

Training of neural machine translation model to apply terminology constraints for language with robust inflection.

Jakub Konieczny

Adam Mickiewicz University
ul. Wieniawskiego 1, 61-712 Poznań, Poland
Email: jakkon6@st.amu.edu.pl

Abstract—The goal of this study is to explore the transformer’s capability of domain translation into a morphologically rich language. Satisfactory translation into Polish requires inflection by tense, number, and person, taking into account six declination cases. The ideal outcome of this study would be to prove that the method proposed by Dinu is capable of training the transformer to translate English to Polish in domain-specific scenarios. Achieving metrics similar to Nowakowski would result in a “zero-shot” translator with a considerably higher translation speed

I. INTRODUCTION

THE goal of this study is to explore the transformer’s capability of domain translation into a morphologically rich language. Satisfactory translation into Polish requires inflection by tense, number, and person, taking into account six declination cases. The ideal outcome of this study would be to prove that the method proposed by Dinu [1] is capable of training the transformer to translate English to Polish in domain-specific scenarios. Achieving metrics similar to Nowakowski [2] would result in a “zero-shot” translator with a considerably higher translation speed.

II. POTENTIAL APPLICATIONS

There is no shortage of uses for machine translation in general. It is reasonable to believe that low latency domain-specific translation will be in demand. Due to the translation speed equal to unconstrained translation, “train-by” models Dinu [1] can be adopted in simultaneous translation, broadening their usability. Institutions like EU Parliament could heavily benefit from such a system.

III. RELATED WORK

Among Slavic languages, Polish is not as well researched as Russian when it comes to machine translation. Nowakowski [2] investigated the application of Constrained Beam Search proposed by Anderson [3] and Hokamp and Liu [4] for English to Polish translation in domain-specific applications. This work examines “train-by” models described by Dinu [1] in the same context. The study, hopefully, will compliment Nowakowski’s [2] findings.

IV. MOTIVATION

The “train-by” approach does not give the premise of achieving translation and inflection quality superior to the Constrained Beam Search method. What it ensures, however, is translation speed equal to that of unconstrained translation. Bearing that in mind, and the fact, that both solutions allow translations of phrases that were not part of training vocabulary, it seems that if capable of ensuring similar quality, the “train-by” approach is worth consideration in most real-life scenarios.

V. EXPERIMENT

The main idea behind the experiment is to determine the approach that, applying lexical constraints, allows to achieve the best correctness of term inflection. To evaluate the solution, the following metrics will be used: BLEU [5], Term Rate [6], as well as Placement Rate, Duplication Rate and Inflection Rate [2]. The training corpus consists of Europarl v8, EUBookshop v2, JRC-Acquis v3.0, TildeMODEL v2018, Wikipedia v1.0, and most of DGT v2019 corpora, filtered with Bicleaner 2 and Bifixer 3 [7]. The corpus includes 3103819 segments. Test sets are made of 1000 and 1104 segment pairs, validation sets consist of 2000 sentences from DGT corpus. The lexical constraints will be extracted from the Compendium of Accounting in Polish & English [8] and will consist of 1197 term pairs. The architecture of choice will be a standard NMT transformer. It is plausible that to reduce the computational power needed for training, parameter sharing will be taken advantage of, as described by Takase and Kiyono [9].

VI. SUMMARY

The goal of the study is to acquire a better understanding of what might be a state-of-the-art approach to translating into Polish in domain-specific scenarios. The experiment will also bring new knowledge about the expectable quality of translation into a language with robust inflection. The low latency of the method that will be used gives a premise, that the model will be feasible to apply in production environments.

REFERENCES

- [1] Georgiana Dinu, Prashant Mathur, Marcello Federico, and Yaser Al-Onaizan. Training neural machine translation to apply terminology constraints. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3063–3068, Florence, Italy, July 2019. Association for Computational Linguistics.
- [2] Jassem Nowakowski. Neural machine translation with inflected lexicon. 2021.
- [3] Peter Anderson, Basura Fernando, Mark Johnson, and Stephen Gould. Guided open vocabulary image captioning with constrained beam search. *CoRR*, abs/1612.00576, 2016.
- [4] Chris Hokamp and Qun Liu. Lexically constrained decoding for sequence generation using grid beam search. *ArXiv*, abs/1704.07138, 2017.
- [5] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics.
- [6] Miriam Exel, Bianka Buschbeck, Lauritz Brandt, and Simona Doneva. Terminology-constrained neural machine translation at SAP. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, pages 271–280, Lisboa, Portugal, November 2020. European Association for Machine Translation.
- [7] Gema Ramírez-Sánchez, Jaume Zaragoza-Bernabeu, Marta Bañón, and Sergio Ortiz Rojas. Bifixer and bicleaner: two open-source tools to clean your parallel data. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, pages 291–298, Lisboa, Portugal, November 2020. European Association for Machine Translation.
- [8] Robert Patterson. Compendium of accounting in polish & english. 2015.
- [9] Sho Takase and Shun Kiyono. Lessons on parameter sharing across layers in transformers. *ArXiv*, abs/2104.06022, 2021.