

5th International Conference on Corpus Linguistics (CILC2013)

An Analysis of Research Production in Corpus Linguistics Applied to Translation

Miguel A. Candel-Mora^{a*}, Chelo Vargas-Sierra^b

^aUniversitat Politècnica de València, Departamento de Lingüística Aplicada, Camino de Vera, 14, 46022 Valencia, Spain

^bUniversidad de Alicante, Departamento de Filología Inglesa, Apdo.99, 03080 Alicante, Spain

Abstract

The aim of the paper is to analyze with data the consolidation of corpus methods in translation and to specify which issues are under research and the features that characterize these studies. To that end, different contributions to corpus linguistics research, teaching and practice of translation were compiled to build a sufficiently representative sample: 389 bibliographic records on corpus linguistics applied to translation. This study deals with the identification and analysis of different data from the records: lines of research, applied and theoretical, language pairs, and in sum, if Corpus Linguistics is a well-established methodology in research in translation and in which way.

© 2013 The Authors. Published by Elsevier Ltd.

Selection and peer-review under responsibility of CILC2013.

Keywords: corpus linguistics, translation, corpus methodology applied to translation

1. Introduction

This paper takes as its starting point the assumption that regular publications, conferences, symposia, books, etc. about a specific topic can be claimed to consolidate a new discipline, namely corpus-based or corpus-driven translation studies. To that end we will use data retrieved from a number of bibliographic records collected from specific databases and will undertake a multi-variable characterization of the corpus methodology applied to translation research.

* Corresponding author. Tel.: +34 96 3870007 Ext. 75341; Fax: +34 963877539.

E-mail address: mcandel@upvnet.upv.es

In a time when corpus-based study methods have proven established in linguistic research and have expanded most of its methods and language analysis techniques to other disciplines such as lexicology, terminology, language teaching and translation, it seems justified to carry out an analysis of the scientific production and consolidated research lines of Corpus Linguistics applied to one of these disciplines: Translation.

Most studies and research on translation range from the observation of different aspects of the translation process to the observation of translation as a product; and for different purposes, such as the study of translation strategies, as a tool for teaching future translators by means of observing solutions adopted by other translators or as a contribution to contrastive studies, the latter motivated by the growing use of computer assisted translation tools. This observation of previous translations is made possible by computerized databases that provide translators with textual reference and consultation material in order to maintain coherence and consistency in their decision making during the translation process –especially for longer texts with a degree of repetition.

This context of new research possibilities is favored by the increase in data processing speed and storage capacity of computer systems in recent years, the availability of texts in electronic format for the development of corpus and its methods and the collection of textual databases.

Corpus Linguistics and Translation Studies are both relatively recent fields of study, although both have managed to build independent and interdisciplinary study areas in recent years. This commitment is manifested in several ways: from their inclusion in university curricula, the organization of academic conferences, workshops and seminars, the implementation of research projects and the direct or indirect application of research results. But above all, the main indicator of the consolidation and quality of the activities generated around an academic discipline is the number of publications arising thereon.

Based on the different contributions to corpus linguistics research, teaching and practice of translation, this paper proposes a study of a sufficiently representative sample of bibliography: 389 bibliographic records on corpus linguistics applied to translation in order to identify different lines of research arising from the binomial Corpus Linguistics-Translation.

This paper is divided into the following sections: after presenting the rationale behind this work and the establishment of goals, the first section briefly reviews the literature on corpus linguistics applied to translation and its main contributions in research. Secondly, the method and tools used in the work are described, along with information from the bibliographic sample, the limitations noted, followed by a section which discusses the variables studied and the percentages and statistics of the variables selected.

2. Corpus applied to translation

Publications on corpora and translation started to emerge in the 1980s, particularly in the Scandinavian countries and in relation to the study of literary texts (Anderman & Rogers 2008: 13). Several authors (Laviosa 2002: 1; Anderman & Rogers 2008: 13; Zanettin 2012: 12, among others) acknowledge Gellerstam in 1986 and Linquist in 1989 as the earliest works on corpus-based translation studies, with their studies comparing a specific aspect (i.e., distribution of words and English adverbials) of original texts with their corresponding translations. Laviosa (2002: 1) emphasizes that corpus-based translation studies have consolidated as a discipline in its own stating that “corpus-based studies of translation rise to a coherent and distinct body of research in both branches of the discipline [theoretical and descriptive research]”. Since Mona Baker’s (1993) seminal article advocated for the use of corpus linguistic methodologies in translation studies, it has been acclaimed as a renewing force by important researchers such as Malmkjær (2003), Rabadán (2009) and Tymoczko (1998) to name a few. Olohan (2002:1) dates the use of corpora as a research tool in translation to no more than twenty years ago and coincides that the line of work in corpus applied to translation seems to be developing its own entity.

Research works on corpus based translation studies are usually classified into three main areas: theoretical studies, empirical results and applications (Laviosa 2002: 22). In turn, the application of corpus to translation is divided into three areas: study of translation as a product and as a process, training of translators and as a tool for contrastive linguistic studies (Baker 1995; Laviosa 2002; Olohan 2004). This research field evolves very rapidly due to the increase in data processing speed and availability of texts in electronic format for the development of corpus. In a professional context, we can also find tools and techniques that contribute to improving the quality and increase the effectiveness of translation work (Laviosa 2002; Vargas-Sierra 2012).

As Laviosa states (2011: 143), corpora are currently playing an important role in every line of work in Translation Studies and the use of corpus material and methodologies is increasingly growing in theoretical, descriptive and contrastive studies, in pedagogical applications, in language teaching and learning, as well as in applications for Computer-Assisted Translation or Machine Translation.

3. Method and tools

The literature collected comprises 389 bibliographic records, ranging from purely empirical studies to the initial classification proposed by several authors between its use for teaching, for research or for professional applications.

In statistics, the sampling method is the study of relations existing between a population and the samples extracted from it. The main function of sampling is determining which part of a population under study should be examined to be able to draw proper inferences about such population. Obtaining an adequate sample means to get a simplified version of the entire population, in that it has to represent its main features. In order to know how large our population was, first we needed to know how many publications of any kind were written in total about corpus and translation. However, to select a representative sample, and as a starting point, we considered our population under study as finite (<100.000 elements to study). The next step was to limit our search and focus on two bibliographic databases with translation references, namely BITRA (Bibliography of Interpreting and Translation) and TSA (Translation Studies Abstracts Online). Then, these databases were queried by using the word 'corpus' in several fields –title, keyword and words in abstract.

After the initial queries to these databases, we decided that our entire population comprised 900 records. Therefore, with a margin of error of less than 5% and a confidence interval of 95%, our sample size had to contain at least 269 records.

Our method of sampling was opportunistic, in the sense given by Leech (2010): the samples were easy to obtain. However, our type of research is qualitative in nature, since it is based on content analysis, i.e., we carried out a close inspection of the results given by the databases in order to decide whether all of them were representative of the object of the study: all the final selected records dealt with corpus and translation. Then we proceeded with the registering and storing of data in a spreadsheet. These data were the following: language(s) of corpora, type (if they were comparable or parallel), corpus size, translation specialization (medical, literary, audiovisual, etc.), corpus orientation (pedagogical, research, or professional), applications and type of publication.

4. Analysis and discussion

4.1. Languages

One of the first indicators studied was the combination of languages in order to identify the origin of the corpus as a significant aspect of research in corpus-based translation studies.

From the bibliographic sample analyzed, 187 records specify the language or combination of languages, while the rest of bibliographic records, a total of 202, do not mention the language or languages of the corpus either because the paper deals with methodological approaches, it refers to the different possibilities of using corpus as a tool for the study of translation or do not use any specific corpus to carry out the study.

From the 187 records that specify the language, 76 records refer to a corpus in English or include English in the language pair of the study, followed by Spanish with 25 records, Chinese 7, Finnish with 6 and Arabic and Swedish 5. The other languages, French, Catalan, Portuguese, German and Italian are mentioned in only 2 records. Japanese, Hebrew, Russian, Dutch, Norwegian, Persian and Xhosa have a representation of 1 record.

4.2. Type of corpus

The reference to the type of corpus used is also a representative indicator of the establishment of the discipline of corpus-based translation studies. For this variable, we had a total of 109 records that mentioned the type of corpus

used in the abstract. As seen in Figure 1, the highest number of records, 63, corresponds to parallel corpus, although there are 16 cases in which both types of corpus are used: comparable and parallel.

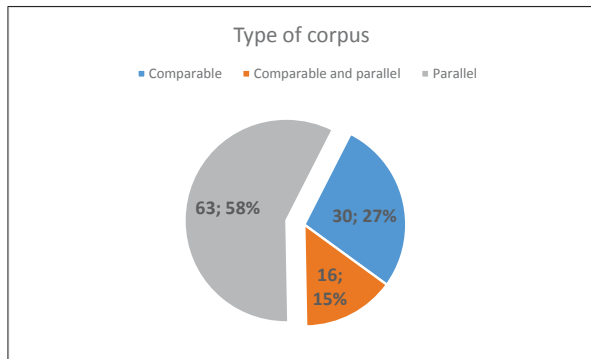


Figure 1: Type of corpus

What the data show is that the studies are consistent with the primary literature consulted for this study (Baker, 1995; Laviosa, 2002; Olohan, 2004) which proposes exploiting parallel corpora as a tool for deciphering most issues in translation. Currently, parallel corpora are of central importance because of the opportunity they offer, once aligned, to gain insights into the many issues of translation, or even to be used as consultation resources for translation equivalences. Results also show this importance in translation research.

4.3. Corpus size

The size of the corpus is a reference parameter in corpus methodology. However, when we tried to analyze this we came across a problem: only 11 records referred to the size of the corpus in the abstract, while 378 do not make any mention in this part of the text. Then this variable cannot be considered relevant for our study, since we cannot draw proper inferences from only 11 records. From the analysis of these 11 records it is concluded that most publications include corpus of over one million words, which is indicative of a medium-large size corpus (Berber Sardinha, 2000: 346). In addition, as shown in Table 1, one of the records used advertisements –a textual genre– to measure the size of the corpus.

Table 1. Corpus size

| Corpus size |
|-------------------|
| 100,000,000 words |
| 8,000,000 words |
| 6,500,000 words |
| 3,000,000 words |
| 2,700,000 words |
| 2,000,000 words |
| 100,000 words |
| 80,000 words |
| 64,000 words |
| 100 ads |

4.4. Translation specialization

In Translation Studies, a common distinction that is made when trying to create a typology of translation is that between specialized translation and literary translation. In our sample, 92 records are aimed at specialized translation while 41 records indicate that the study has been carried out with the purpose of studying literary translation.

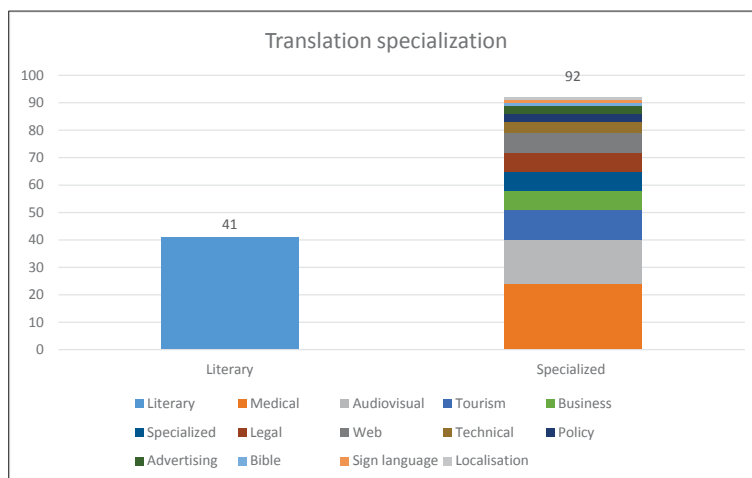


Figure 2: Translation specialization

Table 2: Number of records by subject-field in specialized translation

| Subject-field | Number of records |
|---------------|-------------------|
| Medical | 24 |
| Audiovisual | 16 |
| Tourism | 11 |
| Business | 7 |
| Specialized | 7 |
| Legal | 7 |
| Web | 7 |
| Technical | 4 |
| Policy | 3 |
| Advertising | 3 |
| Bible | 1 |
| Sign language | 1 |
| Localisation | 1 |

Within specialized translation and in accordance with our data, medical, audiovisual, tourism, legal and business occupy the first positions. These data correspond with the current and increasing demand and boom of these specialties in both the professional and the academic context.

4.5. Corpus orientation

According to the classification of the corpus orientation suggested by different authors (Baker, 1995; Laviosa, 2002; Olohan, 2004) the three main applications of corpus in translation are teaching, research and professional applications. From the records analyzed, all of them include this orientation in the title, abstract or keywords. As seen in Figure 3, the majority of records indicate research purposes; however, the boundaries between research and teaching-oriented research are not always clear, therefore this parameter cannot be interpreted appropriately without carrying out an in-depth study of the publication. Nevertheless, the remarkable aspect of this section is that the professional application of findings in academic research is almost negligible when compared to the other two aims.

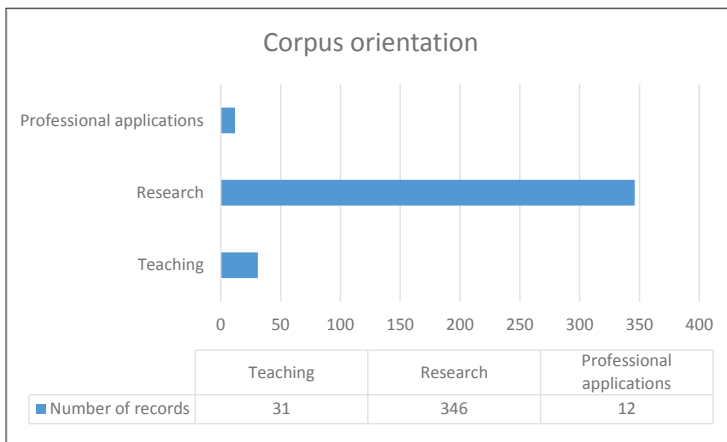


Figure 3: Corpus orientation

Within the classification of corpus use for research purposes, one of the most recurrent topics is the study of universals of translation by explicitly mentioning the term *universals* in either the title, the abstract or the keywords. This denotes a consolidated use of corpus but, at the same time, a conservative and traditional trend since most of the times this same line of research is slightly modified by other more specific research lines such as equivalence, explicitation, translation shifts, simplification and variation, among many others (see Table 3).

Table 3. Research on universals.

| Translation universals | Number of records |
|---|-------------------|
| universals | 8 |
| universals & equivalence | 1 |
| universals & explicitation | 7 |
| universals & translation shifts | 1 |
| universals & measure of distance | 1 |
| universals & non-native and Translated Language | 1 |
| universals & simplification strategies | 1 |
| universals & translation equivalents | 1 |
| universals & translationese | 1 |
| universals & variation | 1 |

Another parameter taken into account within this section on research was the observation of the specific line of research identified. In a total of 105 cases, there was a clear indication of the topic under study, however, with the

exception of universals, with 23 records, terminology, with 8, and phraseology with 4, the rest of records, 70 in total, do not show a consistent research line on a specific topic and range from dictionary making, cross-linguistic asymmetries, ESP, lexical characteristics, to translation norms.

4.6. Applications

The word ‘application’ is used here as the “a particular use that something has”. With regards to the applications of corpus to translation, we found a total of 47 records that mentioned this variable. Figure 4 below shows that quality of translation ranks first among the most common applications. Within this label we included records dealing mainly with observable criteria such as terminology coherence, omission of text passages, and internal consistency. Only one out of five focused on computer-assisted translation and machine translation, even if they are also specific applications of translation studies. The reason may lie in the fact that these topics are frequently treated from other fields of study, such as Natural Language Processing or Computational Linguistics.

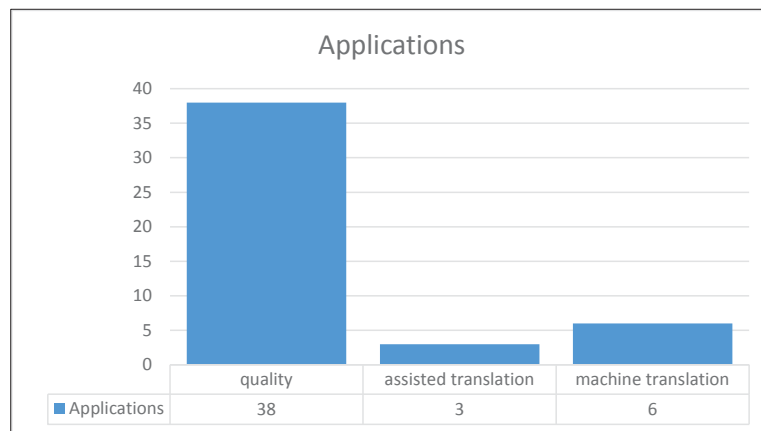


Figure 4: Applications

4.7. Type of publication

Finally, as noted in the introduction, academic publications play a key role in the establishment and consolidation of any new scientific discipline. Therefore, one of the indicators under analysis consisted in tracking and classifying the type of publication of the records selected for this study.



Figure 5: Type of publication

At first glance, there is a balance between book chapters (174) and journal articles (155). It is a well-known fact that researchers on Humanities and Social Sciences publish in a wide variety of media, but have a special predilection for monographs (books and book chapters). Moreover, the increasing number of books along with the number of theses that have emerged in recent years suggest a promising future of this methodology in translation.

5. Conclusions

In this paper, we have presented our sample and then analyzed and discussed several variables extracted for our study. We have attempted to provide a first approach to the study of seven basic variables (languages, type of corpus, corpus size, translation specialization, corpus orientation, applications and types of publications) that show the current focus on corpus methodology within translation studies.

The present study was designed to analyze with data the consolidation of corpus methods in translation and to specify which issues are under research and the features that characterize these studies. The results of this investigation show the rise and consolidation of corpus linguistics research methods applied to translation, and accurately outline the evolution of this multidisciplinary relationship, even with the assimilation of the terminology adapted from Corpus Linguistics. This research will serve as a base for future studies and can be useful as a starting point to know whether corpus linguistic methodology has consolidated in translation studies and which parameters better characterize the studies carried out using a corpus.

This research has thrown up many questions in need of further investigation. It would be interesting to carry out a study on a bibliography with a linguistic approach and observe the treatment given to translation. Further investigation will be aimed at exploiting cross-section analysis that will reveal the types of research conducted taking into account different combinations of indicators such as different language pairs, specialties and type of corpus, for example, in order to see if there is an observable trend.

References

- Baker, M. (1995). Corpora in Translation Studies. An Overview and Suggestions for Future Research. *Target* 7(2): 223-43.
- Baker, M. (1998). Réexplorer la langue de la traduction: une approche par corpus. (Investigating the Language of Translation: A Corpus-based Approach). *Meta* 43(4), 480-485.
- Baker, M. (1999). The Role of Corpora in Investigating the Linguistic Behaviour of Professional Translators. *International Journal of Corpus Linguistics* 4(2), 281-298.
- Berber Sardinha, A. P. (2000). Lingüística de Corpus: histórico e problemática. *D.E.L.T.A.* 16 (2), 323-367. Available at: <http://goo.gl/89IUT>. [June, 2013].
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Kenny, D. (1998). Corpora in Translation Studies. In M. Baker (Ed.), *Routledge Encyclopedia of Translation Studies* (pp. 50-53). London & New York: Routledge.
- Kenny, D. (1998). Creatures of Habit? What Translators Usually Do with Words. *Meta* 43(4), 515-523.
- Laviosa, S. (2011). Corpus Linguistics and Translation Studies. In V. Viana, S. Zyngier & G. Barnbrook (Eds), *Perspectives on Corpus Linguistics* (pp. 131-153). Amsterdam/Philadelphia: John Benjamins.
- Laviosa, S. (2002). *Corpus-based Translation Studies: Theories, Findings, Applications*. Amsterdam/New York: Rodopi.
- Laviosa, S. (1997). How Comparable Can 'Comparable Corpora' Be. *Target* 9(2), 289-319.
- Laviosa, S. (Ed.) (1998). *L'Approche Basée sur le Corpus/The Corpus-based Approach*. Special Issue of *Meta* 43(4).
- Leech, G. (2002). The Importance of Reference Corpora. In *Corpus lingüísticos. Presente y futuro*. Donostia: UZEL. Available at <http://goo.gl/B9FK4> [June, 2013]
- Malmkjær, K. (2003). On a pseudo-subversive use of corpora in translator training. In F. Zanettin, S. Bernardini & D. Stewart (Eds.), *Corpora in Translation Education* (pp. 119-134). Manchester: St. Jerome Publishing.
- Olohan, M. (2004). *Introducing Corpora in Translation Studies*. London/New York: Routledge.
- Rabadán, R. (2009). Corpus-based contrastive analysis and translation universals. A tool for translation quality assessment English -> Spanish. *Babel* 55(4): 303-328.
- Tymoczko, M. (1998). Computerized Corpora and the Future of Translation Studies. *Meta* 43(4), 652-659.
- Vargas-Sierra, C. (2012). La tecnología de corpus en el contexto profesional y académico de la traducción y la terminología: panorama actual, recursos y perspectivas. In M.A. Candel-Mora & E. Ortega-Arjonilla (Eds.), *Tecnología, Traducción y Cultura* (pp.67-99). Valencia: Editorial Tirant lo Blanc.
- Zanettin, F. (2012). *Translation Driven-Corpora. Corpus Resources for Descriptive and Applied Translation Studies*. Manchester: St. Jerome Publishing.