Alignment-Based Tools for Translation Self-Learning

J. Gabriel Pereira Lopes and Marcelo S. Pimenta

Tiago Ildefonso

CITI/DI/FCT/Universidade Nova de Lisboa, Portugal

marcelo.pimenta@inf.ufrgs.br

Abstract: Effective use of translation products and services depends on professional and learner translators familiarity with different translation procedures, strategies and tools, currently available for solving specific translation problems. In this poster we will focus on the use of language independent alignment-based technology, applied to parallel corpora, for effectively aiding translation learning and translation quality improvement, both at-work and at-the-school.

1. Introduction

Translation has undergone considerable changes over the last decade. Translation technology has gained wider acceptance and is currently used by professional translators. Computer-aided translation (CAT) tools entered translation services due to their contribution for improving translators productivity and translation quality. Information Retrieval technology has also affected translation research, as witnessed by the rising interest in corpus-based approaches to translation studies. Such changes have brought new opportunities and new challenges both for translation and for translators training. In this paper, we will show that CAT tools, that have already proved to be effective in providing linguistic help in translation situations, can also effectively support translation self-instruction at-the-work and at-the-school.

Despite the considerable scientific production on CAT, surprisingly, there is little research work devoted to translation learning and to the use of CAT tools in the classroom. But there are exceptions. Somers (2001) discusses strategies for Machine Translation (MT) teaching. Balkan et al. (1997) surveyed tools and techniques for MT teaching. Kenny and Way (2001) report on experiences in MT and translation technologies teaching. Fictumová (2004) discusses the use of the open source e-learning environment LMS Moodle for translators and interpreters. All of them agree on the need to bring translation instruction closer to the real world of professional translation in order to cope with imminent changes and challenges in the translation market. The same applies to the need to learn how to use translation tools (Somers, 2003; Forcada, 2000). However, material for self-learning, for supporting autodidactic translation students and for training professional translators, is rare.

In this poster we will focus on autonomous self-learning of translation, both at-work and at-the-school. For this purpose we will use parallel text alignment-based tools produced in the framework of TRADAUT-PT, a MLIS European project, for supporting lexicon coders to speed up the construction of lexical entries for the various kinds of lexicons used by Systran MT systems. Those tools include a language independent aligner (Ribeiro et al, 2000; Ildefonso et al, 2005), a bilingual concordancer for each pair of languages considered, monolingual concordancers, an extractor of single word and multi-word translations, a web interface for enabling different clients to work, safely and independently, at-the-distance, using these tools and MT systems. As our aligner is language independent and partitions

1 This work has been granted by CAPES/GRICES Brazilian-Portuguese joint project We-Learn.
input parallel texts into parallel segments (text stretches that are translations of each other or a common source text stretch) finer grained than the sentence, it can also align human made and golden standard human translations (figure 1).

2. Using our alignment-based set of tools for translation instruction

In this poster, we will focus on a specific use of our aligner for translation quality evaluation and on the use of our bilingual concordancer for helping the learner to check why her/his translation failed to achieve golden standard quality.

In general, the closer a translation and its golden standard translation are, the better that translation is. As a given sentence may be translated in many ways, a student need to know why his/her own translation diverges from a golden standard. Let us see how this can be achieved using our tools. Consider the excerpt of the translation made by Systran MT system, from Portuguese into English, for the European Council Regulation (EC) No 1239/98 stretch, represented in figure 1, at the right side of the two screen dumps captured. At the left side of those screen dumps it is depicted the aligned golden standard.

```
the 2 the
quantity 9 quantities
of 4 of
each 5 each
species 6 species
cought 7 captured
during 0 during
each 9 each
fishing 10
operation 1 operation
including 12 including
```

Used Golden standard is: “the quantity of each species caught during each fishing operation, including by-catches and discards at sea, in particular cetaceans, reptiles and sea-birds,”. Corresponding Portuguese source text is: “as quantidades de cada espécie capturadas durante cada operação de pesca, incluindo as capturas acessórias e as quantidades devolvidas ao mar, nomeadamente as relativas aos cetáceos, répteis e aves marinhas.”.

In figure 1, first difference is the use of “quantities” instead of “quantity”, for translating “quantidades”. Consulting our bilingual concordancer for “as quantidades de cada espécie captured during each operation”, we get 10 times “the quantity of each species”, 7 times “the quantities of each species”, and twice “the amount of each species”. In this case, we may consider proposed translation as correct. Second difference, related to the use of “captured” instead of “caught”, shows that “harvested” occurs 5 times, “caught” 3 times, “taken” twice, “fished” once, apart from 4 other singular occurrences. Regarding the preference for “fishing operation”, by consulting our bilingual concordancer, we observe that “fishing operation” occurs 8 times and “haul” 7 times. Such observation should lead a student to reflect and, if necessary, to interact with her/his teacher. Fourth difference relates to the literal translation of “quantidades devolvidas ao mar” and the preferred translation “discards at the sea”. Sixth difference relates to the literal translation “as relativas”, not taking into account that, in this case, “as” is a demonstrative pronoun and
so, it might be translated as “those”, giving rise to “namely those concerning cetaceans”. This translation, though different, would be perfectly acceptable. Again plural definite article should be deleted. Last difference, would lead the student to learn, by using the concordancer, that “seabirds” occur 5 times and “sea-birds” twice. Despite these numbers, my spell checker proposes “sea birds” for correcting “seabirds”.

3. Conclusions
We assume that students have a certain competence in the non-native languages they are working on. So, in this paper we focussed on issues related to hands-on translation training. The alignment-based tools we propose help the students to understand basic translation notions (equivalence or lack of it, translatability, loss, compensation, faithfulness, naturalness, form, content, meaning etc.). They also help students to decide, choose and evaluate their own performance during a translation process. Proposed environment additionally introduces students to existing translation technology. Students will draw their conclusions from their practical experience, by finding and solving translation problems and becoming familiar with different translation procedures and strategies available to solve those problems.

Our language independent approach to parallel text alignment gives rise to parallel text segments with a length shorter than the sentence. Moreover, it may be directly applied to the alignment of any translation produced either by a student or by a machine and align it with a golden standard (a parallel text produced by a professional translator). This way, most part of parallel segments will have a length equal to one word. Differences generally occur in larger segments. A traditional aligner, working at the sentence level, would be unable to produce such a low grained alignment, thus making more difficult the visualization of differences. By signalizing differently parallel segments having different content, a student may easily view the differences between a golden standard and his/her own translation.

References