

# Publications

Raphael Rubino

## References

- [1] Bianka Buschbeck, Raj Dabre, Miriam Exel, Matthias Huck, Patrick Huy, Raphael Rubino, and Hideki Tanaka. A multilingual multiway evaluation data set for structured document translation of asian languages. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2022*, pages 237–245, 2022.
- [2] Zhengdong Yang, Wangjin Zhou, Chenhui Chu, Sheng Li, Raj Dabre, Raphael Rubino, and Yi Zhao. Fusion of self-supervised learned models for mos prediction. *arXiv preprint arXiv:2204.04855*, 2022.
- [3] Benjamin Marie, Atsushi Fujita, and Raphael Rubino. Scientific credibility of machine translation research: A meta-evaluation of 769 papers. *arXiv preprint arXiv:2106.15195*, 2021.
- [4] Raphael Rubino, Atsushi Fujita, and Benjamin Marie. Error identification for machine translation with metric embedding and attention. In *Proceedings of the 2nd Workshop on Evaluation and Comparison of NLP Systems*, pages 146–156, 2021.
- [5] Raphaël Rubino, Atsushi Fujita, and Benjamin Marie. Nict kyoto submission for the wmt’21 quality estimation task: Multimetric multilingual pretraining for critical error detection. In *Proceedings of the Sixth Conference on Machine Translation*, pages 941–947, 2021.
- [6] Benjamin Marie, Raphael Rubino, and Atsushi Fujita. Combination of neural machine translation systems at WMT20. In *Proceedings of the Fifth Conference on Machine Translation*, pages 230–238, Online, November 2020. Association for Computational Linguistics.

- [7] Raj Dabre, Raphael Rubino, and Atsushi Fujita. Balancing cost and benefit with tied-multi transformers. *arXiv preprint arXiv:2002.08614*, 2020.
- [8] Benjamin Marie, Raphael Rubino, and Atsushi Fujita. Tagged back-translation revisited: Why does it really work? In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5990–5997, 2020.
- [9] Raphael Rubino. Nict kyoto submission for the wmt’20 quality estimation task: Intermediate training for domain and task adaptation. In *Proceedings of the Fifth Conference on Machine Translation*, pages 1042–1048, 2020.
- [10] Raphael Rubino, Benjamin Marie, Raj Dabre, Atushi Fujita, Masao Utiyama, and Eiichiro Sumita. Extremely low-resource neural machine translation for asian languages. *Machine Translation*, 34(4):347–382, 2020.
- [11] Raphael Rubino and Eiichiro Sumita. Intermediate self-supervised learning for machine translation quality estimation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4355–4360, 2020.
- [12] Rajen Chatterjee, Christian Federmann, Matteo Negri, and Marco Turchi. Findings of the wmt 2019 shared task on automatic post-editing. In *Proceedings of the Fourth Conference on Machine Translation (Volume 3: Shared Task Papers, Day 2)*, pages 11–28, 2019.
- [13] Daria Pylypenko and Raphael Rubino. Dfki-mlt system description for the wmt18 automatic post-editing task. In *Proceedings of the Third Conference on Machine Translation: Shared Task Papers*, pages 836–839, 2018.
- [14] Ahmad Taie, Raphael Rubino, and Josef van Genabith. Infodens: an open-source framework for learning text representations. *arXiv preprint arXiv:1810.07091*, 2018.
- [15] Ondřej Bojar, Christian Buck, Rajen Chatterjee, Christian Federmann, Yvette Graham, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes,

- Philipp Koehn, and Julia Kreutzer. Proceedings of the second conference on machine translation. In *Proceedings of the Second Conference on Machine Translation*, 2017.
- [16] Ondřej Bojar, Rajen Chatterjee, Christian Federmann, Yvette Graham, Barry Haddow, Shujian Huang, Matthias Huck, Philipp Koehn, Qun Liu, Varvara Logacheva, et al. Findings of the 2017 conference on machine translation (wmt17). In *Proceedings of the Second Conference on Machine Translation*, pages 169–214, 2017.
  - [17] Wei Shi, Frances Yung, Raphael Rubino, and Vera Demberg. Using explicit discourse connectives in translation for implicit discourse relation classification. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 484–495, 2017.
  - [18] Hans Uszkoreit, Aleksandra Gabryszak, Leonhard Hennig, Jörg Steffen, Renlong Ai, Stephan Busemann, Jon Dehdari, Josef van Genabith, Georg Heigold, Nils Rethmeier, et al. Common round: Application of language technologies to large-scale web debates. In *Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pages 5–8, 2017.
  - [19] Ondřej Bojar, Rajen Chatterjee, Christian Federmann, Yvette Graham, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Philipp Koehn, Varvara Logacheva, Christof Monz, et al. Marco zampieri (2016). “findings of the 2016 conference on machine translation (wmt16).”. In *Proceedings of the First Conference on Machine Translation. Volume 2: Shared Task Papers, Berlin 2016*, pages 131–198, 2016.
  - [20] Ondrej Bojar, Rajen Chatterjee, Christian Federmann, Yvette Graham, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Philipp Koehn, Varvara Logacheva, Christof Monz, et al. Findings of the 2016 conference on machine translation. In *ACL 2016 FIRST CONFERENCE ON MACHINE TRANSLATION (WMT16)*, pages 131–198. The Association for Computational Linguistics, 2016.
  - [21] T Pirinen, Antonio Toral, and Raphael Rubino. Rule-based and statistical morph segments in english-finnish smt. In *Proceedings of the Second*

*International Workshop on Computational Linguistics for Uralic Languages*, pages 56–69, 2016.

- [22] Raphael Rubino, Stefania Degaetano-Ortlieb, Elke Teich, and Josef van Genabith. Modeling diachronic change in scientific writing with information density. In *Proceedings of the 26th International Conference on Computational Linguistics (COLING 2016)*, 2016.
- [23] Raphael Rubino, Ekaterina Lapshinova-Koltunski, and Josef Van Genabith. Information density and quality estimation features as translationese indicators for human translation classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies*, pages 960–970, 2016.
- [24] Raphaël Rubino, Benjamin Lecouteux, and Georges Linares. *Amélioration des modèles de repli par des sacs de mots et des n-grammes à variables*. PhD thesis, LIG, 2016.
- [25] Antonio Toral, Miquel Esplà-Gomis, Filip Klubička, Nikola Ljubešić, Vassilis Papavassiliou, Prokopis Prokopidis, Raphael Rubino, and Andy Way. Crawl and crowd to bring machine translation to under-resourced languages. *Language Resources and Evaluation*, pages 1–33, 2016.
- [26] Antonio Toral, Miquel Esplà-Gomis, Filip Klubička, Nikola Ljubešić, Vassilis Papavassiliou, Prokopis Prokopidis, Raphael Rubino, and Andy Way. Tourism english-croatian parallel corpus 2.0. *CLARIN.SI data & tools*, 2016.
- [27] Antonio Toral, Sergio ORTIZ-ROJAS, Mikel FORCADA, Nikola Ljubecic, and Prokopis Prokopidis. Abu-matran: Automatic building of machine translation. *Baltic Journal of Modern Computing*, 4(2), 2016.
- [28] Antonio Toral, Raphael Rubino, and Gema Ramírez-Sánchez. Re-assessing the impact of smt techniques with human evaluation: a case study on english–croatian. *Baltic Journal of Modern Computing*, 4(2):368, 2016.
- [29] Pratyush Banerjee, Raphael Rubino, Johann Roturier, and Josef van Genabith. Quality estimation-guided supplementary data selection for

domain adaptation of statistical machine translation. *Machine Translation*, 29:77–100, 2015.

- [30] Miquel Esplà-Gomis, Nikola Ljubesic, Filip Klubicka, Prokopis Prokopidis, Vassilis Papavassiliou, Antonio Toral, Tommi Pirinen, Andy Way, Raphael Rubino, Gema Ramírez-Sánchez, et al. Automatic acquisition of machine translation resources in the abu-matran project. *Procesamiento del Lenguaje Natural*, 1(55):185–188, 2015.
- [31] Raphael Rubino, Tommi A Pirinen, Miquel Espla-Gomis, Nikola Ljubešić, Sergio Ortiz-Rojas, Vassilis Papavassiliou, Prokopis Prokopidis, and Antonio Toral. Abu-matran at wmt 2015 translation task: Morphological segmentation and web crawling. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 184–191, 2015.
- [32] Ondřej Bojar, Christian Buck, Christian Federmann, Barry Haddow, Philipp Koehn, Christof Monz, Matt Post, and Lucia Specia. Proceedings of the ninth workshop on statistical machine translation. In *Proceedings of the Ninth Workshop on Statistical Machine Translation*, 2014.
- [33] Miquel Espla-Gomis, Mikel L Forcada, Nikola Ljubešić, Vassilis Papavassiliou, Prokopis Prokopidis, Sergio Ortiz-Rojas, Tommi Pirinen, Raphaël Rubino, and Antonio Toral. Deliverable d3. 1b acquisition for cycle 2. Technical report, Technical report, Abu-MaTran project, 2014.
- [34] Mikel L. Forcada, Sergio Ortiz-Rojas, Tommi Pirinen, Raphael Rubino, and Antonio Toral. D4. 1b mt systems for the second development cycle, 2014.
- [35] Rasoul Kaljahi, Jennifer Foster, Johann Roturier, and Raphael Rubino. Quality estimation of english-french machine translation: A detailed study of the role of syntax. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 2052–2063, 2014.
- [36] Raphael Rubino, Antonio Toral, Nikola Ljubešić, and Gema Ramírez-Sánchez. Quality estimation for synthetic parallel data generation. In *Language Resources and Evaluation Conference*, 2014.

- [37] Raphael Rubino, Antonio Toral, Victor M Sánchez-Cartagena, Jorge Ferrández-Tordera, Sergio Ortiz-Rojas, Gema Ramírez-Sánchez, Felipe Sánchez-Martínez, and Andy Way. Abu-matran at wmt 2014 translation task: Two-step data selection and rbmt-style synthetic rules. In *Proceedings of the ninth workshop on statistical machine translation*, pages 171–177, 2014.
- [38] Antonio Toral, Raphael Rubino, Miquel Espla-Gomis, Tommi Pirinen, Andy Way, and Gema Ramirez-Sánchez. Extrinsic evaluation of web-crawlers in machine translation: a case study on croatian–english for the tourism domain. In *The Seventeenth Annual Conference of the European Association for Machine Translation (EAMT)*, 2014.
- [39] Pratyush Banerjee, Raphael Rubino, Johann Roturier, and Josef van Genabith. Quality estimation-guided data selection for domain adaptation of smt. In *MT Summit XIV*, 2013.
- [40] Rasoul Samed Zadeh Kaljahi, Jennifer Foster, Raphael Rubino, Johann Roturier, and Fred Hollowood. Parser accuracy in quality estimation of machine translation: a tree kernel approach. 2013.
- [41] Erwan Moreau and Raphael Rubino. An approach using style classification features for quality estimation. In *Proceedings of the Eighth Workshop on Statistical Machine Translation*, pages 429–434, 2013.
- [42] Raphael Rubino, José GC de Souza, Jennifer Foster, and Lucia Specia. Topic models for translation quality estimation for gisting purposes. In *Machine Translation Summit XIV*, 2013.
- [43] Raphael Rubino, Jennifer Foster, Rasoul Samad Zadeh Kaljahi, Johann Roturier, and Fred Hollowood. Estimating the quality of translated user-generated content. In *6th International Joint Conference on Natural Language Processing (IJCNLP)*, pages 1167–1173, 2013.
- [44] Raphael Rubino, Antonio Toral, Santiago Cortés Vaíllo, Jun Xie, Xiaofeng Wu, Stephen Doherty, and Qun Liu. The engl-dcu-prompsit translation systems for wmt13. In *Proceedings of the Eighth Workshop on Statistical Machine Translation*, pages 213–218. Association for Computational Linguistics, 2013.

- [45] Raphael Rubino, Joachim Wagner, Jennifer Foster, Johann Roturier, Rasoul Samad Zadeh Kaljahi, and Fred Hollowood. Dcu-symantec at the wmt 2013 quality estimation shared task. In *Proceedings of the Eighth Workshop on Statistical Machine Translation*, pages 392–397, 2013.
- [46] Hanna Béchara, Raphaël Rubino, Yifan He, Yanjun Ma, and Josef van Genabith. An evaluation of statistical post-editing systems applied to rbmt and smt systems. In *COLING*, 2012.
- [47] Béchara Hanna. Raphaël rubino. yifan he. yanjun ma. josef van genabith. an evaluation of statistical post-editing systems applied to rbmt and smt systems. In *COLING*, volume 21, pages 5–230, 2012.
- [48] Rasoul Samad Zadeh Kaljahi, Raphael Rubino, Johann Roturier, and Jennifer Foster. A detailed analysis of phrase-based and syntax-based mt: The search for systematic differences. In *Proceedings of the 10th Conference of the Association for Machine Translation in the Americas: Research Papers*, 2012.
- [49] Tsuyoshi Okita, Raphaël Rubino, and Josef van Genabith. Sentence-level quality estimation for mt system combination. In *COLING 2012 ML4HMT Workshop*, 2012.
- [50] Raphael Rubino, Jennifer Foster, Joachim Wagner, Johann Roturier, Rasul Samad Zadeh Kaljahi, and Fred Hollowood. Dcu-symantec submission for the wmt 2012 quality estimation task. In *WMT12*, 2012.
- [51] Raphaël Rubino, Stéphane Huet, Fabrice Lefèvre, and Georges Linarès. Post-édition statistique pour l’adaptation aux domaines de spécialité en traduction automatique. In *TALN*, 2012.
- [52] Raphaël Rubino, Stéphane Huet, Fabrice Lefevre, and Georges Linares. Statistical post-editing of machine translation for domain adaptation. In *EAMT*, 2012.
- [53] Marion Potet, Raphaël Rubino, Benjamin Lecouteux, Stéphane Huet, Laurent Besacier, Hervé Blanchon, and Fabrice Lefevre. The liga (lig/lia) machine translation system for wmt 2011. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 440–446, 2011.

- [54] Raphaël Rubino. *Traduction automatique statistique et adaptation à un domaine spécialisé. (Domain Adaptation for Statistical Machine Translation)*. PhD thesis, University of Avignon, France, 2011.
- [55] Raphaël Rubino, Iria Da Cunha, and Georges Linares. Extracción automática de léxico bilingüe: experimentos en español y catalán. In *IEEE 14th International Conference on Interactive Collaborative Learning (ICL)*, 2011.
- [56] Raphaël Rubino and Georges Linares. A multi-view approach for term translation spotting. In *Computational Linguistics and Intelligent Text Processing: 12th International Conference, CICLing 2011, Tokyo, Japan, February 20-26, 2011. Proceedings, Part II 12*, pages 29–40. Springer Berlin Heidelberg, 2011.
- [57] Raphaël Rubino and Georges Linares. Une approche multi-vue pour l’extraction terminologique bilingue. In *Conférence en Recherche d’Information et Applications (CORIA)*, pages 97–111, 2011.
- [58] Grégory Senay, Stanislas Oger, Raphaël Rubino, Georges Linares, and Thomas Parent. Audio indexing on a medical video database: the avison project. In *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*, volume 3, pages 1669–1673. IEEE, 2011.
- [59] Benjamin Lecouteux, Raphaël Rubino, and Georges Linares. Improving back-off models with bag of words and hollow-grams. In *Interspeech*, 2010.
- [60] Raphaël Rubino. Exploring context variation and lexicon coverage in projection-based approach for term translation. In *Proceedings of the Student Research Workshop*, pages 66–70, 2009.