Original scientific paper

Evaluation of Electronic Translation Tools Through Quality Parameters

Vlasta Kučiš
Department for Translation Studies
Faculty of Arts, University of Maribor
Koroška cesta 160, 2000 Maribor, Slovenia
asta.kucis@siol.net, vlasta.kucis@uni-mb.si

Sanja Seljan
Department for Information Sciences
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, 10000 Zagreb, Croatia
sanja.seljan@ffzg.hr

Ksenija Klasnić
Department of Sociology
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, 10000 Zagreb, Croatia
kklasnic@ffzg.hr

Summary

In the paper, the difference of translation quality of texts obtained through traditional reference works and online electronic translation tools (corpus and multilingual terminology database) will be measured in three main categories: lexical, orthographic and punctuation; syntactically and stylistically using paired samples t-test. The translation was made with the support of electronic translation tools, using the example of a Slovenian bilingual corpus called Evrokorpus and the multilingual terminology database Evroterm. In the paper, the hypothesis that modern online translation tools contribute to the quality and consistency of expert translations, as well as to the acquisition of new competitive skills and knowledge is to be examined.

Key words: translation quality, consistency, translation tools, mistake categorisation, t-test

Introduction

The translation profession underwent a metamorphosis at the turn of the last century, embracing new information and communication knowledge and skills, as well as adopting the usage of modern multilingual technologies, according to

Seljan (2007); Vintar (2008); Željko (2004). The acceptance, implementation and application of translation technologies, as well as the exploitation of their potential by translators during the translation process aim to enhance productivity, competitiveness and the quality of the work. They should therefore be perceived as an integral part of a translator's reality necessitated by globalization and the need for swift information flow.

Over the last ten years, the European Union has been intensively preoccupied with the inherent problems of a multilingual environment, which is a demanding and ambitious project. EU translations have to be unambiguous and terminologically consistent. Such unambiguousness can only be achieved through the consistent and synchronized use of terminology databases and other translation tools.

The GILT sector (Globalization, Internationalization, Localization and Translation) has been facing an increase in translation demands. Due to EU enlargement and the use of the English language as lingua franca on one side, and the growing interest for the protection of national cultures and identities on the other side, the development of multilingual services plays a key role in written communication.

Technical innovations, research and quality management aim to compensate for the lack of translators and increased demands within a time constraint. Several key drivers, such as multilingualism and language technology, market changes (commercial translations, web products, localization) and the Internet (e-books, language barriers in communication, multilingual services, web translations, newsletters) have caused considerable changes in the translation process, relating also to expectations in terms of quality, time and consistency.

The importance of translation practices using ICT does not only witness individual experiences, but also examples in large national translation companies, as presented in (Ørsted, 2001), where assessment procedures aim to evaluate the working environment of translators and support services in IT departments, becoming a corporate issue.

Starting from individual education and practice, up to integrated document workflow, translation quality has been a matter of numerous business applications and workflow document changes.

In the paper, the differences in the translation quality among two groups were analyzed and statistically evaluated. The translation quality of texts obtained through traditional reference works and online electronic translation tools (corpus and multilingual terminology database) will be measured through three main categories: lexical, orthographic and punctuation, syntactically and stylistically using t-test.

This paper analyzes the quality and consistency of translations made with the support of electronic translation tools, using the example of a Slovenian bilingual corpus called *Evrokorpus* and the multilingual terminology database *Evroterm,* which are available at http://www.evrokorpus.gov.si and http://www.

evroterm.gov.si respectively. In the paper, the hypothesis that modern online translation tools contribute to the quality and consistency of expert translations, as well as to the acquisition of new competitive skills and knowledge will be examined.

Related work

Quality assurance is also one of the key issues of the language policy of the European Commission's Directorate-General for Translation (DGT). Documents are mostly translated and revised in-house, demanding the quality standards that apply, according to Farkas, to completeness, terminology, clarity, compliance with linguistic and idiomatic requirements of EU legislation, while revisers consider the text from several points of view including meaning, content, language, style, form and editing. Therefore, the DGT is encouraging the use of translation tools through education, in-house open access and document workflow. To ensure a high quality standard, translators are required to use translation tools, memories and databases. Terminological resources and related databases generally include the translation database of the Ministry of Justice, Eurlex or the CELEX database of legal texts, IATE (Inter-Agency Terminology Exchange) and EURAMIS (European Advanced Multilingual Information System).

According to Hemera and Elekes (2008), apart from the growing need for translations within a very short time period, the Central and Eastern European translation markets have faced problems in the translation business in terms of different expectations when it comes to technical aspects, prices and quality levels. While the U.S. and Western European markets had enough time to learn through educational phases, to experiment with business models and to learn business ethics, CEE countries had to learn very fast and under more difficult circumstances, with no time to experiment, but having to meet high and sophisticated quality standards that have become an indispensable issue in information and communication technology, adequate project management and business flexibility.

According to Waddington (2006), there are no standards in the evaluation of translation quality. Often, we judge whether a translation is more or less appropriate. Contrary to right or wrong answers, it is possible to develop non-binary categories that relate to the degree of acceptability, ranging from the least to the most acceptable translation (1 to 5). Like Waddington, Sager (1989) lists five different types of errors: inversion of meaning, omission, addition, deviation, modification, but also linguistic, semantic and pragmatic effects. Another classification relates to the communicative function, evaluating the degree to which it affects communication in the target language. When comparing source and target tests of several software products in order to determine the translation quality, Gerasimov (2007) includes the following errors: inconsistency, inadequately translated terms, omission, identical source and target segments, punc-

tuation, capitalization, number/value formatting errors, incorrect untranslatables and tags.
As this research was conducted on students' assignments, the evaluation was performed through a points system in which every mistake carried one point. Mistakes were classified in three categories: lexical, orthographic and syntactic/stylistic. This kind of text processing was used for easier data processing and an easy-to-survey mistake evaluation.

Goals and operationality

The pilot project was made at the Department for Translation Studies at the Faculty of Arts, University of Maribor. A random sample of 51 students (N=51) from all four years of study was taken. For this purpose, the same group of students translated two texts of similar length from the same domain, differing in the type of tools used.
The students translated two texts from German into Slovene:

- Group A: Text 1 representing part of the *acquis communaitaire*,
- Group B: Text 2 about intercultural communication in the EU

The students were given 45 minutes to translate both texts, which had approximately the same length and were equally as difficult to translate. The first text was 159 words long, the other 140 words. In both experiments, the translation was made from German into Slovene. Both translations were evaluated by a professional bilingual translator, with both German and Slovene as mother tongues and a degree from the Department of German Language.
The students translated the first text (group A) with the help of German-Slovene/Slovene-German electronic dictionaries Debenjak (2003) installed on the computer and a Duden dictionary http://www.duden.de, while also using Goggle and Yahoo search engines. The use of online dictionaries and search engines was provided with the belief that translators without special education are able to use the mentioned tools.
For the translation of the second text (group B), more specialized translation tools were available:

- a Slovenian bilingual corpus called *Evrokorpus* http://www.evrokorpus.gov.si
- the multilingual terminology database *Evroterm* http://www.evroterm.gov.si
- a terminology base integrated into the SDL Trados translation program, with prior 15-minute training (all students were familiar with Trados from the course "Computer-Aided Translation")

Expert evaluation of the translations of both texts was done for each student, with mistakes in the translations measured in three main categories:

- lexical mistakes,
- spelling and punctuation mistakes, and

- syntactic and stylistic mistakes

The basic goal of the research was to determine the differences in translation between both texts with regard to the introduction of additional interactive, computer-aided tools in the translation process. The mentioned research aimed to examine the hypothesis of whether computer-aided translation tools and resources improve the quality and consistency of translation.

As part of this research, the following theses were tested:

1. Differences in average results between translations are to be statistically significant considering lexical mistakes.
2. Differences in average results between translations are to be statistically significant considering spelling/punctuation mistakes.
3. Differences in average results between translations are to be statistically significant considering syntactic/stylistic mistakes.

Sample

The research was done on a sample (N=51) of students from all four years of study. This was a non-probability convenient sample, i.e. one that encompasses a group of individuals available in a certain situation.

There are some methodology issues arising from this sample. First of all, such samples are not representative because they do not encompass the part of the student population interested in attending classes. That is why the interpretation and conclusions arising from this research cannot be generalized against the complete student population. But, the purpose of the research itself is precisely to check whether interactive tools have any influence on the quality of translation.

Moreover, an appropriate sample is the optimal choice because it encompasses a smaller part of the population that can be regarded as being defined by a mutual characteristic (in this case, all respondents work with foreign languages and study translation at university level), which makes it homogenous. With a larger number of respondents, differences would arise only among students of lower and those of higher years of study. It would be expected that translation ability increases with the progress in the years of study due to more experience and practical work in translation.

However, in this research, because of the size of the sample, differences in average results between students of certain years of study will not be taken into account. Another advantage of this sample is the fact that it is economic and easily realized. It is worth repeating that, regardless of the fact that this is a homogenous sample, generalization against the complete student population would not be justified, because the sample is not representative.

Still, it is possible to make certain conclusions regarding the quality and consistency of translation based on statistical processing using t-tests. If the hypotheses prove to be correct, there is justification for the introduction and use of

interactive translation tools that contribute more to quality, speed and consistency in the translation process.

Results

Comparison of total mistakes

Generally speaking, all respondents (N=51) translated two texts from the same domain that were equally as difficult, of similar length, and were translated under similar conditions. When comparing the total number of lexical, spelling, punctuation, syntactic and stylistic mistakes the students made in both texts, we can see that in the first translation there was a total of 958 mistakes, in the second a total of 571 (Table 1). Average number of mistakes in the first translation was 18.78 which decreased in the second translation on the average of 11.20 mistakes. The coefficient of variability presented in Table 1. represents the ratio of the standard deviation of a variable relative to its mean and it measures the degree of variation in each variable. It can be seen than there is a slightly less variability of mistakes in the second translation.

Table 1. Total number of mistakes and paired samples statistics

	Total No. of mistakes	N	Average result	Standard deviation	Coefficient of variability
Group A	958	51	18.78	6.100	32.48
Group B	571	51	11.20	3.742	33.41

Seeing how this is the same sample of respondents in both tests with changed conditions, to test the statistical significance of the difference between the arithmetic means of the samples we used t-test for dependent samples which is a standard parametric test used to test the significance of the change in the average result after the controlled change of conditions. The t-test is based on the comparison of the calculated t-value with the theoretical t-value from the table of critical t-values with respect to different number of degrees of freedom and different risk levels. The calculation of the observed t-value was done using the formula in which the t-value is expressed as the ratio of the difference of arithmetic means and the standard error of difference between means.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}$$

This method is often called a correlated t-test because the Pearson's coefficient of correlation between two measurements is used in the computing of the standard error of difference between means.

$$s_{\bar{X}_1-\bar{X}_2} = \sqrt{s^2_{\bar{X}_1} + s^2_{\bar{X}_2} - 2r_{1,2}s_{\bar{X}_1}s_{\bar{X}_2}}$$

So, to test weather there is a statistically significant change in the average number of mistakes after the repeated testing introduced new parameters and we tracked their influence on the quality of translation, the data were introduced in the formula presented above and the corresponding t-value and border p-value were calculated. The statistical testing was performed two-sided, at risk level α=0.05 and degrees of freedom df=50. Seeing how the border p-value (which represents the probability of the type I error: the rejection of the null hypothesis that is correct) is less than 0.001, we can conclude that the average number of all mistakes has statistically significantly decreased after the introduction of electronic translation tools suggesting the need for adequate education and use of translation tools.

Table 2. Paired samples t-test of statistically significant difference between average number of mistakes

t	p	df
10.553	<.001	50

Lexical mistakes

As has already been pointed out, the same sample of students translated the first text with the help of dictionaries and web search engines, and the second text with the help of web sources, and *Evroterm* and *Evrokorpus*.

The students (N=51) made a total of 479 mistakes in the first text and 302 in the second text. The average result is shown in Table 3. The coefficient of variability is higher in group B, suggesting the bigger variations when using translation tools.

Table 3. Number of lexical mistakes and paired samples statistics

	No. of lexical mistakes	N	Average result	Standard deviation	Coefficient of variability
Group A	479	51	9.39	3.567	37.98
Group B	302	51	5.92	2.489	42.04

As presented in Table 3, the averages of samples differ, and the t-test has determined (t=7.175) that there is a statistically significant difference at the level p<0.001 (Table 4). Therefore, the first hypothesis can be accepted. This means that the comparison of the two translations can lead to the conclusion that interactive tools significantly contributed to the quality of translation, at least when it comes to lexical mistakes, where the number of lexical mistakes was significantly lower using additional interactive tools.

Table 4. Paired samples t-test of statistically significant difference between average number of lexical mistakes

t	p	df
7.175	<.001	50

Spelling and punctuation mistakes

In the same way as in the case of lexical mistakes, spelling mistakes in both translations were analyzed. In total, the number of mistakes students made amounted to 243 in the first text and 131 in the second text (Table 5.) The coefficient variability is considerably bigger in group B.

Table 5. Number of spelling and punctuation mistakes and paired samples statistics

	No. of spelling and punct. mist.	N	Average result	Standard deviation	Coefficient of variability
Group A	479	51	4.76	2.566	53.566
Group B	302	51	2.57	1.814	70.583

The t-test determined that in this case there is also a statistically significant difference between the average number of spelling and punctuation mistakes in two translations (t=5.887). We can conclude that the second hypothesis is accepted as well, i.e. that the use of additional translation tools significantly decreased the number of spelling mistakes ($p<0.001$) (Table 6).

Table 6. Paired samples t-test of statistically significant difference between average number of spelling and punctuation mistakes

t	p	df
5.887	<.001	50

Syntactic and stylistic mistakes

In the same way, we compared syntactic and stylistic mistakes in both translations. The total number of mistakes the students made amounted to 236 in the first text and 138 in the second (Table 7). The coefficient of variability is considerably bigger in group B.

Table 7. Number of syntactic and stylistic mistakes and paired samples statistics

	No. of syntactic and styl. mist.	N	Average result	Standard deviation	Coefficient of variability
Group A	236	51	4.63	2.425	52.375
Group B	138	51	2.71	1.701	62.76

The t-test determined that in this case there is also a statistically significant difference in the average number of syntactic and stylistic mistakes between the two translations. We can conclude that the third hypothesis is accepted as well, i.e. that the use of electronic translation tools has, on average, significantly decreased the number of syntactic and stylistic mistakes ($t=4.43$) with $p<0.001$ (Table 8).

Table 8. T-test of statistically significant differences of syntactic and stylistic mistakes

t-test	p	df
4.43	<.001	50

Interpretation of results

Analyzing the quality of translation and type of mistakes (lexical, spelling and punctuation, syntactic and stylistic), the general conclusion is that the introduction of additional computer-aided translation tools significantly influences the quality and consistency of translation.

Taking into account conditions for translation, time and identical text types, it can be concluded that the use of electronic tools was of significant help to students regarding the quality of their translation, although we cannot make conclusions against the entire population of students of the same departments. In the case of such an analysis, other variables would be important, such as the year of study, success, (lack of) motivation, etc.

T-tests, resulting with t-values 7.175, 5.887 and 4.43 respectively, have all shown statistically significant differences at the level of probability lesser than 0.001 and indicated the acceptance of hypothesis 1, 2 and 3 claiming that translation tools improve the quality of translation at lexical, spelling and punctuation and also syntactic and stylistic level.

In any case, the same sample of students showed significantly better results when using an online corpus and terminology databases. It is important to mention that the introduction of additional electronic tools in translation has, on average, decreased the number of mistakes in all analyzed categories. This means that additional online tools contribute to the quality and consistency of translation on all of the most important levels.

Table 9: Percentage of translation improvements

Mistakes	Group A	Group B	Improvement in %
Lexical	479	302	22.66
Spelling	243	131	29.96
Syntactic / stylistic	236	138	26.20
TOTAL	958	571	25.31

Conclusion

The increasing demand for simultaneous translation and integrated solutions also suggests high quality translations. Adequate education and the use of ICT, i.e. computer-assisted translation tools and their integration into document workflow, could help in the translation process during preparation, translation and revision.

The use of additional translation tools (online terminology base, created terminology base and online corpus) significantly influenced the quality and consistency of translation in general (25.31%), but also on all levels (lexical, spelling and punctuation, syntactic and semantic) ranging from 22.66 – 29.96%. The hypothesis that modern electronic translation tools contribute to the quality and consistency of translation has been accepted with the probability of a type I error being lower than 0.1%. The differences among the results on the three mentioned levels are statistically significant at the level $p<0.001$.

With high expectations regarding the translation quality, time constraints and demand for increased productivity, translators are faced with new challenges in education and in business. The use of translation tools certainly improves the quality of professional translations, but has become a corporate issue, asking for horizontal and vertical integration.

Acknowledgments

This work has been supported by the Ministry of Science, Education and Sports of the Republic of Croatia, under the grant No. 130-1300646-0909.

References

Bonet, Josep. Present and future of machine translation in the Directorate General for Translation of the European Commission. Barcelona, Spain. June 19-21, 2006. http://www.elda.org/tcstar-workshop_2006/pdfs/keynotes/tcstar06_bonet.pdf

BRUNDAGE Jennifer. Translation Quality for Professionals, 3.1.2000. http://www.istworld.org/ProjectDetails...

Debenjak,D. *Veliki slovensko-nemški slovar*, DZS and*Veliki nemško-slovenski slovar*, DZS, 2003.

DGT of the EC. *Translation Tools and Workflow*, 2009. http://ec.europa.eu/dgs/translation/bookshelf/tools_and_workflow_en.pdf

Farkas, Ágnes. Translation at the European Commission: quality criteria and quality assurance. http://www.translationconference.com/2006_eloadasok/FarkasEN.pdf

Gerasimov, Andrei. Review of Translation Quality Assurance Software. 2007. http://www.translatorscafe.com/cafe/article71.htm

Groenewald, H. J.; Fourie, Wildrich. Introducing the Autshumato Integrated Translation Environment. EAMT, 2009. http://www.mt-archive.info/EAMT-2009-Groenewald.pdf

Hager, Astrid. The translation market in ten years' time - a forecast. 6/ 2008, pp. 14. http://www.tekom.de/upload/alg/tcworld_608.pdf

Hemera, Annette; Elekes, György. The Central and Eastern European translation market. Multilingual Computing Inc., March 2008. http://www.leg.eu/images/File/MultiLingual_march.pdf

Joann, Drugan. Multilingual document management and workflow in the European institutions. http://www.leeds.ac.uk/cts/research/publications/leeds-cts-2004-11-drugan.pdf

Muzii, Luigi. Quality Assessment and Economic Sustainability of Translation. ITI conference London, 2009. http://www.openstarts.units.it/dspace/bitstream/10077/2891/1/ritt9_05 muzii.pdf

Ørsted, Jeannette. Quality and Efficiency: Incompatible Elements in Translation Practice? Meta : journal des traducteurs / Meta: Translators' Journal, vol. 46, n° 2, 2001, p. 438-447. http://id.erudit.org/iderudit/003766ar

Seljan, Sanja; Gaspar, Angelina. Primjena prevoditeljskih alata u EU i potreba za hrvatskim tehnologijama: Translation Tools in EU and need for Croatian Language Resources. Jezicna politika i jezicna stvarnost / Language Policy and Language Reality. Zagreb: HDPL, 2009. pp. 617-625.

Seljan, Sanja; Gašpar, Angelina; Pavuna, Damir. Sentence Alignment as the Basis For Translation Memory Database // The Future of Information Sciences: INFuture 2007 - Digital Information and Heritage / Seljan, Sanja ; Stančić, Hrvoje (ur.). Zagreb : Odsjek za Informacijske znanosti, Filozofski fakultet Zagreb, 2007. Str. 299-311.

Seljan, Sanja; Pavuna, Damir. *Why Machine-Assisted Translation (MAT) Tools for Croatian?* // Proceeding of 28th International Information Technology Interfaces Conference - ITI 2006. pp. 469-475

Vintar, Špela. Corpora in translation: a Slovene perspective. *The journal of specialised translation*, 2008. http://www.jotrans.org/issue 10/art vintar.pdf.

Vintar, Špela. Localization, globalisation or Slovenization? *Mostovi,* 2004, 38, pp. 74-82.

Željko, Miran. Evroterm and Evrokorpus – a terminology database and a corpus of translations. In: Humar, Marjeta (ed.) *Terminology at the Time of Globalization*, ZRC Publishing, 2004, pp. 139-149.

Waddington, Christopher. Measuring the effect of errors on translation quality. Lebende Sprachen. Volume 51, Issue 2, pp. 67-71, 2006. http://www.reference-global.com/doi/abs/10.1515/LES.2006.67