

Jörg Tiedemann
Professor, Supervisor for doctoral programme
Department of Digital Humanities
Language Technology
Doctoral Programme in Language Studies
Postal address:
PL 24 (Unioninkatu 40)
00014
Finland
Email: jorg.tiedemann@helsinki.fi
Phone: +358294123261
Mobile: +358503119440



Qualifications

Computational Linguistics, PhD, Uppsala University
2000 → 2003

Computer Science, M.Sc., Otto Von Guericke University, Magdeburg
1991 → 1997

Principal Investigator

Period : 27.01.2016 - 31.12.2017 in Department of Modern Languages 2010-2017

Principal Investigator

Period : 17.02.2018 - * in Department of Digital Humanities

Employment

Professor

Department of Digital Humanities
University of Helsinki
Finland
1 Jan 2018 → present

Language Technology

University of Helsinki
University of Helsinki, Finland
1 Aug 2015 → present

Supervisor for doctoral programme

Doctoral Programme in Language Studies
University of Helsinki
1 Jan 2014 → present

Projects

1. **FoTran: Found in Translation - Natural Language Understanding with Cross-Lingual Grounding**
Tiedemann, J., Celikkanat, H., Raganato, A., Vazquez, R., Sulubacak, U. & Silfverberg, M.
European Research Council (ERC)
01/09/2018 → 31/08/2023
2. **fiskmö: Creation of a parallel corpus of translated documents and machine translation for Finnish and Swedish**
Tiedemann, J., Ginter, F., Papula, N., Aulamo, M., Nieminen, T. & Kanerva, J.
Svenska kulturfonden
01/05/2018 → 31/05/2020
3. **MeMAD: Methods for Managing Audiovisual Data: Combining Automatic Efficiency with Human Accuracy**
Hirvonen, M., Tiedemann, J., Tiittula, L., Sulubacak, U. & Vazquez, R.
European Commission / Horizon 2020
01/01/2018 → 31/12/2020

4. **NLUxLG: NLU with Cross-Lingual Grounding**
Tiedemann, J., Talman, A., Raganato, A. & Celikkanat, H.
SUOMEN AKATEMIA
01/01/2018 → 31/12/2019
5. **Development of Digital Teaching project of the BA Programme in Languages**
Lennes, M., Flanagan, J., Hiltunen, T., Creutz, M., Aunio, L., Buchart, M., Gruzdeva, E., Jakubek, I., Keinänen, N., Kim, J., Kittilä, S., Kopotev, M., Kraenker, S., Länsisalmi, R., Nieuweboer, R., Peterson, E., Tiedemann, J., Tuori, R., Viitamäki, M. & Lindstedt, J.
01/05/2017 → 31/12/2018
6. **NLPL: Nordic Language Processing Laboratory**
Tiedemann, J. & Scherrer, Y.
01/01/2017 → 31/12/2019
7. **CrossNLP: Cross-Linguistic and Multilingual Natural Language Processing with the Focus on Low-Resource Languages and Language Variants**
Tiedemann, J., Östling, R. & Scherrer, Y.
01/01/2016 → 31/12/2018
8. **Erillisrahoitus Helsingin ja Tukholman yliopistojen yhteishanke**
Lehti-Eklund, H., Lindström, J., Miestamo, M., Grünthal, R., Palander-Collin, M., Kalliokoski, J., Östman, J., Ajanki, R., Holopainen, S., Balode, L., Lenk, H. E. H., Kurhila, S., Kotilainen, L., Lindholm, C., Wahlström, M., Solin, A., Fremer, M., Tiittula, L., Hamari, A., Kittilä, S., Huhtamäki, M., Kim, J., Kopotev, M., Peterson, E., Protassova, E., Savijärvi, M. & Tiedemann, J.
Unknown funder
01/01/2015 → 31/12/2017
9. **eSSENCE: Efficient Algorithms for Natural Language Processing Beyond Sentence Boundaries**
Nivre, J., Tiedemann, J. & Stymne, S.
01/09/2012 → 31/08/2014
10. **DiscoMT: Discourse-Oriented Statistical Machine Translation**
Tiedemann, J., Hardmeier, C., Loáiciga, S. & Scherrer, Y.
01/01/2012 → 31/12/2017
11. **LetsMT!: LetsMT! A Platform for Online Sharing of Training Data and Building User Tailored Machine Translation**
Tiedemann, J., Weinitz, P., Saers, M., Zumpe, M. & Schleussner, S.
01/03/2010 → 31/08/2012
12. **OPUS: The Open Parallel Corpus**
Tiedemann, J.
01/06/2004 → ...

Datasets

1. **LetsMT repository**
Tiedemann, J. (Creator), Zenodo, 22 Jan 2018
DOI: 10.5281/zenodo.1157836
Dataset
2. **Uplug**
Tiedemann, J. (Creator), Zenodo, 22 Jan 2018
DOI: 10.5281/zenodo.1157180, <https://github.com/Helsinki-NLP/Uplug/tree/v0.3.9>
Dataset
3. **Lingua-Align**
Tiedemann, J. (Creator), Zenodo, 22 Jan 2018
DOI: 10.5281/zenodo.1157176, <https://github.com/Helsinki-NLP/Lingua-Align/tree/v0.1>
Dataset
4. **OPUS**
Tiedemann, J. (Creator), University of Helsinki, 2017
<http://opus.nlpl.eu>
Dataset
5. **WMT17 test set Finnish-English**
Tiedemann, J. (Creator), Zenodo, 21 Jan 2018
DOI: 10.5281/zenodo.1156326, <https://github.com/Helsinki-NLP/WMT16-test-enfi/tree/v1.0>
Dataset
6. **WMT16 test set Finnish-English**
Tiedemann, J. (Creator), Zenodo, 21 Jan 2018
DOI: 10.5281/zenodo.1156325, <https://github.com/Helsinki-NLP/WMT16-test-enfi/tree/v1.0>
Dataset

7. **Shared Task in Cross-Lingual Parsing**
Tiedemann, J. (Data Manager), Atlassian, 2017
<https://bitbucket.org/hy-crossNLP/wardial2017>
Dataset
8. **DiscoMT 2015 Shared Task on Pronoun Translation**
Tiedemann, J. (Creator), LINDAT/CLARIN, 31 Jan 2016
DOI: <http://hdl.handle.net/11372/LRT-1611>, <http://hdl.handle.net/11372/LRT-1611>
Dataset

Publications

1. Talman, A., Sulubacak, U., Vazquez, R., Scherrer, Y., Virpioja, S., Raganato, A., ... Tiedemann, J. (Accepted/In press). The University of Helsinki submissions to the WMT19 news translation task. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*.
2. Tiedemann, J., & Scherrer, Y. (2019). Measuring Semantic Abstraction of Multilingual NMT with Paraphrase Recognition and Generation Tasks. In A. Rogers, A. Drozd, A. Rumshisky, & Y. Goldberg (Eds.), *Proceedings of the 3rd Workshop on Evaluating Vector Space Representations for NLP* (pp. 35-42). Stroudsburg: Association for Computational Linguistics.
3. Talman, A. J., Yli-Jyrä, A., & Tiedemann, J. (Accepted/In press). Sentence Embeddings in NLI with Iterative Refinement Encoders. *Natural Language Engineering*.
4. Härmäläinen, M., Säily, T., Rueter, J., Tiedemann, J., & Mäkelä, E. (2019). Revisiting NMT for normalization of early English letters. In B. Alex, S. Degaetano-Ortlieb, A. Kazantseva, N. Reiter, & S. Szpakowicz (Eds.), *Proceedings of the 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature* (pp. 71-75). (ACL Anthology; No. W19-25). Stroudsburg: Association for Computational Linguistics.
5. Bjerva, J., Östling, R. M., Han Veiga, M., Tiedemann, J., & Augenstein, I. (Accepted/In press). What do Language Representations Really Represent? *Computational Linguistics*.
6. Stig-Arne, G., Huet, B., Kurimo, M., Laaksonen, J., Merialdo, B., Pham, P., ... Vázquez Carrillo, J. R. (2018). The MeMAD Submission to the WMT18 Multimodal Translation Task. In O. Bojar, R. Chatterjee, C. Federmann, M. Fishel, Y. Graham, B. Haddow, M. Huck, A. J. Yepes, P. Koehn, C. Monz, M. Negri, A. Nèveol, M. Neves, M. Post, L. Specia, M. Turchi, ... K. Verspoor (Eds.), *Proceedings of the Third Conference on Machine Translation (WMT): Shared Task Papers* (pp. 603-611). Stroudsburg: Association for Computational Linguistics.
7. Öhman, E. S., Tiedemann, J., Honkela, T. U., & Kajava, K. (2018). Creating a Dataset for Multilingual Fine-grained Emotion-detection Using Gamification-based Annotation. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 24-30). Stroudsburg: Association for Computational Linguistics.
8. Sulubacak, U., Tiedemann, J., Rouhe, A., Stig-Arne, G., & Kurimo, M. (2018). The MeMAD Submission to the IWSLT 2018 Speech Translation Task. In M. Turchi, J. Niehues, & M. Federico (Eds.), *Proceedings of the 15th International Workshop on Spoken Language Translation (IWSLT 2018)* (pp. 89-94). Bruges.
9. Vazquez, R., Sulubacak, U., & Tiedemann, J. (2018). The University of Helsinki Submission to the WMT19 Parallel Corpus Filtering Task. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*.
10. Ehrentraut, C., Ekholm, M., Tanushi, H., Tiedemann, J., & Dalanis, H. (2018). Detecting hospital-acquired infections: A document classification approach using support vector machines and gradient tree boosting. *Health informatics journal.*, 24(1), 24-42. <https://doi.org/10.1177/1460458216656471>
11. Raganato, A., & Tiedemann, J. (2018). An Analysis of Encoder Representations in Transformer-Based Machine Translation. In L. Tal, G. Chrupala, & A. Alishahi (Eds.), *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP* (pp. 287-297). Stroudsburg: Association for Computational Linguistics.
12. Tiedemann, J. (2018). Emerging Language Spaces Learned From Massively Multilingual Corpora. In E. Mäkelä, M. Tolonen, & J. Tuominen (Eds.), *Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference (DHN 2018)* (Vol. 2084, pp. 188-197). (CEUR Workshop Proceedings). Helsinki: CEUR Workshop Proceedings.
13. Zampieri, M., Malmasi, S., Nakov, P., Ali, A., Shon, S., Glass, J., ... Jain, M. (2018). Language Identification and Morphosyntactic Tagging: The Second VarDial Evaluation Campaign. In M. Zampieri, P. Nakov, N. Ljubešić, J. Tiedemann, S. Malmasi, & A. Ali (Eds.), *Proceedings of the Fifth Workshop on NLP for Similar Languages, Varieties and Dialects* (pp. 1-17). Santa Fe: Association for Computational Linguistics.
14. Härmäläinen, M., Säily, T., Rueter, J., Tiedemann, J., & Mäkelä, E. (2018). Normalizing early English letters to Present-day English spelling. In B. Alex, S. Degaetano-Ortlieb, A. Feldman, A. Kazantseva, N. Reiter, & S. Szpakowicz (Eds.), *Proceedings of the 2nd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature* (pp. 87-96). (ACL Anthology; No. W18-45). Stroudsburg, PA: Association for Computational Linguistics.

15. Lison, P., Tiedemann, J., & Kouylekov, M. (2018). OpenSubtitles2018: Statistical Rescoring of Sentence Alignments in Large, Noisy Parallel Corpora. In N. Calzolari, C. Khalid, C. Christopher, D. Thierry, G. Sara, H. Koiti, I. Hitoshi, M. Bente, M. Joseph, M. Hélène, M. Asuncion, O. Jan, P. Stelios, ... T. Takenobu (Eds.), *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)* (pp. 1742-1748). Paris: European Language Resources Association (ELRA).
16. Bozovic, P., Erjavec, T., Tiedemann, J., Ljubesic, N., & Gorjanc, V. (2018). Opus-MontenegrinSubs 1.0: First electronic corpus of the Montenegrin language. In D. Fišer, & A. Pančur (Eds.), *Proceedings of the conference on Language Technologies & Digital Humanities 2018* (pp. 24-28). Ljubljana: Ljubljana University Press.
17. Zampieri, M., Nakov, P., Ljubesic, N., Tiedemann, J., Malmasi, S., & Ali, A. (Eds.) (2018). *Proceedings of the Fifth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2018)*. Stroudsburg: Association for Computational Linguistics.
18. Raganato, A., Scherrer, Y., Nieminen, T., Hurskainen, A., & Tiedemann, J. (2018). The University of Helsinki submissions to the WMT18 news task. In O. Bojar, R. Chatterjee, C. Federmann, M. Fishel, Y. Graham, B. Haddow, M. Huck, A. J. Yepes, P. Koehn, C. Monz, M. Negri, A. Névóol, M. Neves, M. Post, L. Specia, M. Turchi, ... K. Verspoor (Eds.), *Proceedings of the Third Conference on Machine Translation: Shared Task Papers* (pp. 488-495). Stroudsburg: Association for Computational Linguistics.
19. Shao, Y., Hardmeier, C., Tiedemann, J., & Nivre, J. (2017). Character-based Joint Segmentation and POS Tagging for Chinese using Bidirectional RNN-CRF. In *The Eighth International Joint Conference on Natural Language Processing: Proceedings of the Conference, Vol. 1 (Long Papers)* (pp. 173-183). Taipei: Asian Federation of Natural Language Processing.
20. Tiedemann, J. (Ed.) (2017). *Proceedings of the 21st Nordic Conference on Computational Linguistics (NoDaLiDa)*. (Linköping Electronic Conference Proceedings; No. 131). Gothenburg: Linköping University Electronic Press.
21. Östling, R., & Tiedemann, J. (2017). Continuous multilinguality with language vectors. In *15th Conference of the European Chapter of the Association for Computational Linguistics: Proceedings of Conference, volume 2: Short Papers* (pp. 644-649). Stroudsburg: Association for Computational Linguistics.
22. Tiedemann, J. (2017). Cross-Lingual Dependency Parsing for Closely Related Languages: Helsinki's Submission to VarDial 2017. In *Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial'2017): Proceedings of the Workshop* (pp. 131-136). Stroudsburg: Association for Computational Linguistics. <https://doi.org/10.18653/v1/w17-1216>
23. Zampieri, M., Malmasi, S., Ljubešić, N., Nakov, P., Ali, A., Tiedemann, J., ... Aepli, N. (2017). Findings of the VarDial Evaluation Campaign 2017. In *Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects* (pp. 1-15). Stroudsburg: Association for Computational Linguistics.
24. Tiedemann, J., & van der Plas, L. (2017). Bootstrapping a Dependency Parser for Maltese - A Real-World Test Case. In M. Wieling, M. Kroon, G. Van Noord, & G. Bouma (Eds.), *From Semantics to Dialectometry: Festschrift in honor of John Nerbonne* (pp. 355-365). (Tributes; No. 32). Milton Keynes: College Publications.
25. Loáiciga, S., Stymne, S., Nakov, P., Hardmeier, C., Tiedemann, J., Cettolo, M., & Versley, Y. (2017). Findings of the 2017 DiscoMT Shared Task on Cross-lingual Pronoun Prediction. In *Discourse in Machine Translation (DiscoMT 2017): Proceedings of the Workshop* (pp. 1-16). Stroudsburg: Association for Computational Linguistics.
26. Kotzé, G., Vandeghinste, V., Martens, S., & Tiedemann, J. (2017). Large aligned treebanks for syntax-based machine translation. *Language Resources and Evaluation*, 51(2), 249-282. <https://doi.org/10.1007/s10579-016-9369-0>
27. Tiedemann, J., & Scherrer, Y. (2017). Neural Machine Translation with Extended Context. In *Proceedings of the Third Workshop on Discourse in Machine Translation* (pp. 82-92). Stroudsburg: Association for Computational Linguistics. <https://doi.org/10.18653/v1/w17-4811>
28. Hurskainen, A., & Tiedemann, J. (2017). Rule-based Machine Translation from English to Finnish. In *Proceedings of the Second Conference on Machine Translation (WMT2017)* (pp. 323-329). Copenhagen, Denmark: Association for Computational Linguistics. <https://doi.org/10.18653/v1/w17-4731>
29. Tjong Kim Sang, E., Bollmann, M., Boschker, R., Casacuberta, F., Dietz, F., Dipper, S., ... Zervanou, K. (2017). The CLIN27 Shared Task: Translating Historical Text to Contemporary Language for Improving Automatic Linguistic Annotation. *Computational Linguistics in the Netherlands*, 7, 53-64.
30. Östling, R., Scherrer, Y., Tiedemann, J., & Nieminen, T. (2017). The Helsinki Neural Machine Translation System. In *Proceedings of the Second Conference on Machine Translation (WMT2017)* (pp. 338-347). Stroudsburg: Association for Computational Linguistics. <https://doi.org/10.18653/v1/w17-4733>
31. Malmasi, S., Zampieri, M., Ljubešić, N., Nakov, P., Ali, A., & Tiedemann, J. (2016). Discriminating between Similar Languages and Arabic Dialect Identification: A Report on the Third DSL Shared Task. In *Third Workshop on NLP for Similar Languages, Varieties and Dialects: Proceedings of the Workshop* (pp. 1-14). The COLING 2016 Organizing Committee .
32. Tiedemann, J., Nichols, J., & Sprouse, R. (2016). Tagging Ingush - Language Technology For Low-Resource Languages Using Resources From Linguistic Field Work. In *Language Technology Resources and Tools for Digital Humanities (LT4DH): Proceedings of the Workshop* (pp. 148-155). Osaka.

33. Öhman, E., Honkela, T., & Tiedemann, J. (2016). The Challenges of Multi-dimensional Sentiment Analysis Across Languages. In *Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media (PEOPLES): Proceedings of the Workshop* (pp. 138-142). Osaka, Japan: The COLING 2016 Organizing Committee .
34. Östling, R., & Tiedemann, J. (2016). Efficient word alignment with Markov Chain Monte Carlo. *The Prague Bulletin of Mathematical Linguistics*, 106, 125-146. <https://doi.org/10.1515/pralin-2016-0013>
35. Tiedemann, J. (2016). A Linear Baseline Classifier for Cross-Lingual Pronoun Prediction. In *The 54th Annual Meeting of the Association for Computational Linguistics: Proceedings of the First Conference on Machine Translation (WMT)* (pp. 616-619). Stroudsburg: The Association for Computational Linguistics.
36. Guillou, L., Hardmeier, C., Nakov, P., Stymne, S., Tiedemann, J., Versley, Y., ... Popescu-Belis, A. (2016). Findings of the 2016 WMT Shared Task on Cross-lingual Pronoun Prediction. In *The 54th Annual Meeting of the Association for Computational Linguistics: Proceedings of the First Conference on Machine Translation (WMT)* (pp. 525-542). Stroudsburg: The Association for Computational Linguistics.
37. Tiedemann, J., Cap, F., Kanerva, J., Ginter, F., Stymne, S., Östling, R., & Weller-Di Marco, M. (2016). Phrase-Based SMT for Finnish with More Data, Better Models and Alternative Alignment and Translation Tools. In *The 54th Annual Meeting of the Association for Computational Linguistics: Proceedings of the First Conference on Machine Translation (WMT)* (pp. 391-398). Stroudsburg: The Association for Computational Linguistics.
38. Tiedemann, J., & Agi, Z. (2016). Synthetic Treebanking for Cross-Lingual Dependency Parsing. *Journal of Artificial Intelligence Research*, 55, 209-248. <https://doi.org/10.1613/jair.4785>
39. Smith, A., Hardmeier, C., & Tiedemann, J. (2016). Climbing Mont BLEU: The Strange World of Reachable High-BLEU Translations. *Baltic Journal of Modern Computing*, 4(2), 269-281.
40. Tiedemann, J. (2016). Finding Alternative Translations in a Large Corpus of Movie Subtitles. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, ... S. Piperidis (Eds.), *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC-2016)* (pp. 3518-3522)
41. Tiedemann, J. (2016). OPUS -- Parallel Corpora for Everyone. *Baltic Journal of Modern Computing*, 384.
42. Lison, P., & Tiedemann, J. (2016). OpenSubtitles2015: Extracting Large Parallel Corpora from Movie and TV Subtitles. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, ... S. Piperidis (Eds.), *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC-2016)*
43. Tiedemann, J. (2015). Baseline Models for Pronoun Prediction and Pronoun-Aware Translation. In *Proceedings of the Second Workshop on Discourse in Machine Translation (DiscoMT)* (pp. 108-114). The Association for Computational Linguistics.
44. Tiedemann, J., Ginter, F., & Kanerva, J. (2015). Morphological Segmentation and OPUS for Finnish-English Machine Translation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation* (pp. 177-183). The Association for Computational Linguistics.
45. Callin, J., Hardmeier, C., & Tiedemann, J. (2015). Part-of-Speech Driven Cross-Lingual Pronoun Prediction with Feed-Forward Neural Networks. In *Proceedings of the Second Workshop on Discourse in Machine Translation (DiscoMT)* (pp. 59-64). The Association for Computational Linguistics.
46. Hardmeier, C., Nakov, P., Stymne, S., Tiedemann, J., Versley, Y., & Cettolo, M. (2015). Pronoun-Focused MT and Cross-Lingual Pronoun Prediction: Findings of the 2015 DiscoMT Shared Task on Pronoun Translation. In *Proceedings of the Second Workshop on Discourse in Machine Translation (DiscoMT)*, (pp. 1-16). The Association for Computational Linguistics.
47. Tiedemann, J. (2015). Cross-Lingual Dependency Parsing with Universal Dependencies and Predicted PoS Labels. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)* (pp. 340-349)
48. Shao, Y., Tiedemann, J., & Nivre, J. (2015). Boosting English-Chinese Machine Transliteration via High Quality Alignment and Multilingual Resources. In *Proceedings of the Fifth Named Entity Workshop, joint with 53rd ACL and the 7th IJCNLP* (pp. 56-60)
49. Tiedemann, J. (2015). Improving the Cross-Lingual Projection of Syntactic Dependencies. In *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA 2015)* (pp. 191-199)
50. Agić, Ž., Tiedemann, J., Merkler, D., Krek, S., Dobrovoljc, K., & Može, S. (2014). Cross-lingual Dependency Parsing of Related Languages with Rich Morphosyntactic Tagsets. In *Language Technology for Closely Related Languages and Language Variants (LT4CloseLang)* (pp. 13-24). The Association for Computational Linguistics.
51. Martinez Garcia, E., Tiedemann, J., España-Bonet, C., & Màrquez, L. (2014). Word's Vector Representations meet Machine Translation. In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation* (pp. 132-134)
52. Zampieri, M., Tan, L., Ljubešić, N., & Tiedemann, J. (2014). A Report on the DSL Shared Task 2014. In *Proceedings of the First Workshop on Applying NLP Tools to Similar Languages, Varieties and Dialects* (pp. 58-67). The Association for Computational Linguistics.

53. Tiedemann, J. (2014). Rediscovering Annotation Projection for Cross-Lingual Parser Induction. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*
54. Skadiņš, R., Tiedemann, J., Rozis, R., & Deksne, D. (2014). Billions of Parallel Words for Free: Building and Using the EU Bookshop Corpus. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC-2014)* (pp. 1850-1855)
55. Zampieri, M., Ljubešić, N., Tiedemann, J., & Tan, L. (2014). Merging Comparable Data Sources for the Discrimination of Similar Languages: The DSL Corpus Collection. In *Proceedings of the 7th Workshop on Building and Using Comparable Corpora (BUCC)* (pp. 6-10)
56. Guillou, L., Hardmeier, C., Smith, A., Tiedemann, J., & Webber, B. (2014). ParCor 1.0: A Parallel Pronoun-Coreference Corpus to Support Statistical MT. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)* (pp. 3191-3198)
57. Hardmeier, C., Stymne, S., Smith, A., Tiedemann, J., & Nivre, J. (2014). Anaphora Models and Reordering for Phrase-Based SMT. In *Proceedings of the Ninth Workshop on Statistical Machine Translation* (pp. 122-129). The Association for Computational Linguistics.
58. Stymne, S., Tiedemann, J., & Nivre, J. (2014). Estimating Word Alignment Quality for SMT Reordering Tasks. In *Proceedings of the Ninth Workshop on Statistical Machine Translation* (pp. 275-286). The Association for Computational Linguistics.
59. Tiedemann, J. (2014). Improved Text Extraction from PDF Documents for Large-Scale Natural Language Processing. In A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing* (Vol. 1, pp. 102-112). (Lecture Notes in Computer Science LNCS 8403). Berlin: Springer.
60. Tiedemann, J., Agić, Ž., & Nivre, J. (2014). Treebank Translation for Cross-Lingual Parser Induction. In *Proceedings of the Eighteenth Conference on Computational Language Learning* (pp. 130-140). The Association for Computational Linguistics.
61. Tiedemann, J. (2013). Tools for lexicographic use of parallel and comparable corpora. In R. H. Gouws, U. Heid, W. Schweickard, & H. E. Wiegand (Eds.), *An International Encyclopedia of Lexicography -- Supplementary Volume* (Vol. 4, pp. 1433-1444). (HSK - Dictionaries). De Gruyter Mouton.
62. Tiedemann, J., van der Plas, L., & Villada Moirón, B. (2013). Bitexts as Semantic Mirrors. In *Unknown host publication*
63. Hardmeier, C., Tiedemann, J., & Nivre, J. (2013). Latent Anaphora Resolution for Cross-Lingual Pronoun Prediction. In *Unknown host publication*
64. Hardmeier, C., Stymne, S., Tiedemann, J., & Nivre, J. (2013). Docent: A Document-Level Decoder for Phrase-Based Statistical Machine Translation. In *Unknown host publication* (pp. 193-198)
65. Stymne, S., Hardmeier, C., Tiedemann, J., & Nivre, J. (2013). Feature Weight Optimization for Discourse-Level SMT. In *Unknown host publication* (pp. 60-69)
66. Webber, B., Popescu-Belis, A., Markert, K., & Tiedemann, J. (Eds.) (2013). *Proceedings of the Workshop on Discourse in Machine Translation*. Sofia, Bulgaria: Association for Computational Linguistics.
67. Stymne, S., Hardmeier, C., Tiedemann, J., & Nivre, J. (2013). Tunable Distortion Limits and Corpus Cleaning for SMT. In *Unknown host publication* (pp. 225-231)
68. Pettersson, E., Megyesi, B. B., & Tiedemann, J. (2013). An SMT approach to automatic annotation of historical text. In *Unknown host publication*
69. Stymne, S., Tiedemann, J., Hardmeier, C., & Nivre, J. (2013). Statistical Machine Translation with Readability Constraints. In *Unknown host publication* (pp. 375-386)
70. Tiedemann, J., & Nakov, P. (2013). Analyzing the Use of Character-Level Translation with Sparse and Noisy Datasets. In *Unknown host publication* (pp. 676-684)
71. Tiedemann, J. (2013). Book review of Markus Dickinson, Chris Brew and Detmar Meurers: Language and Computers. *Machine Translation*, 27, 309-312. <https://doi.org/10.1007/s10590-013-9143-7>
72. Tiedemann, J., & Ljubesic, N. (2012). Efficient Discrimination Between Closely Related Languages. In *Unknown host publication* (pp. 2619-2634)
73. Nakov, P., & Tiedemann, J. (2012). Combining Word-Level and Character-Level Models for Machine Translation Between Closely-Related Languages. In *Unknown host publication* (pp. 301-305)
74. Hardmeier, C., Nivre, J., & Tiedemann, J. (2012). Document-Wide Decoding for Phrase-Based Statistical Machine Translation. In *Unknown host publication* (pp. 1179-1190)
75. Vasijevs, A., Skadis, R., & Tiedemann, J. (2012). LetsMT!: Cloud-Based Platform for Do-It-Yourself Machine Translation. In *Unknown host publication* (pp. 43-48)
76. Hardmeier, C., Nivre, J., & Tiedemann, J. (2012). Tree Kernels for Machine Translation Quality Estimation. In *Unknown host publication* (pp. 109-113)

77. Tiedemann, J., Hansen, D. H., Offersgaard, L., Olsen, S., & Zumpe, M. (2012). A Distributed Resource Repository for Cloud-Based Machine Translation. In N. Calzolari, K. Choukri, T. Declerck, M. U. Dogan, B. Maegaard, J. Mariani, J. Odijk, ... S. Piperidis (Eds.), *Unknown host publication* (pp. 2207-2213)
78. Kotzé, G., Vandeghinste, V., Martens, S., & Tiedemann, J. (2012). Large aligned treebanks for syntax-based machine translation. In N. Calzolari, K. Choukri, T. Declerck, M. U. Dogan, B. Maegaard, J. Mariani, J. Odijk, ... S. Piperidis (Eds.), *Unknown host publication* (pp. 467-473)
79. Tiedemann, J. (2012). Parallel Data, Tools and Interfaces in OPUS. In N. Calzolari, K. Choukri, T. Declerck, M. U. Dogan, B. Maegaard, J. Mariani, J. Odijk, ... S. Piperidis (Eds.), *Unknown host publication* (pp. 2214-2218)
80. Tiedemann, J. (2012). Character-Based Pivot Translations for Under-Resourced Languages and Domains. In *Unknown host publication* (pp. 141-151)
81. Ehrentraut, C., Tiedemann, J., Dalianis, H., & Tanushi, H. (2012). Detection of Hospital Acquired Infections in Sparse and Noisy Swedish Patient Records. In *Unknown host publication*
82. Tiedemann, J., Zumpe, M., & Schleussner, S. (2012). Opening OPUS for User Contributions. In *Unknown host publication*
83. Vandeghinste, V., Martens, S., Kotzé, G., Tiedemann, J., Bogaert, J. V. D., Smet, K. D., ... Noord, G. V. (2012). Parse and Corpus-based Machine Translation. In P. Spyns, & J. Odijk (Eds.), *Essential Speech and Language Technology for Dutch: resources, tools and applications* (Theory and Applications of Natural Language Processing). Springer.
84. Hardmeier, C., Tiedemann, J., Saers, M., Federico, M., & Prashant, M. (2011). The Uppsala-FBK systems at WMT 2011. In *Unknown host publication* (pp. 372-378)
85. Plas, L. V. D., Tiedemann, J., & Fahmi, I. (2011). Automatic extraction of medical term variants from multilingual parallel translations. In *Theory and Applications of Natural Language Processing* (Theory and Applications of Natural Language Processing).
86. Tiedemann, J., & Hirst, G. (Ed.) (2011). *Bitext Alignment*. (Synthesis Lectures on Human Language Technologies). San Rafael, California (USA): Morgan & Claypool.
87. Plas, L. V. D., Tiedemann, J., & Manguin, J.-L. (2011). Synonym acquisition across domains and languages. In *Advances in Distributed Agent-based Retrieval Tools* (Studies in Computational Intelligence). Berlin: Springer.
88. van der Plas, L., & Tiedemann, J. (2010). Finding Medical Term Variations using Parallel Corpora and Distributional Similarity. In *Unknown host publication* (pp. 28-37)
89. Tiedemann, J. (2010). Context Adaptation in Statistical Machine Translation Using Models with Exponentially Decaying Cache. In *Unknown host publication* (pp. 8-15)
90. Ill, H. D., Deoskar, T., McClosky, D., Plank, B., & Tiedemann, J. (Eds.) (2010). *Proceedings of the 2010 Workshop on Domain Adaptation for Natural Language Processing*. Uppsala, Sweden: Association for Computational Linguistics.
91. Tiedemann, J. (2010). To Cache or Not To Cache? Experiments with Adaptive Models in Statistical Machine Translation. In *Unknown host publication* (pp. 189-194)
92. Islam, M. Z., Tiedemann, J., & Eisele, A. (2010). English to Bangla Phrase-Based Machine Translation. In *Unknown host publication*
93. Tiedemann, J. (2010). Lingua-Align: An Experimental Toolbox for Automatic Tree-to-Tree Alignment. In N. C. C. Chair, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, ... D. Tapias (Eds.), *Unknown host publication*
94. Plas, L. V. D., Tiedemann, J., & Manguin, J.-L. (2010). Automatic acquisition of synonyms for French using parallel corpora. In *Unknown host publication*
95. Tiedemann, J., & Weijnitz, P. (2010). Let's MT! - A Platform for Sharing SMT Training Data. In *Unknown host publication*
96. Ahrenberg, L., Tiedemann, J., & Volk, M. (Eds.) (2010). *Proceedings of the Workshop on Annotation and Exploitation of Parallel Corpora*. (NEALT Proceedings Series). Northern European Association for Language Technology .
97. Tiedemann, J., & Nabende, P. (2009). Translating Transliterations. *International journal of computing and ICT research* , 3(1), 33-41.
98. Tiedemann, J., & Kotzé, G. (2009). A Discriminative Approach to Tree Alignment. In *Unknown host publication* (pp. 33-39)
99. Tiedemann, J. (2009). Evidence-Based Word Alignment. In I. Ilisei, V. Pekar, & S. Bernardini (Eds.), *Unknown host publication* (pp. 28-32)
100. Tiedemann, J. (2009). Character-based PSMT for Closely Related Languages. In L. Marqués, & H. Somers (Eds.), *Unknown host publication* (pp. 12 - 19)
101. Tiedemann, J. (2009). Translating Questions for Cross-Lingual QA. In L. Marqués, & H. Somers (Eds.), *Unknown host publication* (pp. 112 - 119)

102. Tiedemann, J., & Kotzé, G. (2009). Building a Large Machine-Aligned Parallel Treebank. In *Unknown host publication*
103. Bertels, A., Fairon, C., Tiedemann, J., & Verlinde, S. (2009). Corpus parallèles et corpus ciblés au secours du dictionnaire de traduction. In *Cahiers de lexicologie* (pp. 199-219). (Revue). Classiques Garnier.
104. Tiedemann, J. (2009). News from OPUS - A Collection of Multilingual Parallel Corpora with Tools and Interfaces. In N. Nicolov, K. Bontcheva, G. Angelova, & R. Mitkov (Eds.), *Recent Advances in Natural Language Processing* (Vol. V, pp. 237-248)