EDU WORLD 2018
The 8th International Conference

LEARNER CORPUS-DRIVEN APPROACH TO QUALITY ASSURANCE IN TRANSLATION

Titela Vîlceanu (a)*
*Corresponding author

(a) University of Craiova, 13 A.I. Cuza Street, Craiova, Romania, elavilceanu@yahoo.com

Abstract

The paper underpins a balanced parallel corpus made up of students' translations from the mother tongue (Romanian - L1) into the foreign language (English - L2) with a view to identifying recurrent patterns and strategies, as well as most frequent and serious errors. The paper aim is manifold: to determine the effectiveness and efficiency of the learners'/trainees' routinised ways of achieving equivalence between the source text and the target text, to provide a classification of errors alongside viable pathways of remedial work. Retrospective views will be coupled with prospective views or feedforwarding in the sense of anticipating the translation trainees' main difficulties and making available re-usables. Even if the corpus is not very large, it manages to achieve cumulative representativeness and it is useful to inform language study from a translation-oriented perspective. Besides, although mainstream literature advocates that cumulative representative corpora are highly idiosyncratic, we strongly believe that the research findings may gain generalising force against an evidence-based mechanism and on account of the fact that the corpus data may be exploited fully. The research carried out with 1st and 2nd year translation students is both experimental and exploratory in nature, and qualitative and quantitative data are collected and interpreted in consistent ways. The research complexity is also given by the fact that process- and product-oriented perspectives on translation are combined, students' being asked to reflect on their work in progress and on the end product quality.

Keywords: Learner corpora, translation skills, quality assurance.

1. Introduction

The ever growing interest in real-life communication has shifted attention from the analysis of lexis in isolation (for instance, frequency of single and multiword units) to the detection of the multilayered dependence between lexis and grammar, between lexical and grammatical choices, on the one hand, and co-text and socio-cultural context, on the other hand, integrating morphological, syntactic, semantic and pragmatic aspects (Altenberg & Granger, 2011).

We start from the programmatic statement that corpus linguistics has far reached beyond the confines of the (newly emerged) discipline and beyond descriptive approaches to language in use both in the oral and written modes of communication, extending its areas of application (language teaching, lexicography, translation studies, etc.) and fairly contributing to resource development (i.e., innovative material design and practices) (Adolphs & Lin, 2011).

Over the past decades, we have witnessed technical advances in methods that use unannotated corpora to automatically display representations of meaning, basically, of single or small-sized units, but ever more frequently, of larger or multiple units. Theoretical and applied linguists (to use blanket terms covering interdisciplinary concerns, too) have been focusing on the development of structured linguistic resources. Needless to say that the corpora provide information which is both richer and more reliable that the (subjective) data derived from introspection.

In the same climate of opinion, translation theorists and professionals have joined their efforts to build synergies between two seemingly mutually exclusive approaches to the development of the multilayered translation competence (including language competence, thematic competence, interpersonal competence, technological competence, etc.): corpus-based and corpus-driven methods.

1.1. Overview of corpus-based approaches to teaching translation

Corpus-based approaches underpin a theoretical model endorsing cross-linguistic isomorphisms and anisomorphisms, and use bilingual and multilingual corpora to validate and/or refine the model (deductive workflow). One special mention concerns the fact that *corpus based approach* seems to be used indiscriminately as a blanket term for both *corpus-based* and *corpus-driven methods*.

Corpus-based methods feature the extraction and structuring of multilingual term banks, and the creation of translation memories/workbenches and other support tools.

1.2. Outline of corpus-driven methods to teaching translation

In a narrow sense, corpus-driven approaches exploit the corpus in search for cross-linguistic similarities and dissimilarities so as to arrive at theoretical statements. Corpus-driven methods are generally envisaged as allowing to expand and correct existing linguistic resources; conversely, hand-crafted resources may be said to represent additional sources of monitoring and supervision when learning meaning representations automatically.

Hence, translation corpora are likely not only to reveal patterns which would otherwise be difficult to detect, but also to detect unexpected equivalents functioning as optimal solutions and highlighting the translator's interpretative choices and resourcefulness. A word of caution: an unexpected equivalent should comply with the principles of statistical significance and internal coherence in order to be considered a

viable solution and be safely re-used, in other words, the unexpected rendering should outnumber the less successful attempts and the translator should use it consistently throughout the target text.

2. Problem Statement

The exploitation of corpora to establish correspondences as standardised solutions or recurrent source-target language pairings, mostly associated with specialised translation, as well as equivalences, i.e. contextual renderings resulting from disambiguation, of equal value for human translation and computer-assisted translation, points out to a host of conceptual and methodological difficulties.

2.1. Challenges of building and exploiting learner corpora

The potential and limitations of translation corpora have been widely acknowledged by language and translation theorists as well as by practitioners (language teachers, translators, etc.) One of the main strategic benefits of translation corpora is that they provide sets of cross-linguistic equivalents (equated to the variability factor in translation, i.e. the translator's choice in search of the optimal equivalent) as at the functional and stylistic levels. With reference to strategic costs, translation corpora are said to be marked by negative transfers, i.e. *translationese*, under the influence of the mother tongue. Additionally, by their very nature, translation corpora are made up of written texts belonging to a specific genre and type of discourse, which can be considered both an advantage (in the sense of allowing for in-depth analysis and interpretation of findings as guidelines for quality assurance in translation) and a disadvantage (their application is limited to certain cross-linguistic studies). It becomes obvious that the generalising force of the cross-linguistic insights gained from translation corpora increases with multilingual databases (where the translation of the source text is available in more than two target languages).

Another recurrent issue is linked to text alignment: to secure efficiency and effectiveness, each unit in the source text should be aligned to the corresponding unit in the target text. Customarily, translation corpora are aligned sentence by sentence, although we have to recognise that alignment may also be done at the paragraph, phrase and word levels as minimal units in translation.

2.2. The Romanian context(ing)

To our best knowledge, although all the long-established and high ranking Romanian universities have developed translation Bachelor's programmes, translation teaching still relies heavily on traditional methods, and the learners' potential and achievements are still underexploited. Learner corpora, even if a reality in the Romanian academic environment, are not used to raise wider awareness of common problems faced by translation students and translation trainers, or to build networks and strengthen communities of practice.

Under the circumstances, training translation students becomes a question of developing their translation competence in conjunction with corpus design and use skills, alongside critical thinking skills and networking.

3. Research Questions

The research questions are centred on the following areas:

3.1. Translation type and usability of learner corpora

- Are learner corpora dependent on text and translation type?

- Are translation learner corpora stable or ephemeral?

- Are translation learner corpora rough and ready or representative for a larger text collection?

- Can learner corpora complement other resources in translation quality assurance?

3.2. Degree of learners' involvement in corpus design and use

- Can learner corpora provide evidence of translator's strategic choices?

- Is the learner corpus creation, management and analysis rewarding?

- Are translation learner corpora genuinely comparable?

4. Purpose of the Study

The impact of the digital revolution on language and translation research, as well as its far-reaching influence on the related professions has been widely discussed and long ascertained. Therefore, the aim of the current paper relates to a less explored area: rather than enlarging on the usefulness of dedicated software and other translation tools and toolkits in training translators, I shall focus on the role of the translation trainees in the multi-staged process of learner corpora design and use, starting from relevant data collection and ending with improving reference tools and expert systems.

4.1. Translation trainees engagement with learner corpora design

We strongly believe that human intervention, i.e. translation trainees' and trainers alike, is needed in corpus design and use in order to secure user-friendliness and viable insights. According to Ahmad (2008), the trainees' "being in text" can be said to enable "the text in being" to become representative (p. 60).

4.2. Usability and re-usability as quality label

As already mentioned, learner corpora deal with language use rather than abstract systems, therefore we should consider the relevance and reliability of the corpora in different translation settings when embarking upon learner corpus design.

Likewise, we should allow for learner corpora to be enriched with linguistic and contextual information from a multisided perspective, mapping theoretical landscapes and the industry.

5. Research Methods

The research was carried out with 1st and 2nd year translation students in the second semester of the academic year 2017-2018. The group size was of 20 students, totalling 40 students enrolled in the Translation and Interpretation (English and French) programme at the Faculty of Letters, University of Craiova.

The research was both experimental and exploratory in nature, and qualitative and quantitative data were collected and interpreted in consistent ways. The research complexity was also given by the fact that process- and product-oriented perspectives on translation were combined, students' being asked to reflect on their work in progress and on the end product quality.

More specifically, I envisaged action research as defined by Richards and Schmidt (2002, pp. 8-9): "research that has the primary goal of finding ways of solving problems, bringing about social change or practical action, in comparison with research that seeks to discover scientific principles or develop general laws and theories". I adopted and adapted Nunan's (2006) research cycle, which can be broken down as follows:

5.1. Problem identification

Translation trainees are not accustomed to designing learner corpora and using them to assure quality in translation.

5.2. Preliminary investigation

Translation trainees were showcased the MeLLANGE learner corpus (LTC) (please visit http://corpus.leeds.ac.uk/mellange/ltc.html), which currently consists of 429 student translations, out of which 232 are annotated translations (signalling translation errors), as an example of how to store, systematise and retrieve data and metadata.

5.3. Hypothesis formation

In order to re-purpose this learner corpus, i.e. to transfer the template and accommodate it, students were asked to focus on one type of translation, and they opted out for administrative translation, more precisely on institutional translation as embedding academic marketing via websites (www.ucv.ro, www.incesa.ro).

5.4. Plan intervention

The students were asked to select 5 texts out of a list of 15 texts, text length: about 250 words, and in groups they had to translate the texts from Romanian into English. The translation evaluation was subsequently performed by the tutor (myself in this case) and also involved peer assessment (group work) - see Table 01 below.

Table 01. Action plan

Total number of student groups (group size: 5 members)	Total number of source texts	Total number of target texts
8	5	40

5.5. Initiate and observe outcomes

LTC made use of two broad categories of errors: content-related and language-related - our corpus was annotated with respect to error typology by taking over the two broad types and identifying several sub-types, as indicated by Table 02:

Table 02. Error typology

Language-related		Content-related	
Grammatical accuracy (GA)	**Lexical accuracy (LA)**	**Stylistic conventions (SC)**	**Omissions**
Serious mistake (tense, use of modals, word order, sequence of tenses, etc.)	Near synonym	Register (wrong level of formality)	Impacting heavily
Less serious mistake (articles, plural forms, spelling and punctuation, etc.)	Inadequate word or phrase	Upgrading/downgrading procedure (inadequate use of terminology)	Not resulting in global meaning distortion

5.6. Identification of follow-up problem

The students made observations and comments with reference to cultural gaps and mismanagement of cultural loads in translation, and we decided to add up another category to error typology - see Table 03.

Table 03. Enlarged error typology

Language-related		Content-related		
Grammatical accuracy (GA)	**Lexical accuracy (LA)**	**Stylistic conventions (SC)**	**Cultural intertraffic (CI)**	**Omissions**
Serious mistake (tense, use of modals, word order, sequence of tenses, etc.)	Near synonym	Register (wrong level of formality)	Inadequate domestication strategy	Impacting heavily
Less serious mistake (articles, plural forms, spelling and punctuation, etc.)	Inadequate word or phrase	Upgrading/downgrading procedure (inadequate use of terminology)	Inadequate foreignizing strategy	Not resulting in global meaning distortion

5.7. Second hypothesis

Following the corpus annotation, students become more aware of error translation typology and of the criteria for translation quality assurance, being now able to feedforward, anticipate recurrent problems in the translation of administrative texts.

5.8. Second round action and observation

During the last stage of our action research, the students were asked to provide solutions to the most recurrent problems encountered closely following the established error typology.

6. Findings

The learners' active involvement in corpus design and decision-making processes empowers them,

demonstrating that human intervention is required at a number of stages: data collection, data formatting and derivation of translation guidelines so as to secure fit-for-purpose (institutional) translation.

6.1. Translation guidelines as re-usables

Based on our corpus-driven research, the following guidelines were agreed on:

- identification of text type;
- identification of target readership's expectations and compliance with client's specifications;
- information mining with a view to acquiring thematic competence, terminology included;
- identification of linguistic and cultural gaps;
- process- and product-oriented approach to quality assurance.

6.2. By-product

Although it was not envisaged in the beginning, translation trainees suggested to build a bilingual Romanian-English glossary of institutional terms as resulting from the need of standardising terminology.

7. Conclusion

Learner corpora research is still underrepresented in Romania, especially when incorporating the linguist and the translator's work. It is a fact yet to be acknowledged by many translation scholars and professionals that the digital revolution cannot be equated to the switch from paper-based translation to machine and computer-aided translation. The available multilingual databases, dedicated translation software (free or commercial) have undermined dictionaries as the most reliable sources in translation - in Varantola's (2000) words, translators still rely on dictionary information, but "they also need reassurance when checking their hunches or when they find equivalents they are not familiar with" (p. 118). Corpus linguistics has also made available a wealth of electronic texts, and more recently, corpus design and use has become highly dynamic - similarly to translation, corpora do not exist in a social vacuum, they are demand-driven and they should meet specific requirements.

References

Adolphs, S & Lin, P.M.S. (2011). Corpus Linguistics. In J. Simpson (Ed.), *The Routledge Handbook of Applied Linguistics. The Linguist and the Translator* (pp. 597-610). London: Routledge.

Ahmad, K. (2008). Being in Text and Text in Being: Notes on Representative Texts. In G. Anderman & M. Rogers (Eds.), *Incorporating Corpora. The Linguist and the Translator* (pp. 60-95). Clevedon: Multilingual Matters.

Altenberg, B & Granger, S. (Eds.) (2011). *Lexis in Contrast. Corpus-Based Approaches*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Nunan, D. (2006). *Research Methods in Language Learning*. Cambridge: Cambridge University Press.

Richards, J.C. & Schmidt, R. (Eds.) (2002). *Longman Dictionary of Language Teaching and Applied Linguistics*. London: Longman.

Varantola, K. (2000). Translators, Dictionaries and Text Corpora. In S. Bernardini & F. Zanettin (Eds.), I corpora nella didattica della traduzione/Corpus Use and Learning to Translate (pp. 117-133) Bologna: Cooperativa Libraria Universitaria Editrice.