

# Internship on Context-Aware Neural Machine Translation

*6 months @ GETALP reseach team, Université Grenoble-Alpes*

**Scientific subject:** The field of machine translation has seen substantial improvements in the latest years, mainly related to the advent of the Transformer model [10]. However, some major challenges are still to be won to reach human-level translation performance, particularly when translating full documents, e.g., articles, movie subtitles, books. Among these challenges, there is the capability of contextualizing the current sentence to be translated with the other sentences of the document [6]. In fact, a sentence can be very ambiguous when we don't know what was said previously, and this may impact the quality of its translation [6, 11]. Concrete examples are the improper translation of pronouns with inter-sentential antecedents or inconsistent lexical choices [2, 1]. Although this problem has recently triggered a lot of attention into context-aware translation [4, 9, 7, 12, 13, 5, 14, 3], there still remain a lot of room for improvement. Crucially, one of the limits of the Transformer model with respect to contextual modeling is its computational and memory efficiency, which grows quadratically with the length of the input. In the case of machine translation, this makes it costly to model multiple sentences in a row, one in context of each other. However, a number of research works this year proposed alternatives to the standard Transformer model that aim at solving its efficiency issue [8]. This internship aims at studying the application of these new models to the task of context-aware machine translation for the translation of documents, with the aim of bridging the gap between machine and human translators.

**Tasks:** inspired by the aforementioned studies, the intern will devise a context-aware neuronal translation model using [Fairseq](#), a python framework based on Pytorch for sequence to sequence deep learning. Subsequently, the intern will investigate and analyse the model by running multiple experiments on various datasets, using the GPU servers available in the lab.

**Position:** research internship of 5-6 months starting from February 2021 (flexible) for master students.

**Host:** the [GETALP](#) research team of the *Laboratoire d'Informatique de Grenoble* (LIG), University Grenoble-Alpes.

**Supervisors:** [Marco Dinarelli](#) (CNRS Researcher at LIG) and [Lorenzo Lupo](#) (PhD student).

**Candidate profile:**

- master student (preferably M2) in computer science, applied mathematics or engineering;
- Solid computer science or data science skills, especially Python and data manipulation;
- Prior experience with machine learning models;
- Good understanding of deep learning fundamentals;
- Proactive and autonomous mindset, passionate for machine learning;
- Fluent English proficiency;
- Prior experience with Pytorch is a plus.
- Prior experience in Natural Language Processing is a plus.

**To apply:** send a CV and a motivation letter to [marco.dinarelli@univ-grenoble-alpes.fr](mailto:marco.dinarelli@univ-grenoble-alpes.fr) and [lorenzo.lupo@univ-grenoble-alpes.fr](mailto:lorenzo.lupo@univ-grenoble-alpes.fr).

## References

- [1] R. Bawden, R. Sennrich, A. Birch, and B. Haddow. Evaluating Discourse Phenomena in Neural Machine Translation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1304–1313, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. 00056.
- [2] L. Guillou and C. Hardmeier. Automatic Reference-Based Evaluation of Pronoun Translation Misses the Point. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4797–4802, Brussels, Belgium, Oct. 2018. Association for Computational Linguistics. 00008.
- [3] J. Huo, C. Herold, Y. Gao, L. Dahlmann, S. Khadivi, and H. Ney. Diving Deep into Context-Aware Neural Machine Translation. *arXiv:2010.09482 [cs]*, Oct. 2020. 00000 arXiv: 2010.09482.
- [4] S. Jean, S. Lauly, O. Firat, and K. Cho. Does Neural Machine Translation Benefit from Larger Context? *arXiv:1704.05135 [cs, stat]*, Apr. 2017. 00039 arXiv: 1704.05135.
- [5] M. Junczys-Dowmunt. Microsoft Translator at WMT 2019: Towards Large-Scale Document-Level Neural Machine Translation. In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, pages 225–233, Florence, Italy, Aug. 2019. Association for Computational Linguistics. 00000.
- [6] S. Läubli, R. Sennrich, and M. Volk. Has Machine Translation Achieved Human Parity? A Case for Document-level Evaluation. *arXiv:1808.07048 [cs]*, Aug. 2018. 00038 arXiv: 1808.07048.
- [7] L. Miculicich, D. Ram, N. Pappas, and J. Henderson. Document-Level Neural Machine Translation with Hierarchical Attention Networks. *arXiv:1809.01576 [cs]*, Oct. 2018. 00029 arXiv: 1809.01576.
- [8] Y. Tay, M. Dehghani, D. Bahri, and D. Metzler. Efficient Transformers: A Survey. *arXiv:2009.06732 [cs]*, Sept. 2020. 00008 arXiv: 2009.06732.
- [9] J. Tiedemann and Y. Scherrer. Neural Machine Translation with Extended Context. In *Proceedings of the Third Workshop on Discourse in Machine Translation*, pages 82–92, Copenhagen, Denmark, Sept. 2017. Association for Computational Linguistics. 00040.
- [10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention Is All You Need. *arXiv:1706.03762 [cs]*, Dec. 2017. 05728 arXiv: 1706.03762.
- [11] E. Voita, R. Sennrich, and I. Titov. When a Good Translation is Wrong in Context: Context-Aware Machine Translation Improves on Deixis, Ellipsis, and Lexical Cohesion. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1198–1212, Florence, Italy, July 2019. Association for Computational Linguistics. 00007.
- [12] E. Voita, P. Serdyukov, R. Sennrich, and I. Titov. Context-Aware Neural Machine Translation Learns Anaphora Resolution. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1264–1274, Melbourne, Australia, July 2018. Association for Computational Linguistics. 00049.
- [13] E. Voita, D. Talbot, F. Moiseev, R. Sennrich, and I. Titov. Analyzing Multi-Head Self-Attention: Specialized Heads Do the Heavy Lifting, the Rest Can Be Pruned. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5797–5808, Florence, Italy, July 2019. Association for Computational Linguistics. 00033.
- [14] L. Yu, L. Sartran, W. Stokowiec, W. Ling, L. Kong, P. Blunsom, and C. Dyer. Better Document-Level Machine Translation with Bayes’ Rule. *arXiv:1910.00553 [cs]*, July 2020. 00001 arXiv: 1910.00553.