

Is Machine Translation Ready for Literature?

Antonio Toral

Dublin City University

Andy Way

Dublin City University

ABSTRACT

Given the current maturity of Machine Translation (MT), demonstrated by its growing adoption by industry (where it is mainly used to assist with the translation of technical documentation), we believe now is the time to assess the extent to which MT is useful to assist with translating literary text. Our empirical methodology relies on the fact that the applicability of MT to a given type of text can be assessed by analysing parallel corpora of that particular type and measuring (i) the degree of freedom of the translations (how literal are the translations) and (ii) the narrowness of the domain (how specific or general that text is). Hence, we tackle the problem of measuring the *translatability* of literary text by comparing the degree of freedom of translation and domain narrowness for such texts to texts in two other domains which have been widely studied in the area of MT: technical documentation and news. Moreover, we present a pilot study on MT for literary text where we translate a novel between two Romance languages. The automatic evaluation results (66.2 BLEU points and 23.2 TER points) would be considered, in an industrial setting, as extremely useful for assisting human translation.

1. Introduction

The field of Machine Translation (MT) has evolved very rapidly since the emergence of statistical approaches two decades ago (Brown et al., 1993). MT is nowadays a reality throughout the industry, which continues to adopting this technology as it results in improved translation productivity, at least for technical domains (Plitt and Masselot, 2010).

Having reached this level of maturity, we explore the viability of current state-of-the-art MT for literature, the last bastion of human translation. To what extent is MT useful for literature? At first glance, these two terms (MT and literature) might seem incompatible, but the truth is – to the best of our knowledge – that the applicability of MT to literature has not been studied rigorously from an empirical point of view.

2. Background

The first work on MT for literature we are aware of (Genzel et al., 2010) translates poetry by constraining a SMT system to produce translations that obey to particular length, meter and rhyming constraints. Form is preserved at the price of producing a worse translation. However, this work does not study the viability of MT to assist with the translation of poetry.

The only other work on MT for literature we are aware of (Besacier, 2014) presents a pilot study where MT followed by post-editing is used to translate a short story from English to French. Post-editing is performed by non-professional translators and the author concludes that this pipeline can be useful as a low cost alternative to translate literary works to a broad number of languages at the expense of sacrificing translation quality.

3. Methodology

The applicability of statistical MT (SMT) to translate a given type of text for a given pair of languages can be studied by analysing two properties of the relevant parallel data.

1. Degree of freedom of the translation. While literal translations can be learnt reasonably well by the word alignment component of SMT, free translations result in problematic alignments.
2. Narrowness of the domain. Constrained domains lead to good SMT results. This is due to the fact that in narrow domains lexical selection is not really an issue and relevant terms occur frequently, which allows the SMT model to learn their translations accurately.

We conclude that the narrower the domain and the smaller the degree of freedom of the translation, the more applicable SMT is. This is why SMT performs well on technical documentation while results are substantially worse for more open and unpredictable domains such as news (cf. WMT translation task series¹).

We suggest to study the applicability of SMT to literary text by comparing the degree of freedom and narrowness of parallel corpora for literature to other domains widely studied in the area of MT (technical documentation and news). Such a corpus study can be carried out by using a set of automatic measures. The degree of freedom of the translation can be approximated by the perplexity of the word alignment. The narrowness of the domain can be assessed by using measures such as repetition rate (Bertoldi et al., 2013) and perplexity with respect to a language model (Ruiz and Federico, 2014).

Therefore, in order to assess the *translatability* of literary text with MT, we put the problem in perspective by comparing it to the *translatability* of other widely studied types of text. Instead of considering the *translatability* of literature as a whole, we root the study along two axes:

1. *Relatedness of the language pair*: from pairs of languages that belong to the same family (e.g. Romance languages), through languages that belong to the same group (e.g. Romance and Germanic languages of the Indo-European group) to unrelated languages (e.g. Germanic and Sino-Tibetan languages).
2. *Literary genre*: from novels to poetry.

We hypothesise that the degree of applicability of SMT to literature depends on these two axes. Between related languages, translations should be more literal and complex phenomena (e.g. metaphors) might simply transfer to the target language, while they might have more

¹<http://www.statmt.org/wmt14/translation-task.html>

complex translations between unrelated languages. Regarding literary genres, in poetry the preservation of form might be considered relevant while in novels it may not.

As a preliminary study, we evaluated the translation of a recent best-selling novel for a related language pair (Spanish to Catalan). The scores obtained – 66.2 BLEU (Papineni et al., 2002) points and 23.2 TER (Snover et al., 2006) points – would be considered, in an industrial setting, as very useful for assisting human translation (e.g. by means of post-editing or interactive MT). We expect these scores to generalise to other related language pairs such as Spanish–Portuguese or Spanish–Italian.²

4. Conclusion

In summary, we have proposed a methodology to assess the applicability of MT to literature which aims to give an indication of how well SMT could be expected to perform on literary texts compared to the performance of this technology on technical documentation and news. While we may be far from having MT that is useful to assist with the translation of poetry between distant languages such as English and Chinese, we have provided evidence that state-of-the-art MT can already be useful to assist with the translation of novels between related languages.

References

- Bertoldi, N., Cettolo, M. and Federico, M. 2013. Cache-based Online Adaptation for Machine Translation Enhanced Computer Assisted Translation. In Proceedings of the XIV Machine Translation Summit, pp. 35–42. Nice, France.
- Besacier, L. 2014. Traduction automatisée d'une oeuvre littéraire: une étude pilote. *Traitement Automatique du Langage Naturel (TALN)*. Marseille, France
- Brown, P. F., Pietra, S. A. D., Pietra, V. J. D., and Mercer, R. L. 1993. The mathematics of statistical machine translation. *Computational Linguistics*, 19(2), pp. 263–313.
- Genzel, D., Uszkoreit, J., Och, F. 2010. Poetic statistical machine translation: rhyme and meter. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pp. 158–166. Association for Computational Linguistics.
- Papineni, K., Roukos, S., Ward, T., and Zhu, W. 2002. BLEU: A method for automatic evaluation of Machine Translation. In 40th Annual Meeting of the Association of Computational Linguistics, pp. 311–318. Philadelphia, PA, USA.
- Plitt, M., and Masselot, F. 2010. A productivity test of statistical machine translation post-editing in a typical localisation context. *The Prague Bulletin of Mathematical Linguistics* 93, pp. 7–16.
- Ruiz, N., and Federico, M. 2014. Complexity of Spoken versus Written Language for Machine Translation. In Proceedings of the 17th Annual Conference of the European Association for Machine Translation, pp. 173–180. Dubrovnik, Croatia.
- Snover, M., Dorr, B., Schwartz, R., Micciulla, L., and Makhoul, J. 2006. A Study of Translation Edit Rate with Targeted Human Annotation. In Proceedings of the Conference of the Association for Machine Translation in the Americas Conference, pp. 223–231.

²The lexical similarity between Spanish and Catalan (0.85) is close to that between Spanish and Italian (0.82) and Spanish and Portuguese (0.89). http://en.wikipedia.org/wiki/Lexical_similarity