

Translation Class Instruction as Collaboration in the Act of Translation

Lars Ahrenberg
Department of Computer and
Information Science,
Linköping University
lars.ahrenberg@liu.se

Ljuba Tarvi
University of Helsinki
Helsinki, Finland
ljuba.tarvi@welho.com

Abstract

The paper offers an effective way of teacher-student computer-based collaboration in translation class. We show how a quantitative-qualitative method of analysis supported by word alignment technology can be applied to student translations for use in the classroom. The combined use of natural-language processing and manual techniques enables students to ‘co-emerge’ during highly motivated collaborative sessions. Within the advocated approach, students are proactive seekers for a better translation (grade) in a teacher-centered computer-based peer-assisted translation class.

1 Introduction

Tools for computer-assisted translation (CAT), including translation memories, term banks, and more, are nowadays standard tools for translators. The proper use of such tools and resources are also increasingly becoming obligatory parts of translator training. Yet we believe that translation technology has more to offer translator training, in particular as a support for classroom interaction. Our proposal includes a quantitative analysis of translations, supported by word alignment technology, to enable joint presentation, discussion, and assessment of individual student translation in class. For comparisons with related work, see section 4.

From the pedagogical point of view, the suggested procedure embraces at least four types of evaluation: students’ implied self-evaluation, a preliminary computer evaluation, teacher’s evaluation after manually correcting the imperfect computer alignment and assessment, and peer

evaluation during the collaborative team work in class, when the versions produced by the students are simultaneously displayed, discussed and corrected if necessary.

Theoretically, translations are viewed here as mappings between two languages through emergent conceptual spaces based on an intermediate level of representation (e.g., Honkela et al., 2010). In terms of praxis, the basic approach is rooted in the idea (Vinay & Darbelnet, 1958) of consecutive numbering of the tokens (words) in the original text. This simple technique enables - finding and labeling, in accordance with a chosen set of rules, certain isomorphic correspondences between the source and target tokens. Finding such correspondences is what current machine translation approaches attempt to achieve by statistical means in the training phase.

The quantitative-qualitative technique we use here is the Token Equivalence Method (TEM) (Tarvi 2004). The use of the TEM in translation teaching originated as an argument (involving the second author) in a teacher-student debate over the relevance of a grade. The considerable time spent on the manual preparation of texts for translation using the TEM proved to be fairly well compensated for by the evident objectivity of the grades - the argument that, say, only 65% of the original text has been retained in a translation is difficult to brush aside. Later, the method was applied in research. Tarvi (2004) compared the classical Russian novel in verse by A. Pushkin *Eugene Onegin* (1837) with its nineteen English translations. Figures calculated manually on 10% of the text of the novel showed an excellent fit with the results on the same material obtained elsewhere by conventional comparative methods. Thus, we believe that characterizations of relations between source and target texts in ob-

jective terms is a good thing for translation evaluation.

1.1 The TEM: Basics and Example

Methodologically, the TEM focuses not on translation ‘shifts’ but on what has been kept in translation. The basic frame for analysis in the TEM is the Token Frame (2.2.1), which accounts for the number of the original tokens retained in translations. The other four frames (2.2.2-3, 2.3.1-2), although useful in gauging the comparative merits of the translations and the individual strategies, are optional.

To concisely illustrate the method, one sentence will be used – the famous 13-token opening sentence of Leo Tolstoy’s *Anna Karenina*: *Vse schastlivye semyi pohozi drug na druga, kazhdaya neschastlivaya semya neschastliva po svoemu.* (All happy families resemble one another, every unhappy family is unhappy in its own way.)

Eight English translations of this sentence (2.1) will be used for analysis.

The source text and all its translations are tokenized and analyzed linguistically in different ways. NLP tools such as lemmatizers, part-of-speech taggers and parsers can be applied. Most importantly, however, to support the computation of the Token Frames (2.2.1), they must be word-aligned with the source text (2.6). The teacher or the students are expected to review the alignments and correct them if they are not acceptable. Given the corrected alignments, the aligned data can be used by the teacher and the students in the classroom.

After this introduction of the basics of the theoretical approach and relevant automatic methods for their implementation, the paper is built around the basic structural foci of any instruction unit: before class (2), in class (3), and outside class (4).

2 Before class

This section describes the techniques of processing Source Texts (ST) and Target Texts (TT) by teachers and students.

2.1 Token Numbering and Labeling

The procedure starts with consecutive numbering of the original tokens:

(1)Vse (2)schastlivye (3)semyi (4)pohozi (5)drug (6)na (7)druga, (8)kazhdaya (9)neschastlivaya (10)semya (11)neschastliva (12)po (13)svoemu.

The second step is establishing, via the procedure of (corrected) alignment, the correspondences between the Source tokens (St) and Target tokens (Tt). As a result, every corresponding TT token (Tt), if found, is designated with the number of its source counterpart (St). Besides, since no Tt may remain unlabeled, two types of Tts which have no counterparts in the ST are labeled as Extra tokens (2.3.1) and Formal tokens (2.3.2). Here are the eight translations of the example sentence:

Leo Wiener (1899):

(1)All (2)happy (3)families (4)resemble (5-6-7)one another; (8)every (9)unhappy (10)family (Ft)is (11)unhappy (12)in (Ft)its (13)own way.

Constance Garnett (1901):

(2)Happy (3)families (Ft)are (1)all (4)alike; (8)every (9)unhappy (10)family (Ft)is (11)unhappy (12)in (Ft)its (13)own way.

Rochelle S. Townsend (1912):

(1)All (2)happy (3)families (Ft)are (Et)more (Et)or (Et)less (4)like (5-6-7)one another; (8)every (9)unhappy (10)family (Ft)is (11)unhappy (12)in (Ft)its (13)own (Et)particular way.

Aylmer & Louise Maude (1918):

(1)All (2)happy (3)families (4)resemble (5-6-7)one another, (Et)but (8)each (9)unhappy (10)family (Ft)is (11)unhappy (12)in (Ft)its (13)own way.

Rosemary Edmonds (1954):

(1)All (2)happy (3)families (Ft)are (4)alike; (Et)but (Ft)an (9)unhappy (10)family (Ft)is (11)unhappy (12)after (Ft)its (13)own fashion.

Joel Carmichael (1960):

(2)Happy (3)families (Ft)are (1)all (4)alike; (8)every (9)unhappy (10)family (Ft)is (11)unhappy (12)in (Ft)its (13)own way.

David Magarschack (1961):

(1)All (2)happy (3)families (Ft)are (4)like (5-6-7)one another; (8)each (9)unhappy (10)family (Ft)is (11)unhappy (12)in (Ft)its (13)own way.

Richard Pevear & Larisa Volokhonsky (2000):

(1)All (2)happy (3)families (Ft)are (4)alike; (8)each (9)unhappy (10)family (Ft)is (11)unhappy (12)in (Ft)its (13)own way.

As is seen, two of the versions are clones (Carmichael, Pevear-Volokhonsky), one translation (Garnett) differs from the original only by the choice of the adjective (St 8), while the remaining five versions are more diverse. Note the mode of denoting Tts suggested here: only the

meaningful denotative tokens get labeled, e.g., *are* (4)*alike*, or *is* (11)*unhappy*; if not one Tt but a group of tokens is used as an isomorph to a single St, the whole group is underlined, e.g., (13)*own way*, or (13)*own fashion*.

Although St 4 has been rendered as *are alike* (Edmonds, Pevear-Volokhonsky, Garnett, Carmichael), *are like* (Townsend, Magarschack), and *resemble* (Wiener, the Maudes), all these rendering are viewed as retaining the denotative meaning of the original token. Or, for instance, St 12, whether rendered as *after* (Edmonds) or *in* (all the rest), is also viewed as retained in translation. The connotative shades of meaning most suitable for the outlined goals can be discussed in class (3.2).

This mode of displaying the isomorphisms can be converted to the style of representation used in word alignment systems such as Giza++ (Och and Ney, 2003) as follows: Extra tokens and Formal tokens give rise to null links. Groups of tokens that correspond yield groups of links. Thus, the analysis for Wiener's translation would come out as below:

1-1 2-2 3-3 4-4 5-5 5-6 6-5 6-6 7-5 7-6 8-7 9-8
10-9 0-10 11-11 12-12 0-13 13-14 13-15.

In gauging the content, two types of basic and optional analytical frames, content and formal, are used. Based on the way of calculating the results, the analytical frames will be considered here in two sections, percentage frames (2.2) and count frames (2.3).

2.2 The TEM: Percentage Frames

The results in these frames are calculated as percentages of the ST information retained in translations.

2.2.1 Basic Content Frame (Token Frame)

After finding the isomorphic counterparts, the percentages of the retained tokens are presented in Table 1 (column I). As one can see, Wiener, the Maudes, Magarschack and Townsend translated all thirteen tokens and, hence, scored 100% each; Garnett, Carmichael and Pevear-Volokhonsky omitted Sts 5-6-7 and thus scored 76%, while Edmonds left out four tokens, Sts 5-6-7-8, thus retaining 69% of the original.

2.2.2 Optional Formal Frame 1 (Morphology Frame)

In this frame, if a token is rendered with the same part of speech as in the original, the Tt in

question gets a count. As can be seen in Table 1 (column II), only two translators, Wiener and the Maudes, kept the same type of predicate 1 (St 4) as in the original – *resemble* – while in the remaining six translations the type of predicate 1 has been changed into a compound one: *are alike* (Edmonds, Pevear-Volokhonsky, Garnett, Carmichael), and *are like* (Townsend, Magarschack). Therefore, in this frame, Wiener and the Maudes get a 100% each; Edmonds, with her two changed parts of speech, gets 84%, while the remaining five translators, who changed one part of speech each, score 92%.

2.2.3 Optional Formal Frame 2 (Syntax)

Another possible way of gauging the 'presence' of the original in its translation is monitoring the syntactic changes. If at least two tokens are rendered in the same sequence as in the original and preserve the same syntactic functions, they are considered syntactically kept. Non-translated Sts are viewed as non-kept syntactic positions. Table 1 (column III) shows that Edmonds, who lost four syntactic positions, scores 76%, Garnett, Magarschack and Townsend get 92% each, the rest translators score a 100%.

2.2.4 The Translation Quotient (TQ)

As a result of either manual or computer-assisted translation processing, the teacher gets a tabulated picture (Table 1) of the three analyzed frames (columns I, II, III).

In an attempt to combine the obtained figures in a meaningful whole, the Translation Quotient parameter (TQ, column IV) is used: it is the arithmetic mean of the percentages in the monitored frames. If one adds up the percentage results in all three frames and divides the obtained figure by the number of frames, one gets a TQ, measured in percentage points (pp), which reflects a general quantitative picture of the content-form rendering of the original. This cumulative parameter has shown a perfect fit with the results obtained by other methods of comparative assessment (Tarvi 2004). Table 1 shows four groups of TQ results, from 100% (2 versions) through 97% (2) through 86% (3) to 74% (1).

2.3 The TEM: Count Frames

To further differentiate the translations in their closeness to the original, pure counts of some quantitative parameters can be added to the picture in Table 1: column V (extra tokens, Ets) and VI (formal Tokens, Fts).

2.3.1 Optional Content Frame 1

This frame is a useful tool of assessment, as it shows what has been added to the translation, i.e., the Tts that have no counterparts in the original, labeled as extra Tokens (Et). Table 1 (column V) shows that Wiener, Magarschack, Garnett, Carmichael, and Pevear-Volokhonsky added no extra Tokens (Ets), the Maudes and Edmonds added by one Et each, while Townsend – four.

2.3.2 Optional Formal Frame 3

In this frame, the center of attention is formal Tokens (Fts) – articles, tense markers, etc. Table 1 (column VI) shows that Fts are employed in different quantities: Wiener and the Maudes used two Fts each, Edmonds used four, the rest translators – three Fts each.

2.4 TEM Results: the 13-Token Sentence

The table below gives a cumulative picture of the results in each of the five frames considered:

Table 1. Cumulative Overall Table (13 tokens): Rank Order

I	II	III	IV	V	VI	
TF	MF	SF	TQ	Et	Ft	
(2.2.1)	(2.2.2)	(2.2.3)	(2.2.4)	(2.2.5)	(2.2.6)	
(%)	(%)	(%)	(pp)	(count)	(count)	
Leo Wiener (1899)	100	100	100	0	2	
Aylmer & Louise Maude (1918)	100	100	100	1	2	
David Magarschack (1961)	100	92	100	0	3	
Rochelle S. Townsend (1912)	100	92	97	4	3	
Constance Garnett (1901)	76	92	86	0	3	
Joel Carmichael (1960)	76	92	86	0	3	
Pevear & Volokhonsky (2000)	76	92	86	0	3	
Rosemary Edmonds (1954)	69	84	69	74	1	4

As is seen, there are four groups of the TQ results. In the 100% group, Wiener has a slight advantage (in terms of isomorphism) over the Maudes, since he introduced no extra tokens. In the 97% group, Townsends' translation inferiority (in terms of closeness) is expressed in four extra tokens as compared to no extra tokens in Magarschack's version. In the 86% block, no distinctions can be made because they are word-for-word clones, except for Pevear-Volokhonsky's use of 'each' instead of 'every' (St 8). Edmonds' version (TQ = 74%) has a record (for this sample) number of formal tokens,

four. It does not imply that the translation is bad – this kind of judgment can arise only after a discussion in classroom (3.3).

The one-sentence example, used here pedagogically to explain the TEM techniques, cannot be considered to be fairly representative of the quantitative parameters and their qualitative implications of translated texts. Therefore, we offer the results obtained for a much bigger sample from *Anna Karenina*.

2.4.1 TEM Results: the 405-Token Excerpt

Sheldon (1997) performs a detailed conventional comparative analysis of the four 'focal points' of the novel: the opening sentence considered above (13 tokens), the ball scene (73 tokens), the seduction scene (103 tokens) and the suicide scene (216 tokens). He analyzed the seven translations considered here, except for the version by Pevear and Volokhonsky, which was published three years later. Sheldon concluded that it was Carmichael who showed the best fit with the original.

Here are the quantitative results obtained with the TEM applied to the same excerpts.

Table 2. Cumulative Overall Table (405 tokens): Rank Order

Lost tokens (count)	Kept tokens (count)	TQ (%)	Ft used (count)	Et used (count)	
David Magarschack (1961)	9	396	97,7	96	14
Joe Carmichael (1960)	18	387	95,5	95	15
Constance Garnett (1901)	20	385	95,0	90	8
Aylmer & Louise Maude (1918)	30	375	92,5	91	17
Rosemary Edmonds (1954)	34	371	91,6	87	14
Leo Wiener (1899)	57	348	85,9	74	20
Rochelle S. Townsend (1912)	69	336	82,9	79	42

As is seen, the TQs range from 97,7% to 82,9%. Since the TEM does not cover all aspects of Sheldon's analysis, it favors Magarschack's version, with Carmichael's translation lauded by Sheldon following it closely.

2.5 Language pair independence

In our example with translation from Russian to English, there is an asymmetry in that formal tokens are largely to be seen only on the target side. However, the TEM frames can equally be applied in the reverse direction or to any language pair. Whether or not we choose to exclude some formal tokens from the counting, the

frames are applied in the same way to all translations and their relative differences will be revealed.

2.6 Computational analysis

It has been suggested before that virtual learning environments are useful for translation teaching (e.g., Fictumova (2007)). Our point here is that fine-grained quantitative methods, such as the TEM, can be put to use given support from computational linguistic tools. The proposed environment consists of a central server and a number of client systems for the students. Communication between them is handled as in any e-learning environment, where exercises, grades and other course materials can be stored and accessed. The server includes several modules for monolingual text analysis, such as sentence segmentation, tokenization, lemmatization and PoS-tagging. A parser may also be included to support the computation of the syntax frame. More importantly, there are modules for sentence and word alignments, since this is what is required to support the TEM analysis. In addition, there are modules for reviewing and correcting outputs from all analyzers.

2.6.1 Tokenization

In principle, tokenization, numbering and labeling of tokens (2.1), are processes that computers can handle with ease. It is important, though, that the tokenization is done in a way that supports the purpose to which it will be used. In this case, a tokenization module that only looks at spaces and separators will not be optimal, as the primary unit of TEM is semantic, and may span several text words. Moreover, punctuation marks are not treated as separate tokens in the TEM. This problem could be overcome by tokenizing in two steps. In the first step punctuation marks are removed, lexical tokens are identified using word lists and then formatted as character strings that have no internal spaces. In the second stage spaces are used to identify and number the tokens. Formal tokens can to a large extent be identified as part of this process, using word lists, but extra tokens cannot be identified until after the word alignment.

2.6.2 Sentence alignment

In some cases the translation task may require students not to change sentence boundaries and a one-to-one correspondence between source sentences and sentences of the translations can be assumed to hold when translations are delivered.

If not, a sentence alignment tool such as hunalign (Varga et al., 2005) can be used.

2.6.3 Word alignment

The accuracy of word alignment systems are quite far from 100%. The best performing systems are either statistical, such as Giza++ (Och & Ney, 2003), or hybrid (Moore et al., 2006) and require vast amounts of text to perform well. In the translation class context, the source text will be fairly short, perhaps a few thousand words as a maximum. Even with, say, 20 student translations, the total bitext, consisting of the source text repeated once for each student translation and sentence-aligned with it, will be too short for a statistical aligner to work well. For this reason, a hybrid system that relies on a combination of bilingual resources and statistics for the word alignment seems to be the best choice (cf. Ahrenberg & Tarvi, 2013).

An advantage of having a short source text is that the teacher can develop a dictionary for it in advance to be used by the word aligner. While a teacher cannot predict all possible translations that a student may come up with, this is a resource that can be re-used and extended over several semesters and student groups.

Table 3. Alignment performance on an excerpt from Anna Karenina using different combinations of statistical alignment and lexical resources.

	Prec	Recall	F-score
Giza++	0.499	0.497	0.498
Wordlist based	0.881	0.366	0.517
Combination	0.657	0.610	0.633
Comb + filters	0.820	0.508	0.628

Table 3 shows some results for the Russian-English 405-token excerpt discussed above with different combinations of Giza++-output and lexicon-based alignments. Standard tokenization was used except that punctuation marks were deleted. The source then consists of eight iterations of the excerpt, altogether 3304 tokens¹ and the target text consisting of eight different translations has 4205 tokens. The files were lemmatized before alignment.

The bilingual resources used are a word list of English function words such as articles and possessives that are likely to have no formal counterpart in the source and a bilingual word list created by looking up content words in Google

¹ Standard tokenization does not recognize multitoken units.

Translate. Not all translations suggested by Google have been included. The mean number of translations per Russian lemma is 1.5. In the combinations priority has been given to the alignments proposed by the word lists as they are deemed to have a higher precision.² So, the third row means that Giza++ alignments have been replaced by null links and lexical links induced by the word lists in all cases where there was a contradiction. The fourth row is the result of applying a set of filters based on word frequencies in the corpus and alignment topology to the previous combination.

Obviously, if a complete alignment is called for it is clear that the output of the system must be reviewed and hand-aligned afterwards. There are several interactive word-alignment tools that can be used for this purpose (Tiedemann, 2011), but it will still be time-consuming. However, the burden can be shared between teacher and students, and efforts may be focused on a part of the text only.

2.7 Workflow

After selection of a ST to be used for a translation exercise, the system will have it segmented into sentences, tokenized, and numbered. Then the teacher checks the outcome and corrects it if necessary. The files are then sent to the students. Within the suggested approach, the students are asked to use the client version of the system for translation and then upload their translations to their teacher by a set date before class, or to bring them to class on memory sticks.

When a student translation is in place in the server system, it can be aligned and graded automatically. Of course, the significance of the grades depends on the accuracy of the alignment, but both the student and the teacher can contribute to the reviewing. For instance, the teacher might have marked some words and phrases as especially significant and the student can review the alignments for them in the system for his or her particular translation.

3 In Class

When translations and their alignments are in place in the server system, they can be used as

² The fact that precision is not 100% for wordlist based alignment has two major causes. First, some content words appear two or three times in a sentence and the system does not manage to pick the right occurrence. Also, some common English prepositions get aligned when they shouldn't.

input to various visualization tools. This we see as a further advantage of our approach which will stimulate discussion and reflections among the students. Students' translations can be displayed individually or collectively, on a sentence basis or a phrase basis. Using again the opening sentence of *Anna Karenina* as our example, the outcome for one sentence can look as in Figure 1, where also some of the token frames described above are automatically computed from the alignment.³ Within this format, the teacher is acting as a post-editing human agent who can combine both manners of assessment – computer-assisted and manual.

Since the method combines human and computer resources, it might raise the effectiveness of translation class instruction manifold (Lengyel 2006: 286). The TEM also depersonalizes the problem of grading.

Figure 1. Alignment screenshot for Segment 1 of Translation 1 (Joel Carmichael, 1960) with metrics.

Sentence 1		
Src tokens	Trl tokens	Correspondences
[1] все	Happy	2
[2] счастливые	families	3
[3] семьи	are	0
[4] похожи	all	1
[5] друг	alike	4
[6] на	every	8
[7] друга	unhappy	9
[8] каждая	family	10
[9] несчастливая	is	0
[10] семья	unhappy	11
[11] несчастлива	in	12
[12] по-	its	0
[13] своему	own	13
	way	13
Null aligned		5,6,7

Metrics:	
Basic Content Frame:	76.9
Optional Content Frame:	0
Basic Formal Frame:	3
Optional Formal Frame 1:	...

3.1 From Translation Quotients to Grades

As has been demonstrated, the TEM allows one to get a certain 'cline of fidelity' from the most faithful translation to the freest version. Based on these relative assessments, one can convert the cumulative figures obtained on a number of quantitative parameters to grades. It should be remembered that although the analytical ad-

³ Alignments of the excerpts from *Anna Karenina* can be accessed at <http://www.ida.liu.se/~lah/AnnaK/>

vantage of the frames is that they are minimally subjective, the step from TQ to grades is neither context- nor value-free but depends heavily on the translation task.

Table 4. From TQs to Grades

	TQ	Rank	Grade
Magars hack	97,7	1	Excellent
Carmichael	95,5	2	Good
Garnett	95,0	3	Good
The Maudes	92,5	4	Good -
Edmonds	91,6	5	Good -
Wiener	85,9	6	Satisfactory
Townsend	82,9	7	Satisfactory -

3.2 Gauging Quality

The highlight of the approach is class team work, in the course of which students are expected to have a chance to insert meaningful corrections into their translations and thus improve their ‘home’ grades by the end of class. Because the tokens are numbered, the teacher can easily bring any St, or a group of Sts, on the screen together with all the versions of its or their translations.

It is at this stage that the qualitative side of the TEM comes into play with the aim of improving the final quantitative grade. Let us, for instance, consider the way a group of two tokens from the sentence-example has been rendered. As can be seen here in the manual (Table 5) and computer (Figure 2) versions, this pair of source tokens has been rendered in three different ways. In a computer-equipped class, the required changes can be immediately introduced into the translated texts under discussion.

3.3 Final Grading

As was mentioned in Section 1, the suggested procedure embraces the four basic types of translation evaluation. The method generates absolute score (overall estimates) based on relative scores in separate frames (Table 1).

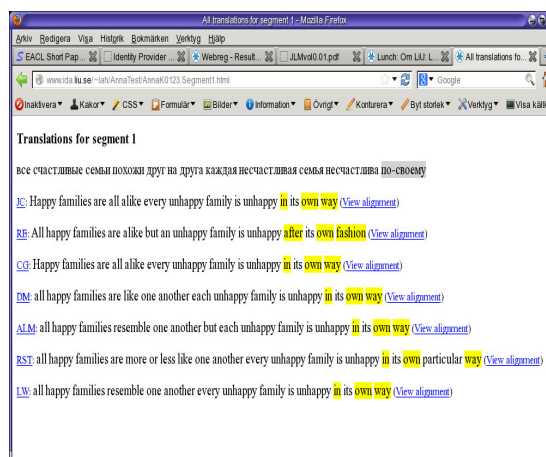
The first monitoring gives a quantitative estimate of students’ homework. After class discussion, which is supposed, like any post-editing, to change the home translations for the better, one more monitoring is carried out, using the same frames. If the system is made incremental, the final grade, which is an arithmetic mean of the home and class grades, can be registered automatically. If, at the end of class, the final grades

are exhibited on screen in their ranking order, it might be the best possible motivation for students to work diligently both at home and in class.

Table 5. Renderings of Source tokens 12-13

LW: (12)in (Ft)its (13)own way.
 CG: (12)in (Ft)its (13)own way.
 ALM: (12)in (Ft)its (13)own way.
 JC: (12)in (Ft)its (13)own way.
 DM: (12)in (Ft)its (13)own way.
 RPLV: (12)in (Ft)its (13)own way.
 RE: (12)after (Ft)its (13)own fashion.
 RT: (12)in (Ft)its (13)own (Et)particular way.

Figure 2. Renderings of Source tokens 12-13 (computed alignments)



4 Outside Class

Within machine translation research, work has been going on for several years, and is still very active, for the search of metrics that assess the similarity of a system translation with human reference translations. Metrics, such as BLEU (Papineni et al.. 2002), TER (Snover et al.. 2006), and Meteor (Lavie and Denkowski: 2009), could also be included in the proposed environment. Published translations or translations that the teacher recognizes as particularly good can be used as reference translations. However, the scores of these metrics do not give as much qualitative information as the TEM frames.

The role of corpora in translation and translation training is a topic of some interest (e.g. Zanettin et al.: 2003). In translator training, the corpora are mostly seen as resources for the student to use when practicing translation (Lopez-

Rodríguez and Tercedor-Sánchez: 2008). This is orthogonal to what we are proposing here, i.e., enabling immediate comparisons and assessments of students' translations as a class-based activity. A system with a similar purpose is reported in Shei and Pain (2002: 323) who describe it as an "intelligent tutoring system designed to help student translators learn to appreciate the distinction between literal and liberal translation". Their system allows students to compare their own translations with reference translations and have them classified in terms of categories such as literal, semantic, and communicative. The comparisons are made one sentence at a time, using the Dice coefficient, i.e., by treating the sentences as bags of words. Our proposal, in contrast, uses more advanced computational linguistics tools and provides text level assessment based on word alignment.

Michaud and McCoy (2013) describe a system and a study where the goal, as in our proposal, is to develop automatic support for translator training. They focus on the inverted TERp metric (Snover et al., 2009) for evaluation of student translations. TERp requires a reference translation but can represent the difference between a given translation and the reference in terms of editing operations such as insertion, deletion, change of word order and matches of different kinds. A weak positive correlation with instructor-based grades (using Pearson's r) could be demonstrated in the study and the authors argue that TERp is sufficiently reliable to provide feed-back to students in a tutoring environment.

The main difference between their proposal and ours is that we start with a metric that has been developed for the task of grading human translations, while TERp is originally an MT metric. Thus, TEM does not require reference translations, but on the other hand its computation has not been automated and so, that is where our current efforts are focused. It should be emphasized that the teacher's load within this approach remains quite heavy but the reviewing work may be shared between teachers and students.

Both the TEM and TERp provide quantitative measurements that can lay the foundation for qualitative discussions and feedback to students but as the TEM does not require a reference it gives the students more freedom in improving their work.

As a more or less objective way of measuring the quantity with allowances made for quality, the method can also be used by teachers at ex-

ams, by editors for choosing a translation, by managers recruiting new in-house translators, by translators for self-monitoring, etc. The computer-generated figures are obtained right on the spot – they may not be exactly accurate but they give a rough general picture at the level of content-form 'presence' of the original in its translations.

Acknowledgement

We are grateful to the anonymous reviewers who provided useful comments and additional references.

References

- Ahrenberg, Lars and Tarvi, Ljuba. 2013. Natural language processing for the translation class. Proceedings of the second workshop on NLP for computer-assisted language learning at NODALIDA 2013 (May 22, Oslo). NEALT Proceedings Series 17 / Linköping Electronic Conference Proceedings 86: 1–10.
- Carmichael, Joel. 1960. *Anna Karenina*, by Lev Tolstoy. New York: Bantam Books, 1980.
- Edmonds, Rosemary. 1954. *Anna Karenina*, by Lev Tolstoy. London: The Folio Society, 1975.
- Garnett, Constance. 1901. *Anna Karenina*, by Leo Tolstoy, with revisions by Leonard J. Kent and Nina Berberova. New York: Modern Library, 1993.
- Honkela, Timo *et al.* 2010. *GIGA: Grounded Intersubjective Concept Analysis: A Method for Enhancing Mutual Understanding and Participation*, Espoo: Aalto University School of Science and Technology.
- Lavie, Alon and Denkowski, Michael J. 2009. 'The Meteor metric for automatic evaluation of machine translation,' *Machine Translation*, Vol 23 (2-3) 105-115.
- Lengyel, István. 2006. 'Book reviews. Ljuba Tarvi: *Comparative Translation Assessment: Quantifying Quality*,' *Across Languages and Cultures* 7 (2) 2006, 284-286.
- Liang, Percy, Taskar, Ben, and Klein, Dan. 2006. 'Alignment by Agreement.' In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 2006, 104-111.
- López-Rodríguez, Clara Inés and Tercedor-Sánchez, María Isabel. 2008. 'Corpora and Students' Autonomy in Scientific and Technical Translation training,' *Journal of Specialized Translation (JoSTrans)*, Issue 09 (2008), 2-19.

- Magarschack, David. 1961. *Anna Karenina*, by Lev Tolstoy. New York: The New American Library.
- Maude, Louise and Maude, Aylmer. 1918. *Anna Karenina*, by Lev Tolstoy, a Norton Critical Edition, ed. George Gabian, with the Maude translation as revised by George Gibian. 2d edition. New York: Norton and Co, 1995.
- Michaud, Lisa N. and McCoy, Patricia Ann. 2013. Applying Machine Translation Metrics to Student-Written Translations. *Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications (BEA8)*, 2013, 306-311.
- Moore, Robert C, Yih, Wen-tau, and Bode, Anders. 2006. 'Improved Discriminative Bilingual Word Alignment.' In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL*, 2006, 513-520.
- Papineni, Kishore, Roukos, Salim, Ward, Todd, and Zhu, Wei-Jing. 2002. 'BLEU: a method for automatic evaluation of machine translation.' In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, 311-318.
- Pevear, Richard & Volokhonsky, Larissa. 2000. *Leo Tolstoy. Anna Karenina*. Penguin Books.
- Shei, Chi-Chiang and Pain, Helen. 2002. 'Computer-Assisted Teaching of Translation Methods.' *Literary & Linguistic Computing*, Vol, 17, No 3 (2002), 323-343.
- Sheldon, Richard. 1997. 'Problems in the English Translation of Anna Karenina.' *Essays in the Art and Theory of Translation*, Lewiston-Queenston-Lampeter: The Edwin Mellen Press, 1997
- Snover, Matthew, Dorr, Bonnie, Schwartz, Richard, Micciulla, Linnea and Makhoul, John. 2006. 'A Study of Translation Edit Rate with Targeted Human Annotation.' *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas*, 223-231.
- Snover, Matthew, Madnani, Nitin, Dorr, Bonnie J., and Schwartz, Richard. 2009. Fluency, adequacy, or HTER? Exploring different judgments with a tunable MT metric. *Proceedings of the EACL Fourth Workshop on Statistical Machine Translation, Athens, Greece, March 30-31, 2009*: 259-268.
- Tarvi, Ljuba. 2004. *Comparative Translation Assessment: Quantifying Quality*, Helsinki: Helsinki University Press.
- Tiedemann. 2011. *Bitext Alignment*. Morgan & Claypool Publishers.
- Townsend, Rochelle S. 1912. *Anna Karenina*, by Count Leo Tolstoy. London & Toronto: J.M. Dent & Sons; New York: E.P. Dutton and Co, 1928.
- Varga, Daniel, Németh, Laszlo, Halácsy, Peter, Kornai, Andras, Trón, Viktor, and Viktor Nagy, 2005. Parallel corpora for medium density languages. *Proceedings of RANLP 2005*.
- Vinay, Jean-Paul & Darbelnet, Jean. 1995 [1958]. *Comparative Stylistics of French and English. A Methodology for Translation*, Amsterdam: John Benjamins.
- Wiener, Leo. 1899. *Anna Karenina*, by Lyof N. Tolstoy, vols II-IV: *The Novels and Other Works of Lyof N. Tolstoy*. New York: Charles Scribner's Sons, 1904.
- Zanettin, Federico, Bernardini, Silvia, and Stewart, Dominic (eds.). 2003. *Corpora in Translator Education*, Manchester.