proceedings of the LFG '13 Conference*, Stanford, CA, 2013. CSLI Publications. URL `http://www.stanford.edu/group/cslipublications/cslipublications/LFG/18/lfg13.html`.

[10] Cheikh M. Bamba Dione. Pruning the search space of the Wolof LFG grammar using a probabilistic and a constraint grammar parser. In Nicoletta (conference chair) Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asunción Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 2863–2870, Reykjavik, Iceland, May 2014. European Language Resources Association (ELRA). URL `http://www.lrec-conf.org/proceedings/lrec2014/pdf/62_Paper.pdf`.

[11] Senka Drobac, Miikka Silfverberg, and Anssi Mikael Yli-Jyrä. Implementation of replace rules using preference operator. *Proceedings of the International Workshop on Finite State Methods in Natural Language Processing 2012*, pages 55–59, 2012. URL `http://www.aclweb.org/anthology/W12-6210`.

[12] Senka Drobac, Krister Lindén, Tommi A. Pirinen, and Miikka Silfverberg. Heuristic hyper-minimization of finite state lexicons. In Nicoletta (conference chair) Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asunción Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3319–3324, Reykjavik, Iceland, May 2014. European Language Resources Association (ELRA). URL `http://www.lrec-conf.org/proceedings/lrec2014/pdf/784_Paper.pdf`.

[13] Jakob Elming, Anders Johannsen, Sigrid Klerke, Emanuele Lapponi, Héctor Martínez Alonso, and Anders Søgaard. Down-stream effects of tree-to-dependency conversions. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 617–626, Atlanta, Georgia,

June 2013. Association for Computational Linguistics. URL
`http://www.aclweb.org/anthology/N13-1070`.

[14] Binyam Gebrekidan Gebre, Peter Wittenburg, and Przemyslaw Lenkiewicz. Towards automatic gesture stroke detection. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, May 2012. European Language Resources Association (ELRA). ISBN 978-2-9517408-7-7. URL
`http://www.lrec-conf.org/proceedings/lrec2012/pdf/454_Paper.pdf`.

[15] Binyam Gebrekidan Gebre, Peter Wittenburg, and Tom Heskes. Automatic signer diarization – the mover is the signer approach. In *Proceedings of CVPR Workshop on Language for Vision*, 2013. URL `http://pubman.mpdl.mpg.de/pubman/item/escidoc:1740130:3/component/escidoc:1740129/PID2778045.pdf`.

[16] Binyam Gebrekidan Gebre, Peter Wittenburg, and Tom Heskes. Automatic signer diarization – the mover is the signer approach. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference*, pages 283–287. IEEE, 2013. doi: 10.1109/CVPRW.2013.49.

[17] Binyam Gebrekidan Gebre, Peter Wittenburg, and Tom Heskes. The gesturer is the speaker. In *Proceedings of ICASSP 2013*, pages 3751–3755. IEEE, 2013. URL
`http://pubman.mpdl.mpg.de/pubman/item/escidoc:1583089:6/component/escidoc:1711860/Gebre_ICASSP_2013.pdf`.

[18] Binyam Gebrekidan Gebre, Peter Wittenburg, and Tom Heskes. Automatic sign language identification. In *Proceedings of ICIP 2013*, pages 2626–2630, 2013. URL
`http://pubman.mpdl.mpg.de/pubman/item/escidoc:1745698:9/component/escidoc:1867912/published_paper.pdf`.

[19] Binyam Gebrekidan Gebre, Marcos Zampieri, Peter Wittenburg, and Tom Heskes. Improving native language identification with tf-idf weighting. In *Proceedings of the 8th NAACL Workshop on Innovative Use of NLP for Building Educational Applications (BEA8)*, 2013. URL
`http://pubman.mpdl.mpg.de/pubman/item/escidoc:1740046:9/component/escidoc:1740045/cologne_nijmegen_naaclhlt2013.pdf`.

[20] Binyam Gebrekidan Gebre, Peter Wittenburg, Tom Heskes, and Sebastian Drude. Motion history images for online speaker/signer diarization. In *Proceedings of the 2014 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2014.

[21] Nathan Green. Effects of noun phrase bracketing in dependency parsing and machine translation. In *Proceedings of the ACL 2011 Student Session*, pages 69–74, Portland, OR, USA, June 2011. Association for Computational Linguistics. URL
`http://www.aclweb.org/anthology/P11-3013`.

[22] Nathan Green. Dependency parsing. In *WDS 2011 Proceedings of Contributed Papers*, pages 137–142, Prague, Czech Republic, June 2011.

[23] Nathan Green. Building parallel corpora through social network gaming. In Nancy Ide, Collin Baker, Christiane Fellbaum, and Rebecca Passoneau, editors, *Proceedings of the Collaborative Resource Development and Delivery Workshop at LREC 2012*, pages 22–25. European Language Resources Association (ELRA), May 2012. URL
`http://www.lrec-conf.org/proceedings/lrec2012/workshops/22.LREC%202012%20Collaborative%20Resources%20Proceedings.pdf`.

[24] Nathan Green, Septina Dian Larasati, and Zdeněk Žabokrtský. Indonesian dependency treebank: Annotation and parsing. In *Proceedings of the 26th Pacific Asia Conference on Language, Information, and Computation (PACLIC)*, pages 137–145, Bali, Indonesia, November 2012. Faculty of Computer Science, Universitas Indonesia. URL `http://www.aclweb.org/anthology/Y12-1014`.

[25] Nathan Green, Loganathan Ramasamy, and Zdeněk Žabokrtský. Using an SVM Ensemble System for improved Tamil dependency parsing. In *Proceedings of the ACL Joint Workshop on Statistical Parsing and Semantic Processing of Morphologically Rich Languages*, pages 72–77, Jeju, Republic of Korea, July 2012. Association for Computational Linguistics. URL `http://aclweb.org/anthology-new/W/W12/W12-3410.pdf`.

[26] Nathan Green and Zdeněk Žabokrtský. Ensemble parsing and its effect on machine translation. Technical Report 48, Charles University, Prague, Czech Republic, 2012.

[27] Nathan Green and Zdeněk Žabokrtský. Hybrid combination of constituency and dependency trees into an Ensemble Dependency Parser. In *Workshop on Innovative Hybrid Approaches to the Processing of Textual Data*, pages 19–26, Avignon, France, April 2012. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/W/W12/W12-0503`.

[28] Nathan Green and Zdeněk Žabokrtský. Improvements to syntax-based machine translation using ensemble dependency parsers. In Marta R. Costa-jussà, Reinhard Papp, Patrik Lambert, Kurt Eberle, Rafael E. Banchs, and Bogdan Babych, editors, *Proceedings of the Second Workshop on Hybrid Approaches to Translation (at the 51st Annual Meeting of the Association for Computational Linguistics)*, pages 19–24, Sofia, Bulgaria, August 2013. Bălgarska akademija na naukite, Association for Computational Linguistics. ISBN 978-1-937284-63-3.

[29] Nathan David Green. *Improvements to Syntax-based Machine Translation using Ensemble Dependency Parsers*. PhD thesis, Charles University in Prague, Prague, Czech Republic, September 2013.

[30] Julia Hancke, Sowmya Vajjala, and Detmar Meurers. Readability classification for German using lexical, syntactic, and morphological features. In Martin Kay and Christian Boitet, editors, *Proceedings of the 24th International Conference on Computational Linguistics (COLING)*, volume Technical Papers, pages 1063–1080, Mumbai, India, December 2012. The COLING 2012 Organizing Committee. URL `http://aclweb.org/anthology-new/C/C12/C12-1065.pdf`.

[31] Mans Hulden. Constraint grammar parsing with left and right sequential finite transducers. In *Proceedings of the International Workshop on Finite State Methods in Natural Language Processing 2011*, July 2011. URL `http://dingo.sbs.arizona.edu/~mhulden/fsmnlp_cg_2011.pdf`.

[32] Mans Hulden. Finite-State Technology. In Ruslan Mitkov, editor, *The Oxford Handbook of Computational Linguistics*. Oxford University Press, 2nd edition, 2011.

[33] Mans Hulden. Parsing CFGs and PCFGs with a Chomsky-Schützenberger representation. In *Human Language Technology. Challenges for Computer Science and Linguistics*, volume 6562 of *Lecture Notes in Artificial Intelligence*, pages 151–160. Springer, 2011. URL `http://dingo.sbs.arizona.edu/~mhulden/hulden_ltc_2009.pdf`.

[34] Mans Hulden, Iñaki Alegria, Izaskun Etxeberria, and Montse Maritxalar. Learning word-level dialectal variation as phonological replacement rules using a limited parallel corpus. In *Proceedings of the First Workshop on Algorithms and Resources for Modelling of Dialects and Language Varieties*, pages 39–48, Edinburgh, Scotland, July 2011. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/W11-2605`.

[35] Anders Johannsen, Héctor Martínez Alonso, Christian Rishøj, and Anders Søgaard. Shared task system description: Frustratingly hard compositionality prediction. In *Proceedings of the Workshop on Distributional Semantics and Compositionality*, pages 29–32, Portland, Oregon, USA, June 2011. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/W11-1305`.

[36] Anders Johannsen, Héctor Martínez, Sigrid Klerke, and Anders Søgaard. EMNLP@CPH: Is frequency all there is to simplicity? In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics*, volume 2, pages 408–412, Montréal, Canada, June 2012. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/S12-1054`.

[37] Septina Dian Larasati. Improving word alignment by exploiting adapted word similarity. In *Proceedings of the Workshop on Monolingual Machine Translation at AMTA 2012*, pages 41–45, San Diego, CA, Octoberctober 2012. The Association for Machine Translation in the Americas.

[38] Septina Dian Larasati. IDENTIC corpus: Morphologically enriched Indonesian-English parallel corpus. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, May 2012. European Language Resources Association (ELRA). ISBN 978-2-9517408-7-7. URL `http://www.lrec-conf.org/proceedings/lrec2012/pdf/644_Paper.pdf`.

[39] Septina Dian Larasati. Handling Indonesian clitics: A dataset comparison for an Indonesian-English statistical machine translation system. In *Proceedings of the 26th Pacific Asia Conference on Language, Information, and Computation (PACLIC)*, pages 146–152, Bali, Indonesia, November 2012. Faculty of Computer Science, Universitas Indonesia. URL `http://www.aclweb.org/anthology/Y12-1015`.

[40] Septina Dian Larasati. Towards an Indonesian-English SMT system: A case study of an under-studied and under-resourced language, Indonesian. In *WDS'12 Proceedings of Contributed Papers*, volume I, pages 123–129, Prague, Czech Republic, May 2012. MatfyzPress. URL `http://www.mff.cuni.cz/veda/konference/wds/proc/pdf12/WDS12_121_i3_Larasati.pdf`.

[41] Anna Lenkiewicz and Sebastian Drude. Automatic annotation of linguistic 2D and Kinect recordings with the Media Query Language for Elan. In *Proceedings of the Digital Humanities 2013 Conference*, 2013.

[42] Anna Lenkiewicz, Magdalena Lis, and Przemyslaw Lenkiewicz. Linguistic concepts described with Media Query Language for automated annotation. In J. C. Meiser, editor, *Proceedings of the Digital Humanities 2012 Conference*, pages 477–479, Hamburg, July 2012. URL `http://hdl.handle.net/11858/00-001M-0000-000E-78CF-8`.

[43] Przemyslaw Lenkiewicz, Peter Wittenburg, Binyam Gebrekidan Gebre, Anna Lenkiewicz, Oliver Schreer, and Stefano Masneri. Application of video processing methods for linguistic research. In Zygmunt Vetulani, editor, *Human language technologies as a challenge for computer science and linguistics. Proceedings of the 5th Language and Technology Conference (LTC)*, pages 561–564, Poznań, Poland, November 2011. Adam Mickiewicz University. URL `http://pubman.mpdl.mpg.de/pubman/item/escidoc:1239560:5/component/escidoc:1239573/Lenkiewicz_et_al_2011_LTC_2011.pdf`.

[44] Przemyslaw Lenkiewicz, Binyam Gebrekidan Gebre, Oliver Schreer, Stefano Masneri, Daniel Schneider, and Sebastian Tschöpel. AVATecH — automated annotation through audio and video analysis. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, May 2012. European Language Resources Association (ELRA). ISBN 978-2-9517408-7-7. URL `http://www.lrec-conf.org/proceedings/lrec2012/pdf/303_Paper.pdf`.

[45] Przemyslaw Lenkiewicz, Dieter Van Uytvanck, Peter Wittenburg, and Sebastian Drude. Towards automated annotation of audio and video recordings by application of advanced web-services. In *13th Annual Conference of the International Speech Communication Association (InterSpeech)*, Portland, OR, USA, 2012. ISCA.

[46] Krister Lindén, Erik Axelson, Senka Drobac, Sam Hardwick, Juha Kuokkala, Jyrki Niemi, Tommi Pirinen, and Miikka Silfverberg. HFST — a system for creating NLP tools. In Cerstin Mahlow and Michael Piotrowski, editors, *Systems and Frameworks for Computational Morphology*, volume 380 of *Communications in Computer and Information Science*, pages 53–71. Springer Verlag, 2013. URL `http://link.springer.com/content/pdf/10.1007%2F978-3-642-40486-3_4.pdf`.

[47] Krister Lindén, Erik Axelson, Senka Drobac, Sam Hardwick, Miikka Silfverberg, and Tommi Pirinen. Using HFST for creating computational linguistic applications. In Adam Przepiórkowski, Maciej Piasecki, Krzysztof Jassem, and Piotr Fuglewicz, editors, *Computational Linguistics — Applications*, volume 458 of *Studies in Computational Intelligence*, pages 3–25. Springer Verlag, 2013. URL `http://link.springer.com/content/pdf/10.1007%2F978-3-642-34399-5_1.pdf`.

[48] Magdalena Lis. Annotation scheme for multimodal communication: Employing plWordNet 1.5. In *Proceedings of the Formal and Computational Approaches to Multimodal Communication Workshop. 24th European Summer School in Logic, Language and Information (ESSLLI 2012)*, Opole, Poland, August 2012. CEUR-WS.org. URL `http://xerxes.carleton.ca/~giorgolo/gesture-workshop/paper5.pdf`.

[49] Magdalena Lis. Influencing gestural representation of eventualities: insights from ontology. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction (ICMI'12)*, pages 281–288, New York, NY, 2012. ACM. ISBN 978-1-4503-1467-1. doi: 10.1145/2388676.2388733. URL `http://doi.acm.org/10.1145/2388676.2388733`.

[50] Magdalena Lis. Polish multimodal corpus — a collection of referential gestures. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, May 2012.

European Language Resources Association (ELRA). ISBN 978-2-9517408-7-7. URL
`http://www.lrec-conf.org/proceedings/lrec2012/pdf/737_Paper.pdf`.

[51] Magdalena Lis. *Multimodal representation of entities: A corpus-based investigation of co-speech hand gesture.* PhD thesis, University of Copenhagen, Copenhagen, Denmark, 2014.

[52] Magdalena Lis and Costanza Navarretta. Classifying the form of iconic hand gestures from the linguistic categorization of co-occurring verbs. In *Proceedings of the 1st European Symposium on Multimodal Communication (MMSym'13)*, Valletta, Malta, forthcoming.

[53] Magdalena Lis and Fey Parrill. Referent type and its verbal and gestural representation: A test on English multimodal corpus and wordnet3.1. In *Proceedings of 1st European Symposium on Multimodal Communication (MMSym'13)*, Valletta, Malta, forthcoming.

[54] Gunn Inger Lyse, Carla Parra Escartín, and Koenraad De Smedt. Applying current metadata initiatives: The META-NORD experience. In Victoria Arranz, Daan Broeder, Bertrand Gaiffe, Maria Gavrilidou, Monica Monachini, and Thorsten Trippel, editors, *Proceedings of the Workshop on Describing LRs with Metadata: Towards Flexibility and Interoperability in the Documentation of LRs*, pages 13–20, Istanbul, Turkey, May 2012. European Language Resources Association (ELRA). URL `http://www.lrec-conf.org/proceedings/lrec2012/workshops/11.LREC2012%20Metadata%20Proceedings.pdf`.

[55] Jianqiang Ma. Automatic annotation of syntactic categories in Chinese word structures. In Nicoletta (conference chair) Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asunción Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 4087–4092, Reykjavik, Iceland, May 2014. European Language Resources Association (ELRA). URL
`http://www.lrec-conf.org/proceedings/lrec2014/pdf/1158_Paper.pdf`.

[56] Jianqiang Ma and Dale Gerdemann. Phrase-based approach for adaptive tokenization. In *Proceedings of the Twelfth Meeting of the Special Interest Group on Computational Morphology and Phonology*, pages 17–25, Montréal, Canada, June 2012. Association for Computational Linguistics. URL
`http://aclweb.org/anthology/W/W12/W12-2303.pdf`.

[57] Jianqiang Ma and Dale Gerdemann. Distributional evidence and beyond: the success and limitations of machine learning in Chinese word segmentation. In *Proceedings of the 14th International Conference on Intelligent Text Processing and Computational Linguistics (Poster Session)*, Samos, Greece, March 2013. Springer.

[58] Jianqiang Ma and Dale Gerdemann. Distributional evidence and beyond: the success and limitations of machine learning in Chinese word segmentation. *Research in Computing Science*, 70:17–30, 2013.

[59] Jianqiang Ma, Chunyu Kit, and Dale Gerdemann. Semi-automatic annotation of Chinese word structure. In *Proceedings of the Second CIPS-SIGHAN Joint Conference on Chinese Language Processing*, pages 9–17, Tianjin, China, December 2012. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/W12-6304`.

[60] David Mareček, Martin Popel, Loganathan Ramasamy, Jan Štěpánek, Daniel Zeman, Zdeněk Žabokrtský, and Jan Hajič. Cross-language study on influence of coordination style on dependency parsing performance. Úfal Technical Report TR-2013-49, Charles University, Prague, Czech Republic, 2013. URL `http://ufal.mff.cuni.cz/techrep/tr49.pdf`.

[61] Héctor Martínez Alonso. *Annotation of regular polysemy: an empirical assessment of the underspecified sense.* PhD thesis, University of Copenhagen (Denmark) and Pompeu Fabra University (Spain), December 2013.

[62] Héctor Martínez Alonso and Lauren Romeo. Crowdsourcing as a preprocessing for complex semantic annotation tasks. In Nicoletta (conference chair) Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asunción Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 229–234, Reykjavik, Iceland, May 2014. European Language Resources Association (ELRA).

[63] Héctor Martínez Alonso, Núria Bel, and Bolette Sandford Pedersen. Identification of sense selection in regular polysemy using shallow features. In Bolette Sandford Pedersen, Gunta Nešpore, and Inguna Skadiņa, editors, *Proceedings of the 18th Nordic Conference of Computational Linguistics, Riga (NoDaLiDa 2011)*, number 11 in NEALT Proceedings Series, pages 18–25. Northern European Association for Language Technology, 2011. URL `http://hdl.handle.net/10062/17285`.

[64] Héctor Martínez Alonso, Núria Bel, and Bolette Sandford Pedersen. A voting scheme to detect semantic underspecification. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, May 2012. European Language Resources Association (ELRA). ISBN 978-2-9517408-7-7. URL `http://www.lrec-conf.org/proceedings/lrec2012/pdf/225_Paper.pdf`.

[65] Héctor Martínez Alonso, Bolette Sandford Pedersen, and Núria Bel. Annotation of regular polysemy and underspecification. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, 4-9 August 2013, Sofia, Bulgaria*, volume 2, pages 725–730, Sofia, Bulgaria, August 2013. Association for Computational Linguistics. URL `http://aclweb.org/anthology/P/P13/P13-2127.pdf`.

[66] Germán Mira and Pedro Patiño. La traducción de la fraseología especializada. In Marisela Colín, editor, *Manual de traducción de textos especializados. Nuevos enfoques, nuevas metodologías*, pages 20–35. UNAM, México, forthcoming.

[67] Costanza Navarretta and Magdalena Lis. Transfer learning of feedback head expressions in comparable multimodal Danish and Polish corpora. In Nicoletta (conference chair) Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asunción Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3597–3603, Reykjavik, Iceland, May 2014. European Language Resources Association (ELRA). URL `http://www.lrec-conf.org/proceedings/lrec2014/pdf/525_Paper.pdf`.

[68] Costanza Navarretta and Magdalena Lis. Multimodal feedback expressions in Danish and Polish feedback expressions. In Jens Allwood, Elisabeth Ahlsén, Patrizia Paggio, Costanza Navarretta, and Kristiina Jokinen, editors, *Proceedings of the 4th Nordic Symposium on Multimodal Communication*, pages 55–62, Linköping, Sweden, 2013. NEALT, LiU Electronic Press. URL `http://www.ep.liu.se/ecp/093/008/ecp12093008.pdf`.

[69] Silvia Necşulescu. Automatic acquisition of possible contexts for low frequent words. In *Proceedings of the Student Research Workshop at the 8th International Conference on Recent*

Advances in Natural Language Processing (RANLP 2011), pages 121–126, Hissar, Bulgaria, September 2011. RANLP 2011 Organising Committee. URL `http://aclweb.org/anthology-new/R/R11/R11-2019.pdf`.

[70] Silvia Necşulescu, Núria Bel, Muntsa Padró, Montserrat Marimon, and Eva Revilla. Towards the automatic merging of language resources. In *Proceedings of the International Workshop on Lexical Resources (WoLeR 2011)*, Ljubljana, Slovenia, 2011. URL `http://alpage.inria.fr/~sagot/woler2011/WoLeR2011/Program_files/WoLeR%202011%20-%20Necsulescu%20Bel%20Padro%20Marimon%20Revilla.pdf`.

[71] Silviac Necşulescu, Sara Mendes, and Núria Bel. Combining dependency information and generalization in a pattern-based approach to the classification of lexical-semantic relation instances. In Nicoletta (conference chair) Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asunción Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 4308–4315, Reykjavik, Iceland, Mayay 2014. European Language Resources Association (ELRA). URL `http://www.lrec-conf.org/proceedings/lrec2014/pdf/837_Paper.pdf`.

[72] Muntsa Padró, Núria Bel, and Silvia Necşulescu. Towards the automatic merging of lexical resources: Automatic mapping. In *Proceedings of the 8th International Conference on Recent Advances in Natural Language Processing (RANLP 2011)*, pages 296–301, Hissar, Bulgaria, September 2011. RANLP 2011 Organising Committee. URL `http://aclweb.org/anthology-new/R/R11/R11-1041.pdf`.

[73] Muntsa Padró, Miguel Ballesteros, Héctor Martínez, and Bernd Bohnet. Finding dependency parsing limits over a large Spanish corpus. In *Proceedings of 6th International Joint Conference on Natural Language Processing*, Nagoya, Japan, October 2013. URL `http://www.taln.upf.edu/system/files/biblio_files/ijcnlp_final_padro_et_al_2013.pdf`.

[74] Muntsa Padró, Núria Bel, and Silvia Necşulescu. Towards the fully automatic merging of lexical resources: A step forward. In *LREC 2012 Workshop on Language Resource Merging; 2012 May 22, Istanbul, Turkey*, volume abs/1303.1929, pages 8–14. European Language Resources Association (ELRA), 2013. URL `http://repositori.upf.edu/bitstream/handle/10230/20417/Bel_LREC2012_4.pdf`.

[75] Carla Parra Escartín. Historia de la traducción automática. *La Linterna del Traductor, nueva época*, (6), January 2012. URL `http://www.lalinternadeltraductor.org/n6/traduccion-automatica.html`.

[76] Carla Parra Escartín. Design and compilation of a specialized Spanish-German parallel corpus. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2199–2206, Istanbul, Turkey, May 2012. European Language Resources Association (ELRA). ISBN 978-2-9517408-7-7. URL `http://www.lrec-conf.org/proceedings/lrec2012/pdf/577_Paper.pdf`.

[77] Carla Parra Escartín. Encoding a parallel corpus: The TRIS corpus experience. *Bergen Language and Linguistics Studies (BELLS)*, 3(1):61–80, 2013. URL `https://bells.uib.no/index.php/bells/article/view/362/376`.

[78] Carla Parra Escartín. Chasing the perfect splitter: A comparison of different compound splitting tools. In Nicoletta (conference chair) Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asunción Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3340–3347, Reykjavik, Iceland, May 2014. European Language Resources Association (ELRA). URL http://www.lrec-conf.org/proceedings/lrec2014/pdf/909_Paper.pdf.

[79] Carla Parra Escartín, Gyri Smørdal Losnegaard, Gunn Inger Lyse Samdal, and Pedro Patiño García. Representing Multiword Expressions in Lexical and Terminological Resources: An Analysis for Natural Language Processing Purposes. In I. Kosem, J. Kallas, P. Gantar, S. Krek, M. Langemets, and M. Tuulik, editors, *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference*, pages 338–357, Tallinn, Estonia, 17-19 October 2013. Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut (Ljubljana/Tallinn). URL http://eki.ee/elex2013/proceedings/eLex2013_23_Parra-Escartin+etal.pdf.

[80] Carla Parra Escartín, Stephan Peitz, and Hermann Ney. German compounds and statistical machine translation. can they get along? In *Proceedings of the 10th Workshop on Multiword Expressions (MWE 2014)*, Gothenburg, Sweden, 2014.

[81] Pedro Patiño. A specialized parallel corpus of English and Spanish Free Trade Agreements for the study of specialized collocations. *Synaps*, 26:85–89, 2011. URL http://www.nhh.no/Files/Filer/institutter/fsk/Synaps/Artikkel%2010%20(26).pdf.

[82] Pedro Patiño. Towards a definition of specialized collocation. In G. Quiroz and Pedro Patiño, editors, *LSP in Colombia: advances and challenges*, Linguistic Insights — Studies in Language and Communication. Peter Lang, Bern, 2013.

[83] Pedro Patiño García. FTA corpus: a parallel corpus of English and Spanish Free Trade Agreements for the study of specialized collocations. *Bergen Language and Linguistics Studies (BELLS)*, 3(1):81–92, 2013. ISSN 1892-2449. URL https://boap.uib.no/index.php/bells/article/view/363/377.

[84] Martin Popel, David Mareček, Nathan Green, and Zdeněk Žabokrtsky. Influence of parser choice on dependency-based MT. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 433–439, Edinburgh, Scotland, July 2011. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W11-2153.

[85] Gabriel Quiroz and Pedro Patiño. Foreword. In Gabriel Quiroz and Pedro Patiño, editors, *LSP in Colombia: advances and challenges*, Linguistic Insights — Studies in Language and Communication. Peter Lang, Bern, 2013. forthcoming.

[86] Loganathan Ramasamy, Ondřej Bojar, and Zdeněk Žabokrtský. Morphological processing for English-Tamil statistical machine translation. In *Proceedings of the Workshop on Machine Translation and Parsing in Indian Languages*, pages 113–122, Mumbai, India, December 2012. The COLING 2012 Organizing Committee. URL http://www.aclweb.org/anthology/W12-5611.

[87] Loganathan Ramasamy, Zdeněk Žabokrtský, and Sowmya Vajjala. The study of effect of length in morphological segmentation of agglutinative languages. In *Proceedings of the First Workshop on Multilingual Modeling*, pages 18–24, Jeju, Republic of Korea, July 2012.

Association for Computational Linguistics. URL
`http://aclweb.org/anthology//W/W12/W12-3903.pdf`.

[88] Loganathan Ramasamy and Zdeněk Žabokrtský. Tamil Dependency Treebank (TamilTB) —
Annotation manual. ÚFAL/CKL Technical Report TR-2011-42, Univerzita Karlova v Praze,
Institute of Formal and Applied Linguistics, 2011. URL
`http://ufal.mff.cuni.cz/~ramasamy/papers/TamilTB-TR.pdf`.

[89] Loganathan Ramasamy and Zdeněk Žabokrtský. Tamil dependency parsing: Results using
rule based and corpus based approaches. In Alexander Gelbukh, editor, *Computational
Linguistics and Intelligent Text Processing, 12th International Conference, (CICLing 2011),
Tokyo, Japan, February 20-26, 2011*, volume 6608 of *Lecture Notes in Computer Science*, pages
82–95, Berlin/Heidelberg, 2011. Springer-Verlag. ISBN 978-3-642-19399-6. URL
`http://dl.acm.org/citation.cfm?id=1964808`.

[90] Loganathan Ramasamy and Zdeněk Žabokrtský. Prague dependency style treebank for
tamil. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente
Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eighth
International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey,
May 2012. European Language Resources Association (ELRA). ISBN 978-2-9517408-7-7. URL
`http://www.lrec-conf.org/proceedings/lrec2012/pdf/456_Paper.pdf`.

[91] Lauren Romeo, Héctor Martínez Alonso, and Núria Bel. Class-based word sense induction
for dot-type nominals. In *Proceedings of the 6th International Conference on Generative
Approaches to the Lexicon*, Pisa, Italy, September 2013. URL
`http://glcon2013.org/wp-content/uploads/2013/09/paper_7.pdf`.

[92] Anne-Kathrin Schumann. A bilingual study of knowledge-rich context extraction in Russian
and German. In Zygmunt Vetulani, editor, *Proceedings of the 5th Language and Technology
Conference (LTC 2011)*, pages 516–520. Fundacja Uniwersytetu im. A. Mickiewicza, 2011.
URL `http://hnk.ffzg.hr/bibl/ltc2011/book/papers/LRT3-4.pdf`.

[93] Anne-Kathrin Schumann. Extraction of knowledge-rich contexts in Russian — a study in the
automotive domain. In Bolette Sandford Pedersen, Gunta Nešpore, and Inguna Skadiņa,
editors, *Proceedings of the 18th Nordic Conference of Computational Linguistics, Riga
(NoDaLiDa 2011)*, number 11 in NEALT Proceedings Series, pages 311–314, 2011. URL
`http://hdl.handle.net/10062/17361`.

[94] Anne-Kathrin Schumann. A case study of knowledge-rich context extraction in Russian. In
Kyo Kageura and Pierre Zweigenbaum, editors, *Proceedings of the 9th International
Conference on Terminology and Artificial Intelligence (TIA 2011)*, pages 143–146, Paris,
November 2011. INALCO. URL
`http://tia2011.crim.fr/Proceedings/pdf/TIA26.pdf`.

[95] Anne-Kathrin Schumann. Towards the automated enrichment of multilingual terminology
databases with knowledge-rich contexts — Experiments with Russian EuroTermBank data. In
Tatiana Gornostay, editor, *Proceedings of the 2nd Workshop on the Creation, Harmonization
and Application of Terminology Resources (CHAT 2012)*, volume 72 of *Linköping Electronic
Conference Proceedings*, pages 27–34, 2012. URL
`http://www.ep.liu.se/ecp/072/004/ecp12072004.pdf`.

[96] Anne-Kathrin Schumann. Towards the automated enrichment of multilingual terminology
databases with knowledge-rich contexts. In Alexander E. Kibrik, editor, *Proceedings of*

Dialogue 2012, pages 559–567, Moscow, Russia, 2012. Российский государственный гуманитарный университет (Russian State University of Humanities). URL `http://dialog-21.ru/digests/dialog2012/materials/pdf/152.pdf`.

[97] Anne-Kathrin Schumann. Knowledge-rich context extraction and ranking with KnowPipe. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 3626–3630, Istanbul, Turkey, May 2012. European Language Resources Association (ELRA). URL `http://www.lrec-conf.org/proceedings/lrec2012/pdf/678_Paper.pdf`.

[98] Sebastian Sulger, Miriam Butt, Tracy Holloway King, Paul Meurer, Tibor Laczkó, György Rákosi, Cheikh Bamba Dione, Helge Dyvik, Victoria Rosén, Koenraad De Smedt, Agnieszka Patejuk, Özlem Çetinoglu, I Wayan Arka, and Meladel Mistica. ParGramBank: The ParGram parallel treebank. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 550–560, Sofia, Bulgaria, August 2013. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/P13-1054`.

[99] Raymond Hendy Susanto, Septina Dian Larasati, and Francis M. Tyers. Rule-based machine translation between Indonesian and Malaysian. In *Proceedings of the Workshop on South and Southeast Asian Natural Language Processing at Coling 2012*, Mumbai, India, December 2012. Coling 2012 Organizing Committee.

[100] Anders Søgaard, Héctor Martínez Alonso, Jakob Elming, and Anders Johannsen. Using crowdsourcing to get representations based on regular expressions. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1476–1480, Seattle, Washington, USA, October 2013. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/D13-1154`.

[101] Larraitz Uria, Mans Hulden, Izaskun Etxeberria, and Iñaki Alegria. Recursos y métodos de sustitución léxica en las variantes dialectales en euskera. In *Proceedings of the Workshop on Iberian Cross-Language NLP tasks (ICL-2011)*, pages 70–76. CEUR-WS, September 2011. URL `http://ceur-ws.org/Vol-824/paper10.pdf`.

[102] Sowmya Vajjala and Kaidi Lõo. Role of morpho-syntactic features in estonian proficiency classification. In *Proceedings of the 8th workshop on Innovative Use of NLP for Building Educational Applications (BEA8)*, pages 63–72. Association for Computational Linguistics, 2013. URL `http://aclweb.org/anthology-new/W/W13/W13-1708.pdf`.

[103] Sowmya Vajjala and Detmar Meurers. On improving the accuracy of readability classification using insights from second language acquisition. In *Proceedings of the 7th Workshop on Innovative Use of NLP for Building Educational Applications (BEA7)*, pages 163–173, Montréal, Canada, June 2012. Association for Computational Linguistics. URL `http://aclweb.org/anthology-new/W/W12/W12-2019.pdf`.

[104] Sowmya Vajjala and Detmar Meurers. On the applicability of readability models to web texts. In *Proceedings of the Second Workshop on Predicting and Improving Text Readability for Target Reader Populations*, pages 59–68, Sofia, Bulgaria, August 2013. Association for Computational Linguistics. URL `http://aclweb.org/anthology/W13-2907.pdf`.

[105] Sowmya Vajjala and Detmar Meurers. On assessing the reading level of individual sentences for text simplification. In *Proceedings of the 14th Conference of the European Chapter of the*

Association for Computational Linguistics, Gothenburg, Sweden, 2014. Association for Computational Linguistics. URL `http://purl.org/dm/papers/Vajjala.Meurers-14-eacl.html`.

[106] Sowmya Vajjala and Detmar Meurers. Readability assessment for text simplification: From analyzing documents to identifying sentential simplifications. *International Journal of Applied Linguistics, Special Issue on Current Research in Readability and Text Simplification*, 2014. URL `http://purl.org/dm/papers/Vajjala.Meurers-14-ijal.html`.

[107] Sowmya Vajjala and Detmar Meurers. Exploring measures of "readability" for spoken language: Analyzing linguistic features of subtitles to identify age-specific tv programs. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR) at EACL-14*, Gothenburg, Sweden, 2014. Association for Computational Linguistics. URL `http://purl.org/dm/papers/Vajjala.Meurers-14-pitr.html`.

[108] Peter Wittenburg, Przemyslaw Lenkiewicz, Erik Auer, Anna Lenkiewicz, Binyam Gebrekidan Gebre, and Sebastian Drude. AV processing in eHumanities — A paradigm shift. In J. C. Meiser, editor, *Proceedings of the Digital Humanities Conference 2012*, pages 538–541, Hamburg, July 2012. URL `http://hdl.handle.net/11858/00-001M-0000-000E-78C0-6`.

[109] Marcos Zampieri and Binyam Gebrekidan Gebre. Automatic identification of language varieties: The case of Portuguese. In *Proceedings of KONVENS2012*, pages 233–237, Vienna, Austria, 2012.

[110] Marcos Zampieri, Binyam Gebrekidan Gebre, and Sascha Diwersy. Classifying pluricentric languages: Extending the monolingual model. In *Proceedings of the Fourth Swedish Language Technology Conference (SLTC2012)*, pages 79–80, Lund, Sweden, 2012.

[111] Marcos Zampieri, Binyam Gebrekidan Gebre, and Sascha Diwersy. N-gram language models and pos distribution for the identification of spanish varieties. In *Proceedings of TALN2013*, pages 580–587, 2013.

[112] Daniel Zeman, David Mareček, Martin Popel, Loganathan Ramasamy, Jan Štěpánek, Zdeněk Žabokrtský, and Jan Hajič. HamleDT: To parse or not to parse? In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, May 2012. European Language Resources Association (ELRA). ISBN 978-2-9517408-7-7. URL `http://www.lrec-conf.org/proceedings/lrec2012/pdf/429_Paper.pdf`.