Machine Translation Technology in Advanced Language Teaching and Translator Training: A Corpus-based Approach to Post-editing MT Output

Bogdan BABYCH, Anne BUCKLEY, Richard HUGHES, Svitlana BABYCH
Centre for Translation Studies
University of Leeds
Leeds, UK
{b.babych, a.buckley, r.j.hughes, s.babych}@leeds.ac.uk

1. Introduction

At advanced stages of language learning students are refining their skills in producing idiomatic and stylistically appropriate texts, with acceptable usage patterns, conventional terminology and a broad range of lexical resources, such as synonyms, collocations and complex lexico-grammatical constructions. Specifically, this level of proficiency is particularly important in translator training (Kuebler, 2011; Aston, 1999).

Creating or finding teaching materials and exercises for advanced learners is a challenging task for language tutors, because at this level traditional textbooks are not sufficient. Tutors usually rely on authentic texts and give students feedback on their productive skills. However, there are a number of practical and ethical issues which limit the use of this critical feedback: (a) it needs to be confidential: usually students can access only their own corrected work, which makes it difficult to discuss individual students' errors with the whole class; (b) students have limited possibilities to distance themselves from the text they produce and review the errors critically.

In our paper we explore techniques for developing non-native linguistic intuition via post-editing of imperfect Machine Translation (MT) output, using concordance, collocation and terminological searches in large-scale monolingual and bilingual corpora with the goal of arriving at a more fluent and coherent text. The texts, originally written in English, are automatically translated into another language such as German or Russian with an MT system (e.g. Systran) and then back-translated with MT into English. Alternatively we use a human translation of a text as input for MT, and use the text originally written in English as a reference.

We use our methodology during the one-semester module *English for Translators* taught at the Centre for Translation Studies of the University of Leeds, as part of the Leeds MA programme in Applied Translation Studies. The module covers a range of text types and genres (journalistic, administrative, technical and literary) – specifically those that are discussed within other modules on Specialised Translation within the programme. It is taught to non-native speakers of English from a variety of different language backgrounds and focuses on specific aspects relevant to professional translator training, such as developing contrastive knowledge of source and target language systems, legitimate variation in translation, universal and language-specific aspects of text typology, and specifics of translation into the non-native language. However, the methodology can be applied in a wider range of contexts at advanced stages of language learning, where students take control over their learning process and develop their productive skills in autonomous learning.

Machine Translation technology was originally designed to generate rough translation for people who do not understand the source language, or to produce editable draft for bilingual professional translators (Hutchins & Somers, 1992). It has been suggested that there are similarities between errors made by MT systems and those made by non-native speakers (Lee et al, 2007). For our purposes it is important that MT output contains the original message, but with its fluency disrupted on the lexical, collocational or stylistic levels. The advantage of using MT in the classroom is that students can critically review the MT output, discuss potential solutions in a group with the tutor, and check their decisions by doing corpus-based research. The learning objective is to match the construction, lexicon, terminology and stylistic resources of the reference text or come up with alternative acceptable solutions, using our corpus-based tools built around the Corpus Workbench search engine (Christ, 1994; Evert & Hardy, 2011). These tools include monolingual and bilingual concordances, terminological and collocation searches in large corpora (between 100 million and 1 billion words). Students explore usage patterns of specific linguistic expressions beyond the given text, try to find appropriate solutions to non-trivial editing problems, and then discuss their findings with the tutors and other native speakers. This type of work develops language skills at an advanced level via discovery learning and promoting learner autonomy. This gives the students more confidence in their ability to identify non-fluent phrases and find good near-native-speaker expressions for correcting these dysfluencies.

In the paper we explore the advantages of using post-editing of machine translation output combined with corpus-based translator training in the classroom for advanced language learners. We propose this new type of task for translator training and describe how the MT technology has to be adapted to better meet the requirements of this task, by being embedded into corpus-based resources and offering alternative translations for segments. For this purpose having non-idiomatic output is surprisingly an advantage, not a shortcoming, of the MT system.

2. Context of our research

When learners of a foreign language acquire productive skills (speaking and writing) they need to be able to reflect on their performance, identify and critically analyse errors, and come up with better suggestions acceptable by native speakers. This critical approach forms an important part of speakers' linguistic intuition in a foreign language. In developing this intuition students initially rely on tutors' feedback for identifying their errors (*negative linguistic evidence*) and producing acceptable alternative suggestions for expressing their intended propositions. Later on in the learning process they gradually become more aware of potentially problematic expressions and independently correct them into acceptable variants, increasingly using input from listening and reading (*positive linguistic evidence*). However, availability of the negative linguistic evidence still remains an important clue for all stages of foreign language learning. In advanced stages, for example, such negative evidence is crucial to verify and refine usage patterns for complex and infrequent linguistic constructions.

The availability and importance of negative linguistic evidence for language learning has been discussed in a wide range of contexts, e.g., studies of L1 learning in the context of Chomsky's poverty of stimulus hypothesis and further experimental attempts to verify it (e.g. Chouinard & Clark, 2003), or in the context of machine learning – computational models of

language acquisition (e.g. Lewis & Elman, 2001) – and consistently demonstrates that negative feedback is essential for successful language learning. Similar results are reported in experimental studies of L2 learning, e.g., White (1991) shows that including negative evidence is more effective for learning the correct position of English adverbs than positive input alone; Aljaafreh and Lantolf (1994) analyse negative feedback as regulation within an experimental framework and arrive at similar conclusions.

Even though showing that ill-formed L2 can be counterproductive in the initial stages of foreign language learning (Somers, 2004), in the advanced stages (and especially for professional translator training), negative linguistic evidence becomes very useful, since students are already aware of contrastive differences between languages and consciously take control over developing their productive skills in autonomous learning.

However, obtaining negative linguistic evidence is relatively 'expensive' for language learners for several practical reasons: mainly because it is usually given in classroom situations by language tutors, and the amount of these contact hours is necessarily limited. In autonomous learning or in everyday situations outside the classroom it is relatively hard for language learners to obtain negative linguistic evidence, even if they have sufficient contact with native speakers. This is mainly due to the fact that correcting non-native errors is not perceived as a natural way of communication in most cultures, and even though native speakers can easily identify these errors, they are usually too embarrassed to correct non-native speakers. In addition, the number of topics and genres which can be discussed (and corrected) in the classroom cannot possibly cover the complete range of topics present in real-world communication. Corrected written assignments also cannot cover all text types that language learners will produce in practical situations. As a result, some errors that are specific to certain text types and genres may remain uncorrected for long periods of time, and non-native speakers will continue to make these errors even when they achieve relatively high language proficiency and start to use language for professional communication. Another practical limitation in using critical feedback (mentioned in Section 1) is the need to keep individual students' feedback confidential, so the opportunity to critically analyse feedback given to others is restricted.

Because of this, the amount of useful negative evidence is inherently limited by the number of texts produced by language learners and assessed by tutors: here students do not have any chance to actively engage with the body of negative data generated in a systematic way by some external source and arrive at systematic ways of analysing and correcting the errors and developing appropriate linguistic intuitions for speaking and writing.

Another limitation of traditional approaches to giving negative linguistic evidence to groups of students is that they typically discuss errors in individual sentences without looking into the wider context. Discourse-level errors, such as lexical choices that rely on a wider context beyond the sentence level, can easily be overlooked in this approach.

Language learners would benefit from a methodology of systematically generating negative linguistic evidence and critically analysing and correcting this data for in-class analysis with tutors and for autonomous learning. This linguistic evidence needs to be generated both for potential sentence-level and text-level errors and to inform learners' correction decisions in wider discourse-level contexts.

3. Methodology

In this paper we propose to address these problems by combining traditional in-class discussion of negative linguistic evidence with two state-of-the art technologies: Machine Translation and concordance and collocation-based searches in large linguistically annotated corpora. An overview of previous work in the area of using MT for language learning, as well as useful suggestions for work in this area, is given in Somers (2004). Compared to existing approaches our methodology develops these ideas in two main directions: (1) it combines post-editing and corpus-based searches for idiomatic solutions in the framework of autonomous language learning at advanced stages, and (2) it focuses on learners' autonomy and addresses the specific needs of professional translator training and advanced language learners, as part of an MA programme in Applied Translation Studies.

Modern Machine Translation systems (such as Systran, Lucy, Reverso, Lingenio MT, Google Translate, Language Weaver) in many cases produce understandable sentences and phrases that preserve most of the intended propositional content, but are usually non-idiomatic and/or linguistically wrong.

Such imperfections in output are often perceived as a weakness of MT systems. However, a variety of legitimate uses of imperfect MT have been identified. If MT is used for *assimilation* (understanding texts written in an unknown source language), e.g., in order to extract specific information, users can tolerate dysfluencies that do not distort comprehension and the original meaning of the text (Hartley, 2008: 121). If MT is used for *dissemination* (publication produced by professional translators with good knowledge of the source language), the study by Plitt & Masselot (2010) reports that post-editing increases translators' productivity for certain combinations of translation directions, subject domains and text types by 74% on average, although this figure varies widely between different translators (between 20% and 131%) depending on their attitude and experience with post-editing MT-generated texts.

The idea behind our methodology is that we can create negative linguistic evidence for foreign language learners in a systematic way using MT-generated sentences that preserve propositional content but are linguistically wrong, unidiomatic, non-fluent or stylistically unacceptable. This material can be further integrated into classroom-based exercises and homework assignments with the goal of developing learners' linguistic intuition in a foreign language. These types of task involve identification of problematic expressions; detection of syntactic, lexical and stylistic problems; and finding acceptable solutions using concordance and collocation searches in large linguistically annotated corpora.

We cover the same range of text types that our students translate in their Specialised Translation modules: journalistic, administrative, technical, literal and academic. To generate texts for correction we typically use two-stage MT: we translate an original English text into another language, e.g., German or Russian, back-translate it into English, manually check for possible content omissions or potential incomprehensibility (we refer to such cases as 'content black holes') and finally – reconstruct the original intended meaning of the sentences and phrases which became unclear during the MT phase. An alternative way of generating content is to take an original text written in a foreign language and to translate it into English via a single-stage MT. Since the majority of our texts need to be based upon authentic British content, we normally use two-stage MT.

For creation of negative linguistic evidence we tested two MT systems: the statistical MT (SMT) system Google Translate, which was automatically built from a large collection of human translations, and the rule-based MT (RBMT) system Systran that uses hand-crafted rules for source text analysis, cross-lingual transfer and target text generation. Each of the systems has its advantages for different usage scenarios and translation directions (Callison-Burch et al., 2009) but in our experiments RBMT architecture showed a clear advantage: it produces translation that is more faithful to the original and less fluent or idiomatic linguistically. Such dysfluencies are traditionally seen as a shortcoming of the rule-based systems, but become a surprising advantage for the purposes of generating negative linguistic evidence. On the other hand, SMT systems usually generate much more fluent and natural sounding output, but occasionally mistranslate or omit important lexical items. Google output was often found linguistically 'too perfect' to be used as a source of negative linguistic evidence, but more often produced 'content black holes', which had to be corrected manually. As a result, we chose the Systran RBMT system for generating negative linguistic evidence for language learners. We use correction exercises as part of non-assessed homework, for classroom discussion in groups, and also as part of the final module assessment.

The types of negative linguistic evidence found in the output of MT systems in many respects resemble the errors of non-native speakers. Similar methods can be used for automatically identifying imperfect MT output and errors made by language learners (Lee et al, 2007). Certain error types, such as collocation and word order errors, are common for language learners and MT (especially RBMT) systems. There are also important differences between language learners and MT systems. Non-native speakers more often make errors on the local level (agreement, choices of correct word senses), while MT errors frequently happen on a more global textual level and concern cohesion and coherence, non-local dependencies, co-reference of pronouns, etc. However, our approach does not need to exactly replicate MT error types made by foreign language learners. Even if MT creates negative linguistic evidence of a different type, in combination with corpus-based approaches it is still useful for developing linguistic intuition for non-native language learners. If an MT-generated text is not useful for language learners, tutors can modify it to better address certain language issues important for learners. This post-editing is still more efficient than generating negative evidence from learners' materials (e.g., from written assignments).

4. Translation exercises and corpus-based searches

In this section we present an end-to-end example of how language learners post-edit MT output and verify their solutions using concordance and collocation searches as homework. Further, in class they propose critical feedback on MT errors and discuss their suggestions with tutors. An example of an administrative text (in MT-translated and original versions) and selected students' suggestions are shown in Figure 1.

MT: Child poverty act of 2010

Original text:
Child Poverty Act 2010

MT: Goals regarding children poverty

Original text:
Targets relating to child poverty

Suggestions:

1. Targets regarding child poverty

2. Child poverty targets

MT: (1) must <u>the Secretary of State</u>, as soon as sensibly feasible after the end of <u>the 2010 target year</u> and on any case not later that 30 June 2012, situate before the parliament a report on whether the goal was hit 2010.

Original text:
(1) The Secretary of State must, as soon as reasonably practicable after the end of the 2010 target year and in any event not later than 30 June 2012, lay before Parliament a report on whether the 2010 target has been met.

Students' suggestions:

1. (1) The Secretary of State must submit a report on whether the target has been met to the parliament once sensibly feasible after the end of the 2010 target year and on any case no later than 30 June 2012.

2. (1) After the end of the 2010 target year, the Secretary of State must submit as soon as possible a sensibly feasible report to the parliament on whether the goal set in 2010 has been achieved no later than 30 June 2012.

MT: (2) Are the goal 2010 at <u>the financial year</u> beginning with 1 April 2010: 1,7 million children or fewer sit in qualifying households in the united kingdom, which fell within <u>the relevant income group</u> for the purpose of section 3 (<u>the relative low income target</u>).

Original text:
(2) The 2010 target is that in the financial year beginning with 1 April 2010, 1.7 million children or fewer live in qualifying households in the United Kingdom that fell within the relevant income group for the purposes of section 3 (the relative low income target)

Students' suggestions:

1. The 2010 target is that in the financial year beginning with 1 April 2010, there should be no more than 1.7 million children in qualifying households in the United Kingdom, who fall within the relevant low income group for the purpose of section 3 (the relative low income target).

2. (2) Target for financial year 2010 beginning as of April 1st, 2010: For the purpose of Section 3, at least 1.7 million children from relative low-income households in the UK shall be helped (See Section 3 for 'Relative Low Income Target');

Figure 1. *MT-translated text, original text and students' suggestions discussed in class.*

Students are given MT output as homework, and they are asked to identify phrases that do not work in English and suggest alternative acceptable solutions initially without any help from native speakers. However, they are encouraged to use corpus-based resources and run concordance and collocation-based searches in two interfaces to large monolingual and parallel corpora developed at the University of Leeds: CSAR and IntelliText (http://corpus.leeds.ac.uk/internet.html; http://corpus.leeds.ac.uk/itweb/).

Some of the concordance and collocation searches helpful in addressing some of the problems identified in Figure 1 are shown in Figure 2. These searches test which verbs can go in the context *before .. parliament*, and allow students to successfully identify the idiomatic administrative term *lay before parliament*.

Query translation in IntelliText interface:

```
any word with part of speech tag "V.*" then up to 2 words of any type then word
"before" thenup to 2 words of any type then word "parliament"
```

Queries (IntelliText and CSAR):

```
[pos="V.*"] [word!="\W"]{0,2} [word="before"] [word!="\W"]{0,2} [word="parliament"]
(meet (meet [pos="V.*"] [word="before"] -0 3) [word="parliament"] -3 6)
```

Concordance:

publication of draft orders to	be **laid** before parliament	were awaited. QEB considered a
and therefore did not need to	defend its decisions before parliament	. The idea of presenting a yearly
embodied in the Education Reform Bill	**placed** before parliament	in November 1987. While these
companies since the regulations	were first **laid** before parliament	in 2002, has the power to stop
no such plan exists. Proposals	**put** before parliament	yesterday would make EDF and GDF
Labour - the five-year plan to	be **put** before parliament	this week by the Education Secretary
Party leader said his party would	**put** new legislation before the Berlin parliament	this autumn seeking approval for

Collocations:

```
MU (meet [word="before"] [word="parliament"] -0 3)
Colloc: left=3, right=0; Filter: V*
```

Collocation	Joint	Freq1	Freq2	LL score	Concordance
before parliament ~~ come	9	1643900		10.65	Examples
before parliament ~~ go	9	2365620		9.09	Examples
before parliament ~~ put	5	763486		6.34	Examples
before parliament ~~ legislate	1	4391		3.03	Examples
before the ~~ go	4	2365620		2.53	Examples
before parliament ~~ appear	2	491369		2.08	Examples
before the ~~ come	3	1643900		2.00	Examples
before his ~~ promise	1	129252		1.35	Examples
before parliament ~~ lay	1	134237		1.33	Examples

Figure 2. *Concordance and collocation search results for correcting 'situate before parliament'.*

Students then submit their suggestions on-line via the VLE (Virtual Learning Environment), and in the next class we analyse their solutions and discuss them, showing the original English variant and analysing alternative acceptable variants and suggested expressions which do not work in English.

5. Conclusions and future work

In this paper we have presented a methodology that combines generation of negative linguistic evidence for language learners using MT output, independent corpus-based searches and critical feedback and discussion of the errors with tutors. This methodology allows tutors to generate linguistic evidence systematically and develop learners' intuitions in the non-native language. Future work will include formal evaluation of the methodology in classroom settings and collaborative work with developers of MT to better adapt their rule-based systems to the needs of language learners.

References

Aljaafreh, A. & Lantolf, J. P. (1994). 'Negative Feedback as Regulation and Second Language Learning in the Zone of Proximal Development'. *The Modern Language Journal*, 78 (4). 465–483.

Aston, G. (1999). 'Corpus use and learning to translate'. *Textus*, 12. 289-313.

Callison-Burch, C., Koehn, P., Monz, C. & Schroeder, J. (2009). EuroMatrix1.2b: Evaluation Campaign. Available from: http://euromatrix.net/deliverables/deliverable-1-2b.pdf [29/05/2012].

Chouinard, M. & Clark, E. (2003). 'Adult reformulations of child errors as negative evidence'. *Journal of Child Language*, 30. 637-669.

Christ, O. (1994). A modular and flexible architecture for an integrated corpus query system. COMPLEX '94, Budapest.

Evert, S. & Hardie, A. (2011). 'Twenty-first century Corpus Workbench: Updating a query architecture for the new millennium'. In: *Proc. of Corpus Linguistics 2011*, University of Birmingham, UK.

Hartley, A. (2008). 'Technology and translation'. In: Munday, J (ed.) (2008). *The Routledge Companion to Translation Studies*. Routledge. 106-127.

Hutchins, W.J. & Somers, H.L. (1992). *An introduction to machine translation*. London: Academic Press, 1992.

Kuebler, N. (2011). 'Working with corpora for translation teaching in a French-speaking setting'. In: Frankenberg-Garcia, A., Aston, G. & Flowerdew, L. (eds.) (2011). *New Trends in Corpora and Language Learning (Corpus and Discourse)*. Continuum. 62-79.

Lee, J., Zhou, M. & Liu, X. (2007). 'Detection of non-native sentences using machine-translated training data'. In: *Proc. of NAACL HLT 2007*, Companion Volume. 93–96.

Lewis, J. D. & Elman, J. L. (2001). 'Learnability and the Statistical Structure of Language: Poverty of Stimulus Arguments Revisited'. In: *Proc. of the 26th Annual Boston University Conference on Language Development*.

Plitt, M. & Masselot, F. (2010). 'A Productivity Test of Statistical Machine Translation: Post-Editing in a Typical Localisation Context'. In: *The Prague Bulletin of Mathematical Linguistics*, 93. 7–16.

Somers, H. (2004). 'Does machine translation have a role in language learning?' *UNTELE 2004: L'Autonomie de l'Enseignantet de l'Apprenant face aux Technologies de l'Informationet de la Communication*, Compiegne, France.

White, L. (1991). 'Adverb placement in second language acquisition: some effects of positive and negative evidence in the classroom'. *Second Language Research*, 7(2). 133-161.