

Grzegorz Chrupała  
Associate Professor  
Cognitive Science & AI  
Cognitive Science and AI  
**Type of address: Visiting address.**  
Warandelaan 2  
Dante Building, room D 127  
5037 AB  
Tilburg  
Netherlands  
**Email:** G.A.Chrupala@tilburguniversity.edu



## Research output

Sarti, G., Chrupała, G., Nissim, M., & Bisazza, A. (2024). *Quantifying the Plausibility of Context Reliance in Neural Machine Translation*. <https://doi.org/10.48550/arXiv.2310.01188>

Shen, G., Watkins, M., Alishahi, A., Bisazza, A., & Chrupała, G. (2024). *Encoding of lexical tone in self-supervised models of spoken language*. 4250-4261. Paper presented at Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers). <https://doi.org/10.18653/v1/2024.naacl-long.239>

Mohebbi, H., Chrupała, G., Zuidema, W., & Alishahi, A. (2023). Homophone Disambiguation Reveals Patterns of Context Mixing in Speech Transformers. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 8249–8260). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.emnlp-main.513>

Chrupała, G. (2023). Putting natural in natural language processing. In *Findings of the Association for Computational Linguistics: ACL 2023* (pp. 7820-7827). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.findings-acl.495>

Mohebbi, H., Zuidema, W., Chrupała, G., & Alishahi, A. (2023). Quantifying Context Mixing in Transformers. In *In Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 3378-3400). Association for Computational Linguistics. <https://aclanthology.org/2023.eacl-main.245/>

Shen, G., Alishahi, A., Bisazza, A., & Chrupała, G. (2023). Wave to Syntax: Probing spoken language models for syntax. In *Proc. INTERSPEECH 2023* (pp. 1259-1263). (Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH). <https://doi.org/10.21437/Interspeech.2023-679>

Afsharizadeh, M., Ebrahimpour-Komleh, H., Bagheri, A., & Chrupala, G. (2022). A Survey on Multi-document Summarization and Domain-Oriented Approaches. *Journal of Information Systems and Telecommunication*, 10(37), 68-79. <https://doi.org/10.52547/jist.16245.10.37.68>

Nikolaus, M., Alishahi, A., & Chrupała, G. (2022). Learning English with Peppa Pig. *Transactions of the Association for Computational Linguistics*, 10, 922-936. [https://doi.org/10.1162/tacl\\_a\\_00498](https://doi.org/10.1162/tacl_a_00498)

Emmery, C., Kádár, A., Chrupała, G., & Daelemans, W. M. P. (2022). *Cyberbullying Classifiers are Sensitive to Model-Agnostic Perturbations*. Paper presented at Language Resources and Evaluation Conference, Marseille, France. <https://arxiv.org/abs/2201.06384>

Chrupala, G. (2022). Visually grounded models of spoken language - A survey of datasets, architectures and evaluation techniques. *Journal of Artificial Intelligence Research*, 73, 673-707. <https://doi.org/10.1613/jair.1.12967>

Higy, B., Gelderloos, L., Alishahi, A., & Chrupała, G. (2021). Discrete representations in neural models of spoken language. In *Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP* (pp. 163-176). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.blackboxnlp-1.11>

- Emmery, C., Kádár, Á., & Chrupala, G. (2021). Adversarial Stylometry in the Wild: Transferable Lexical Substitution Attacks on Author Profiling. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers* (pp. 2388-2402). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.eacl-main.203>
- Alishahi, A., Chrupala, G., Cristià, A., Dupoux, E., Higy, B., Lavechin, M., Räsänen, O., & Yu, C. (2021). ZR-2021VG: Zero-Resource Speech Challenge, Visually-Grounded Language Modelling track, 2021 edition. *CoRR*, *abs/2107.06546*. <https://arxiv.org/abs/2107.06546>
- Alishahi, A., Belinkov, Y., Chrupala, G., Hupkes, D., Pinter, Y., & Sajjad, H. (Eds.) (2020). *Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*. Association for Computational Linguistics. <https://aclanthology.org/2020.blackboxnlp-1.0>
- Higy, B., Elliott, D., & Chrupala, G. (2020). Textual supervision for visually grounded spoken language understanding. In *Findings of the Association for Computational Linguistics: EMNLP 2020* (pp. 2698-2709). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.findings-emnlp.244>
- Chrupala, G., Higy, B., & Alishahi, A. (2020). Analyzing analytical methods: The case of phonology in neural models of spoken language. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 4146-4156). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.381>
- Gelderloos, L., Chrupala, G., & Alishahi, A. (2020). Learning to Understand Child-directed and Adult-directed Speech. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 1-6). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.1>
- Linzen, T., Chrupala, G., Belinkov, Y., & Hupkes, D. (2019). *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*. Association for Computational Linguistics. <https://www.aclweb.org/anthology/W19-4800>
- Alishahi, A., Chrupala, G., & Linzen, T. (2019). Analyzing and Interpreting Neural Networks for NLP: A Report on the First BlackboxNLP Workshop. *Natural Language Engineering*, 25(4), 543-557. Article 135132491900024. <https://doi.org/10.1017/S135132491900024X>
- Chrupala, G., & Alishahi, A. (2019). Correlating Neural and Symbolic Representations of Language. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 2952-2962). Association for Computational Linguistics. <https://www.aclweb.org/anthology/P19-1283>
- Chrupala, G. (2019). Symbolic Inductive Bias for Visually Grounded Learning of Spoken Language. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 6452-6462). Association for Computational Linguistics. <https://www.aclweb.org/anthology/P19-1647>
- Kádár, Á., Chrupala, G., Alishahi, A., & Elliott, D. (2019). Bootstrapping Disjoint Datasets for Multilingual Multimodal Representation Learning. Unpublished. <https://arxiv.org/abs/1911.03678>
- Chrupala, G., Gelderloos, L., Kádár, Á., & Alishahi, A. (2019). On the difficulty of a distributional semantics of spoken language. In *Proceedings of the Society for Computation in Linguistics* (Vol. 2) <https://doi.org/10.7275/extq-7546>
- Linzen, T., Chrupala, G., & Alishahi, A. (2018). *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*. Association for Computational Linguistics. <https://www.aclweb.org/anthology/W18-5400>
- Emmery, C., Manjavacas, E., & Chrupala, G. (2018). Style Obfuscation by Invariance. In *COLING 2018 Association for Computational Linguistics*. <https://arxiv.org/abs/1805.07143>

- Emmery, C., Manjavacas, E., & Chrupala, G. (2018). *Attribute Obfuscation with Gradient Reversal*. Abstract from Computational Linguistics in the Netherlands, Nijmegen, Netherlands. <http://aclweb.org/anthology/C18-1084>
- Kádár, Á., Elliott, D., Côte, M.-A., Chrupala, G., & Alishahi, A. (2018). Lessons Learned in Multilingual Grounded Language Learning. In *Proceedings of the 22nd Conference on Computational Natural Language Learning* (pp. 402-412). Association for Computational Linguistics. <http://aclweb.org/anthology/K18-1039>
- Kádár, Á., Côte, M.-A., Chrupala, G., & Alishahi, A. (2018). Revisiting the Hierarchical Multiscale LSTM. In *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 3215-3227). Association for Computational Linguistics. <http://aclweb.org/anthology/C18-1272>
- Chrupala, G., Alishahi, A., & Gelderloos, L. (2017). *Emergence of language structures from exposure to visually grounded speech signal*. Abstract from Computational Linguistics in the Netherlands 27, Leuven, Belgium.
- Alishahi, A., Barking, M., & Chrupala, G. (2017). Encoding of phonology in a recurrent neural model of grounded speech. In R. Levy, & L. Specia (Eds.), *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)* (pp. 368-378). Association for Computational Linguistics. <https://doi.org/10.18653/v1/K17-1037>
- Gelderloos, L., Alishahi, A., Chrupala, G., & Fernández, R. (2017). *Modeling relations in a referential game*. Abstract from Computational Linguistics in the Netherlands, Nijmegen, Netherlands. <http://clin28.cls.ru.nl/#abstract-69>
- Kadar, A., Chrupala, G., & Alishahi, A. (2017). Representation of linguistic form and function in recurrent neural networks. *Computational Linguistics*, 43(4), 761-780. [https://doi.org/10.1162/COLI\\_a\\_00300](https://doi.org/10.1162/COLI_a_00300)
- Chrupala, G., Gelderloos, L., & Alishahi, A. (2017). Representations of language in a model of visually grounded speech signal. In *Proceedings of the 55th of the Annual Meeting of the Association for Computational Linguistics* (pp. 613-622 ). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P17-1057>
- Emmery, C., Chrupala, G., & Daelemans, W. M. P. (2017). Simple Queries as Distant Labels for Detecting Gender on Twitter. In *Proceedings of the 3rd Workshop on Noisy User-generated Text* (pp. 50-55). Association for Computational Linguistics. <http://noisy-text.github.io/2017/pdf/WNUT07.pdf>
- Lazaridou, A., Chrupala, G., Fernández, R., & Baroni, M. (2016). Multimodal Semantic Learning from Child-Directed Input . In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 387-392). Association for Computational Linguistics. <http://clac.cimec.unitn.it/marco/publications/lazaridou-et-al-multimodal-learning-from-cdi-naacl2016.pdf>
- Trovato, G., Chrupala, G., & Takanishi, A. (2016). Application of the Naive Bayes Classifier for Representation and Use of Heterogeneous and Incomplete Knowledge in Social Robotics. *Robotics*, 5(1). <https://doi.org/10.3390/robotics5010006>
- Gelderloos, L. J., & Chrupala, G. (2016). From phonemes to images: levels of representation in a recurrent neural model of visually-grounded language learning. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers* (pp. 1309-1319). International Committee on Computational Linguistics. <http://aclweb.org/anthology/C/C16/C16-1124.pdf>
- Chrupala, G. (2015). *Architectures and representations for string transduction*. Abstract from The 25th Meeting of Computational Linguistics in the Netherlands (CLIN25), Belgium. <http://www.clips.uantwerpen.be/clin25/abstracts>
- Chrupala, G., Kadar, A., & Alishahi, A. (2015). Learning language through pictures. In C. Zong, & M. Strube (Eds.), *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)* (pp. 112-118). Association for Computational Linguistics. <http://www.aclweb.org/anthology/P/P15/P15-2019.pdf>

Kadar, A., Alishahi, A., & Chrupala, G. (2015). Learning word meanings from images of natural scenes. *Traitement Automatique des Langues*.

Kadar, A., Chrupala, G., & Alishahi, A. (2015). *Linguistic Analysis of Multi-Modal Recurrent Neural Networks*. Abstract from Workshop on Vision and Language 2015 (VL'15): Vision and Language Integration Meets Cognitive Systems, Lisbon, Portugal. <http://www.emnlp2015.org/proceedings/VL/pdf/VL04.pdf>

Baltadzhieva, A., & Chrupala, G. (2015). Predicting the quality of questions on Stackoverflow. In G. Angelova, K. Bontcheva, & R. Mitkov (Eds.), *Proceedings of Recent Advances in Natural Language Processing* (pp. 32-40) <file:///home/gchrupala/repos/grzegorz.chrupala.me/papers/ranlp-2015.pdf>

Baltadzhieva, A., & Chrupala, G. (2015). Question Quality in Community Question Answering Forums: A survey. *ACM SIGKDD Explorations Newsletter*, 17(1), 8-13.

Kadar, A., Chrupala, G., & Alishahi, A. (2015). *Towards learning domain-general representations for language from multi-modal data*. Abstract from The 26th Meeting of Computational Linguistics in the Netherlands (CLIN26), Amsterdam, Netherlands.

Barman, U., Wagner, J., Chrupala, G., & Foster, J. (2014). DCU-UVT. Word-Level Language Classification with Code-Mixed Data. In *First Workshop on Computational Approaches to Code Switching* (pp. 127-132). Association for Computational Linguistics (ACL). <http://www.aclweb.org/anthology/W/W14/W14-39.pdf#page=135>

Alishahi, A., & Chrupala, G. (2014). *Induced lexical categories enhance cross-situational learning of word meanings*. Abstract from 24th Meeting of Computational Linguistics in The Netherlands (CLIN 2014), Leiden, Netherlands.

Chrupala, G. (2014). *Language-agnostic processing of microblog data with text embeddings*. Poster session presented at 24th Meeting of Computational Linguistics in The Netherlands (CLIN 2014), Leiden, Netherlands.

Chrupala, G. (2014). *Learning character-wise text representations with Elman nets*. Abstract from 23rd Annual Belgian-Dutch Conference on Machine Learning (BENELEARN 2014), Brussels, Belgium.

Chrupala, G. (2014). *Learning to normalize text from few examples*. Abstract from ATILA 2014, Ghent, Belgium. <http://grzegorz.chrupala.me/papers/atila-2014.pdf>

Chrupala, G. (2014). Normalizing tweets with edit scripts and recurrent neural embeddings. In K. Toutanova, & H. Wu (Eds.), *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (Short Papers ed., Vol. 2, pp. 680-686). Association for Computational Linguistics (ACL).

Roth, B., Barth, T., Chrupala, G., Gropp, M., & Klakow, D. (2014). RelationFactory: A Fast, Modular and Effective System for Knowledge Base Population. In S. Wintner, S. Goldwater, & S. Riezler (Eds.), *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 89-92). Association for Computational Linguistics (ACL).

Roth, B., Barth, T., Chrupala, G., Gropp, M., & Klakow, D. (2014). RelationFactory: End-to-end relation extraction and knowledge base population pipeline. Software <https://github.com/beroth/relationfactory>

Deng, H., & Chrupala, G. (2014). Semantic approaches to software component retrieval with English queries. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Conference on International Language Resources and Evaluation (LREC)* (pp. 3248-3252). European Language Resources Association (ELRA). [http://www.lrec-conf.org/proceedings/lrec2014/pdf/106\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/106_Paper.pdf)

- Chrupala, G. (2013). *Text segmentation with character-level text embeddings*. Paper presented at Workshop on Deep Learning for Audio, Speech and Language Processing, ICML 2013, Atlanta, United States.
- Evang, K., Basile, V., Chrupala, G., & Bos, J. (2013). Elephant: Sequence labeling for word and sentence segmentation. Software <http://gmb.let.rug.nl/elephant/about.php>
- Evang, K., Basile, V., Chrupala, G., & Bos, J. (2013). Elephant: Sequence Labeling for Word and Sentence Segmentation . In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 1422-1426). Association for Computational Linguistics (ACL). <http://www.aclweb.org/anthology/D13-1146>
- Chrupala, G. A. (2013). Learning text representations from character-level data. In M. Theune, & A. Nijholt (Eds.), *Book of abstracts of the 23rd meeting of computational linguistics in the Netherlands (CLIN)* (pp. 18). Universiteit Twente.
- Alishahi, A., & Chrupala, G. A. (2012). Concurrent acquisition of word meaning and lexical categories. In *Proceedings of the Conference on Empirical Methods on Natural Language Processing and Computational Natural Language Learning* (pp. 643-654). Association for Computational Linguistics. <http://www.informatik.uni-trier.de/~ley/pers/hd/a/Alishahi:Afra.html>
- Roth, B., Chrupala, G., Wiegand, M., & Singh, M. (2012). Generalizing from Freebase and Patterns using Distant Supervision for Slot Filling. In *Text Analysis Conference*
- Chrupala, G. A. (2012). Hierarchical clustering of word class distributions. In T. Cohn, P. Blunsom, & J. Graca (Eds.), *Proceedings of the NAACL-HLT Workshop on the Induction of Linguistic Structure, Montreal* (pp. 100-104). Association for Computational Linguistics.
- Chrupala, G. A. (2012). Learning from evolving data streams: Online triage of bug reports. In W. Daelemans (Ed.), *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 613-622). Association for Computational Linguistics.
- Chrupala, G. (2011). Efficient induction of probabilistic word classes with LDA. In *International Joint Conference on Natural Language Processing*
- Xu, F., Kazalski, S., Chrupala, G., Roth, B., Zhao, X., Wiegand, M., & Klakow, D. (2011). Saarland University Spoken Language Systems Group at TAC KBP 2011. In *Text Analysis Conference*
- Chrupala, G., & Klakow, D. (2010). A Named Entity Labeler for German: exploiting Wikipedia and distributional clusters. In *Proceedings of the Conference on International Language Resources and Evaluation (LREC)* (pp. 552-556)
- Chrupala, G., Dinu, G., & Roth, B. (2010). Enriched syntax-based meaning representation for answer extraction. In *SIGIR Workshop: Query Representation and Understanding*
- Chrupala, G. A. (2010). Lemmatization and lexicalized statistical parsing of morphologically rich languages: the case of French. In *Statistical Parsing of Morphologically Rich Languages Workshop* Association for Computational Linguistics.
- Chrupala, G. (2010). Morfette: A tool for supervised learning of morphology. Software <https://code.google.com/p/morfette/>
- Chrupala, G., & Alishahi, A. (2010). Online entropy-based model of lexical category acquisition. In *Fourteenth Conference on Computational Natural Language Learning* (pp. 182-191)
- Dinu, G., & Chrupala, G. (2010). Relatedness curves for acquiring paraphrases. In *Proceedings of the 2010 Workshop on GEometrical Models of Natural Language Semantics* (pp. 27-32)

- Chrupala, G., Momtazi, S., Wiegand, M., Kazalski, S., Xu, F., Roth, B., Balahur, A., & Klakow, D. (2010). Saarland University Spoken Language Systems at the Slot Filling Task of TAC KBP 2010. In *Text Analysis Conference*
- Chrupala, G. (2010). Sequor. Software <https://bitbucket.org/gchrupala/sequor>
- Alishahi, A., & Chrupala, G. (2009). Lexical category acquisition as an incremental process. In *Proceedings of the CogSci 2009 Workshop on Psycho Computational Models of Human Language Acquisition*
- Chrupala, G. A., Dinu, G., & van Genabith, J. (2008). Learning morphology with Morfette. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*
- Wiegand, M., Momtazi, S., Kazalski, S., Xu, F., Chrupala, G., & Klakow, D. (2008). The Alyssa System at TAC QA 2008. In *Text Analysis Conference*
- Chrupala, G. (2008). *Towards a machine-learning architecture for lexical functional grammar parsing*. Dublin City University. School of Computing.
- Chrupala, G. A. (2007). Better training for function labeling. In *Proceedings of Recent Advances in Natural Language Processing*
- Chrupala, G. A. (2007). Using very large corpora to detect raising and control verbs. In *Proceedings of LFG07 CSLI Publications*.
- Chrupala, G. A. (2006). Improving treebank-based automatic LFG induction for Spanish. In *Proceedings of the LFG '06 Conference CSLI Publications*.
- Chrupala, G. A. (2006). Simple data-driven context-sensitive lemmatization. *Procesamiento del Lenguaje natural, Revista*, (37).
- Chrupala, G. A. (2006). Using machine-learning to assign function labels to parser output for Spanish. In *Proceedings of the COLING/ACL on Main Conference Poster Sessions Association for Computational Linguistics*.
- Pym, A., & Chrupala, G. (2005). The quantitative analysis of translation flows in the age of an international language. In A. Branchadell, & L. M. West (Eds.), *Less Translated Languages* (pp. 27-38). John Benjamins Publishing Company.
- Carreras, X., Màrquez, L., & Chrupala, G. (2004). Hierarchical recognition of propositional arguments with perceptrons. In *Proceedings of CoNLL'2004 Shared Task*
- Chrupala, G. (2003). Acquiring verb subcategorization from spanish corpora.
- Chrupala, G. (2003). Perl Scripting in Translation Project Management. *Across Languages and Cultures*, 4(1), 109-132.
- Chrupala, G. (1998). Biblioteca in fabula, the library motive in La biblioteca de Babel, The british museum is falling down and Il nome della rosa.