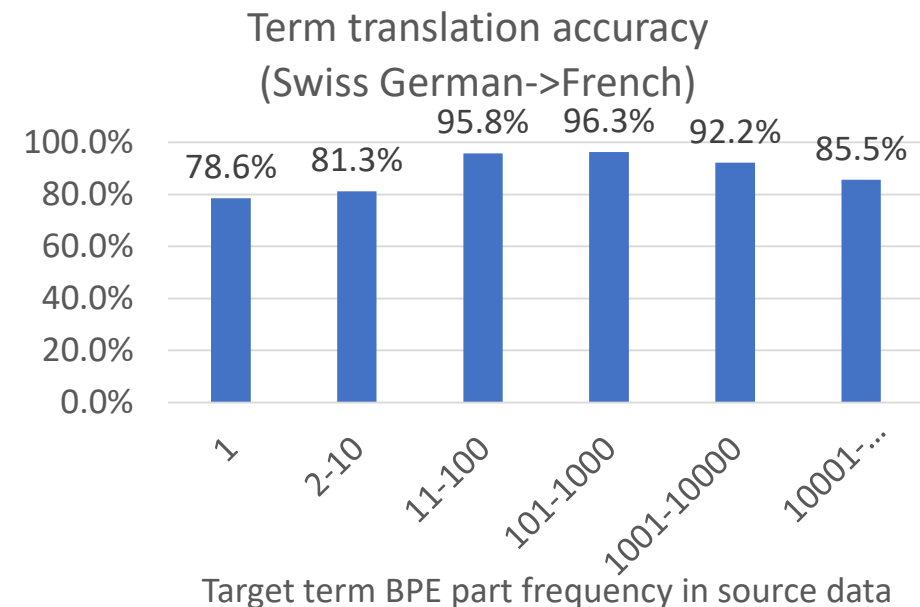


Terminology Translation: From Research to Production

- **Challenge** – some terms consist of rare BPE parts and are translated poorly
- **Solution 1** – make sure that training data TLA contain BPE parts relevant to terms used at the test time
- **Solution 2** – filter term collections such that out-of-vocabulary terms are ignored
- **Solution 3** – use character representations of TLA (Niehues, 2021)



Main Takeaway

- Terminology integration is a **cascade** of terminology creation, curation, identification and only then translation using MT.
- Terminology creation and curation is and should be done by **professional translators and domain experts**.
- **Poor terminology management choices will be propagated in downstream processes** – terminology identification and terminology translation, and will impede the final translation quality.

Main Takeaway

To mitigate error propagation, pay attention to how terminology is managed and prepared for MT such that it is MT-ready

- Make sure that terminology is consistent
- Make sure that terminology is domain-specific
- Do not overexaggerate with needless wordiness
 - Online/dynamic learning, and translation memories may be better suited for such data
- Provide enough metadata such that your term identification method is able to function properly

References

Sennrich, Rico, and Barry Haddow. "Linguistic Input Features Improve Neural Machine Translation." *Proceedings of the First Conference on Machine Translation: Volume 1, Research Papers*. 2016.

Post, Matt, and David Vilar. "Fast Lexically Constrained Decoding with Dynamic Beam Allocation for Neural Machine Translation." *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. 2018.

Dinu, Georgiana, et al. "Training Neural Machine Translation to Apply Terminology Constraints." *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2019.

Bergmanis, Toms, and Mārcis Pinnis. "Facilitating Terminology Translation with Target Lemma Annotations." *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. 2021.

Niehues, Jan. "Continuous Learning in Neural Machine Translation using Bilingual Dictionaries." *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. 2021.