Improving translator productivity with MT: a patent translation case study

John Tinsley   john@iconictranslation .com

Iconic Translation Machines Ltd. Invent DCU, Glasnevin, Dublin 9, Ireland

Abstract

When evaluating the suitability of MT for post-editing, there are a lot of variables to consider that could have an impact on its overall effectiveness, including: the languages in question, the documents being translated, the translator workflow, and not least the individual translators themselves. Add to that the various ways we can carry out evaluation - automatic measures, subjective assessments of fluency and adequacy, etc. - and we have got a lot of data on our hands. Despite all of these data points, we are ultimately trying to answer a simple question: is the MT useful and to what extent? In this article, we introduce Iconic Translation Machines and describe a case study on a large-scale post-editing evaluation involving more than 20 translators working on Chinese to English patent translation. We discuss how various evaluations were carried out – from initial MT engine development to translator productivity – and discuss the implications of these findings on the real-world application of MT.

1. Introduction

Iconic Translation Machines (Iconic) is a machine translation software and service provider that specialises in domain-adapted MT. We focus on developing MT engines for specialist content types that are particularly challenging for translation and require more than a pure data-driven approach. Our flagship solution, IPTranslator, was an MT service adapted specifically for patent and intellectual property-related content, and in additional to this, Iconic now develops domain-adapted engines in the areas of finance, life-sciences, and e-commerce.

2. Machine Translation with Subject Matter Expertise

Iconic's approach to MT extends on the classic data-driven approach of statistical MT by incorporating aspects of syntax-based MT and rule-based MT. These approaches are combined with domain-specific processes that have been developed to address particular stylistic conventions of different content types, such as extremely long sentences with complex alpha-number sequences in patent data. Distinct processes, addressing the language, domain, and style, are combined together in Iconic's Ensemble Architecture^{TM} which selects the most effective combination of processes for a particular input type at runtime.

3. RWS : A Case Study

RWS Group is a world-leading language service provider that specialises in patent translation. Iconic worked with RWS to customise a domain-adapted MT engine for Chinese to English

translation. The goal of using MT at RWS was to improve the productivity of translators through post-editing and this was taken into account during the development and automatic evaluation of the engines.

Iconic used its existing baseline for Chinese—English and supplemented it with additional in-domain data suited to RWS content. This adaptation achieved significant improvements in BLEU and TER scores as shown in Figure 1.
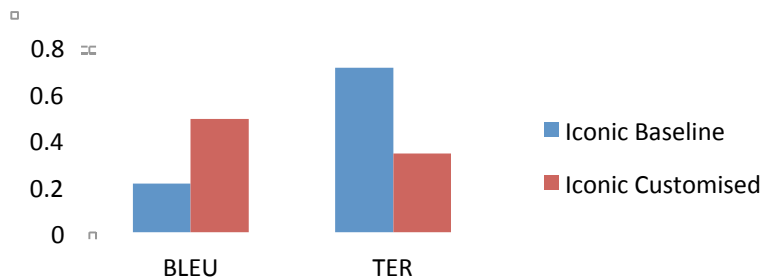

Figure 1 Improvements over the baseline

The TER (translation edit rate) scores give a good indication that the MT output is of sufficient quality to facilitate faster post-editing. In the next section, we describe the evaluations we carried out in order to validate and quantify this in a practical, real-world scenario with professional translators.

3.1. Evaluation Setup: mitigating the variables

In order to quantify the increase in productivity of translators post-editing MT output as opposed to translating from scratch, we carried out an evaluation using the TAUS Dynamic Quality Framework (DQF)[1]. Using this tool, translators post-edit and translate (from scratch) alternating segments in a given test document and the amount of time spent on each segment is measured. The total time spent post-editing vs. translating from scratch is calculated, allowing us to calculate the percentage increase in speed between the two tasks (with the assumption that post-editing will be faster).

There are a number of variables in such an evaluation that could have an impact on the veracity of the results. These variables, and the steps we have taken in our evaluation setup in order to mitigate their impact, are show below:

- **Translator attitude towards MT:** Translators may have a certain bias as relates to MT. We used 24 translators for this evaluation to reduce potential noise from outlying results.
- **Lack of familiarity with the task:** Translators were provided with written instructions, an instructional video, and the opportunity to test the tool prior to beginning the task.
- **Ability of the translator:** Some translators may adapt faster to the task of post-editing than others. We used translators with varying levels of experience to reduce this effect.
- **Difficulty of the test set:** A particularly challenging test set for MT/translation could produce skewed results. We used four different test sets to avoid this.

Each translator translated more than 200 segments each and the findings are presented in the next section.

---

[1] https://evaluate.taus.net/evaluate/dqf/dynamic-quality-framework

3.2. Results : a resounding success

The main finding from the practical evaluations was that 83% of the translators were faster when post-editing MT output for Chinese to English translation. Figure 2 below shows that almost 40% of translators were more than 30% faster, while only 4 translators were slower while post-editing.

Productivity Range	#Translators	% of total
> 30%	9	39%
20-29%	3	13%
10-19%	3	13%
1-9%	4	17%
< 0%	4	17%

Figure 2 Range of improvement across translators

Post-editing was shown to be effective for both experienced and inexperienced translators, and consistent improvements were seen across all test sets.

4. Conclusions

The concept of a one-size-fits-all solution for MT is far-fetched. In order to achieve good quality results, especially for difficult languages and domains, engines must be highly tuned with a number of syntactic and content-specific processes, on top of a baseline SMT architecture.

We have demonstrated the effectiveness of this approach in improving machine translation quality for Chinese to English patent translation. The quality and *usability* of the MT output was validated in translation evaluations which showed a significant increase in the productivity of translators who were post-editing the MT output in a practical translation scenario.