

# The corpus approach: a common way forward for Contrastive Linguistics and Translation Studies?

Sylviane Granger

University of Louvain

## Abstract

*In this introductory chapter, Granger traces the development of Contrastive Linguistics and Translation Studies over the last decades to the present day, focusing on the role of the computer corpus in giving new impetus to each field and bringing them closer together. She discusses the different types of monolingual and multilingual corpora being used in CL and TS research, proposing at the same time a common corpus terminology. She then relates the contribution of the different corpus types to the major research interests of each discipline, highlighting the complementarity of the research and calling for increased cross-fertilization and resource-pooling. She then examines the practical issues of corpus exploitation, with a review of corpus analysis tools of particular value for CL and TS and finally, the contribution of corpus-based CL and TS research in the teaching of foreign languages and translation.*

## 0. Contrastive Linguistics and Translation Studies, two converging disciplines

Although the disciplines of Contrastive Linguistics (CL) and Translation Studies (TS) cover partly common ground, it is only recently, with the emergence of corpora, that they have started to converge. This rapprochement is apparent from recent publications<sup>1</sup> and conferences that have brought together specialists from the two fields, bearing witness to the vitality of multilingual studies in general.

The history of Contrastive Linguistics has been characterized by a pattern of success-decline-success. CL was originally a purely applied enterprise, aiming to produce more efficient foreign language teaching methods and tools. Based on the general assumption that difference equals difficulty, CL, which in those days was called Contrastive Analysis (CA), consisted in charting areas of similarity and difference between languages and basing the teaching syllabus on the contrastive findings. Advances in the understanding of Second Language Acquisition (SLA) mechanisms led to a questioning of the very basis of CA. Interlingual factors were found to be less prevalent than other factors, among which intralingual mechanisms such as the overgeneralization of target rules and external factors such as the influence of teaching methods or personal factors like motivation. This led to the decline of CA, but not to its death. At first, it gave rise to some drastic pedagogical decisions, which in some cases culminated in a total ban of the mother tongue in FL teaching. But research (see Odlin 1989, Selinker 1992, James 1998) re-established transfer as a major – if not **the** major – factor in SLA, which in turn led to a progressive – albeit limited – return of contrastive

considerations in teaching. More importantly, the questioning of the contrastive approach to FL teaching did not impede its extension to other fields. The globalisation of society led to an increased awareness of the importance of interlingual and intercultural communication and played a major role in the revival of CL. Another factor which helped boost contrastive studies was the emergence and rapid development of corpus linguistics and natural language processing, which are increasingly focusing on cross-linguistic issues. Large bilingual corpora such as the English/French Hansard Corpus<sup>2</sup> gave contrastive linguists and NLP specialists a much more solid empirical basis than had ever been previously available. Previous research had been largely intuition-based. Vinay & Darbelnet (1977) and Malblanc (1968) are well-known exemplars of this type of approach. As the authors had an excellent knowledge of the languages they compared, these books contain a wealth of interesting contrastive statements. However, intuitions can be misleading and a few striking differences can lead to dangerous over-generalisations. For instance, the absence in English of connectors corresponding to the French 'or' or 'en effet' has led to the general conclusion that French favours explicit linking while English tends to leave links implicit (Vinay & Darbelnet 1977: 222, Newmark 1988: 59, Hervey & Higgins 1992: 49). This contrastive claim still awaits empirical investigation and was not confirmed by a preliminary investigation based on an English-French *journal* corpus (Anthonie 1996). Contrastive linguists now have a way of testing and quantifying intuition-based contrastive statements in a body of empirical data that is vastly superior – both qualitatively and quantitatively – to the type of contrastive data that had hitherto been available to them.

Although the practice of translation, and reflection on the nature of the translation process, go far back in history, to the times of Cicero and St Jerome, Translation Studies did not emerge as a separate academic discipline until the 1980s. It was Holmes (1988) that put forward an overall framework for the new field in his seminal article 'The name and nature of translation studies'. Holmes and many other translation specialists expressed the same dissatisfaction with introspective methods as that felt by contrastive linguists and advocated recourse to large bodies of translated texts. The recent emergence of a corpus-based trend within the field of translation studies can be seen as a direct consequence of a major change of perspective in TS that displaced the research focus from the source text to the target text. While for years research had focused on equivalence with the source text (semantic, connotative, stylistic, etc.), theorists like Toury (1980, 1995) advocated a target orientation and turned translation studies into a descriptive endeavour to specify probabilistic laws of translation. This major changeover is described as follows by Venuti (2000: 123): "The literature on equivalence formulates linguistic and textual models and often prescribes a specific translation practice (pragmatic, functional, communicative). The target orientation, in contrast, focuses on actual translations and submits them to detailed description and orientation. It inspires research projects that involve substantial corpora of translated texts".

It was Mona Baker who pioneered the corpus-based trend in the early 90s. She laid down the agenda for corpus-based TS (1993 and 1995) and started

collecting corpora of translated texts with a view to uncovering the distinctive patterns of translation. Her investigations brought to light a number of potential 'translation universals' (Baker 1993) which further corpus studies are helping to confirm or disprove (see Puurtinen this volume).

In short, researchers in CL and TS have come to rely on corpora to verify, refine or clarify theories that hitherto had had little or no empirical support and to achieve a higher degree of descriptive adequacy. At the same time however, they do not always use exactly the same types of corpus and do not have the same research objectives, differences which are examined in section 1 below.

## 1. Corpora in cross-linguistic research

In the corpus, scholars of contrastive linguistics and translation studies now have a common resource. Unfortunately, like in many new scientific fields, the terminology has not yet been firmly established, leading to a great deal of confusion.

Contrastive linguists distinguish between two main types of corpus for use in cross-linguistic research:

- corpora consisting of original texts in one language and their translations into one or more languages – let us call these *translation corpora*;
- corpora consisting of original texts in two or more languages, matched by criteria such as the time of composition, text category, intended audience, etc. – let us call these *comparable corpora*. (Johansson & Hasselgård 1999).

It should be noted however, that even among contrastive linguists the terminology is not entirely consistent. The term *parallel corpus* is sometimes used to refer to a comparable corpus (Aijmer et al 1996: 79, Schmied & Schäffler 1996: 41), a translation corpus (Hartmann 1980: 37) or a combined comparable/translation corpus (Johansson et al 1996).

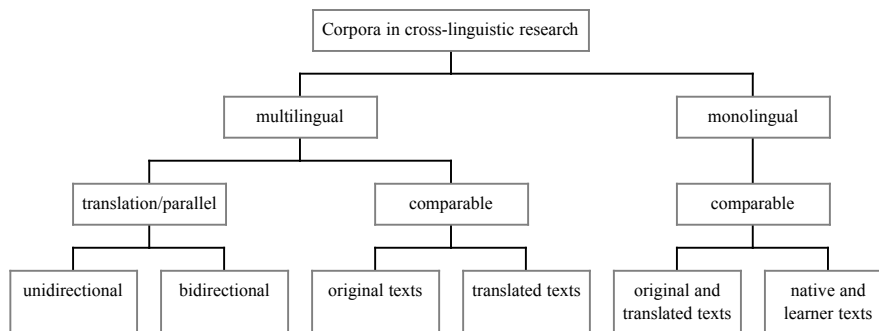
Contrastive linguists are keen to point out that the two types of corpus should be used concurrently as each has its advantages and disadvantages (see Johansson this volume). Comparable corpora have the major advantage of representing original texts in the two or more languages under comparison, i.e. language spontaneously produced by native speakers of those languages. They are therefore in principle free from the influence of other languages,<sup>3</sup> which is obviously not the case of translation corpora as the original source text is in a different language and will quite naturally exert some kind of influence on the target text. The main drawback of comparable corpora lies in the difficulty of establishing comparability of texts. Some types of text are culture-specific and simply have no exact equivalent in other languages. Translation corpora are an ideal resource for establishing equivalence between languages since they convey the same semantic content. The main drawback of translation corpora, however,

is that they often display traces of the source text and therefore cannot really be considered as reliable data as regards the target language, especially in frequency terms. In addition, it is not always possible to find translations of all texts, either because of the text type – letters and e-mail messages, for instance, are not usually translated – or because there are more translations in one direction (English to Norwegian, for instance) than in another (Norwegian to English).

TS researchers use the terms *translation corpus*, *parallel corpus* and *comparable corpus* to cover different types of texts. The term *comparable corpus* is used to refer to ‘two separate collections of texts in the same language: one corpus consists of original texts in the language in question and the other consists of translations in that language from a given source language or languages’ (Baker 1995: 234). The term *translation (or translational) corpus* is used to refer to the corpus of translated texts (see Baker 1999 and Puurtinen this volume). While in standard CL terminology, comparable corpora are usually multilingual (comparable original texts in different languages), in TS terminology they are usually monolingual (original and translated texts in the same language). Within the TS framework the term *parallel corpus* usually refers to ‘corpora that contain a series of source texts aligned with their corresponding translations’ (Malmkjaer 1998: 539), in other words what contrastive linguists usually refer to as translation corpora.

Over and above the terminological difference, there is a more fundamental discrepancy between the two cross-linguistic approaches. In the TS framework, translated texts are considered as texts in their own right, which are analysed in order to “understand what translation is and how it works” (Baker 1993: 243). In the CL framework they are often presented as unreliable as the cross-linguistic similarities and differences that they help establish may be ‘distorted’ by the translation process, i.e. may be the result of interference from the source texts.

The present volume reflects the terminological diversity that characterises current cross-linguistic research. However, we feel that unified terminology is desirable and would like to suggest the general typology illustrated in Figure 1. In this typology, a primary distinction is made between multilingual and monolingual corpora. *Multilingual corpora* involve more than one language. They may be of two main types: (a) *translation (or: parallel) corpora*<sup>4</sup> (which contain source texts and their translations and may be unidirectional – from language X to language Z – or bi/multidirectional) and (b) *comparable corpora* (which contain non-translated or translated texts of the same genre). The *monolingual corpora* relevant for cross-linguistic research are all comparable corpora. They may contain (a) *original and translated* texts in one and the same language or (b) *native and learner texts* in one and the same language.<sup>5</sup> In this typology, the term *parallel corpus* is used unambiguously to refer to corpora of source texts and their translations. It is not used to refer to comparable corpora or as a generic term to refer to any type of multilingual corpus (Teubert 1996: 245).



**Figure 1.** Corpora in cross-linguistic research

This diagram does not include the many extralinguistic features that influence the data and therefore need to be carefully recorded, such as the translator’s status (professional or student) or the direction of the translation process (into the translator’s mother tongue or not).

**2. Types of corpus-based comparison**

With these different corpus types, a variety of comparisons can be undertaken. Table 1 presents an overview of the different types of cross-linguistic comparison and the disciplines within which they are undertaken (see also Johansson, this volume)

	Type of comparison	Type of corpus	Discipline
1	$OL_x \Leftrightarrow OL_y$	Multilingual comparable corpus of original texts	CL
2	$SL_x \Leftrightarrow TL_y$	Multilingual translation/parallel corpus	CL & TS
3	$SL_x \Leftrightarrow TL_x$	Monolingual comparable corpus of original and translated texts	TS & CL
4	$TL_x \Leftrightarrow TL_y$	Multilingual comparable corpus of translated texts	TS

OL = original language  
 SL = source language  
 TL = translated language

**Table 1.** Types of corpus-based cross-linguistic comparison

The first type of comparison, between corpora of original texts in different languages (x and y), is the CL domain of expertise par excellence. However, there is a growing awareness among TS researchers of the interest of this type of research for translation studies. The second type of comparison is the most obvious meeting point between CL and TS. Researchers in both fields use the same resource but to different ends: uncovering differences and similarities between two (or more) languages for CL and capturing the distinctive features of the translation process and product for TS. The third type of comparison, which contrasts original and translated varieties of one and the same language, is the ideal method for uncovering the distinctive features of translated texts and hence seems at first sight to fall exclusively within TS. However, this type of comparison is increasingly being used by CL researchers who interpret differences between OL and TL as indirect evidence of differences between the languages involved (see Johansson & Hasselgård 1999 and Johansson this volume). Finally, the comparison of translated varieties in different languages is quite clearly the prerogative of TS. However, it is essential that contrastive linguists pay attention to this type of study. Failing to properly understand the nature of translated texts might lead them to attribute some difference between OL and TL to interference from OL when in fact the phenomenon may simply be a manifestation of a translation universal.

### 3. Limitations of a corpus-based approach in cross-linguistic studies

While it is undeniable that both CL and TS stand to gain from using corpora, it is important to bear in mind that a corpus-based methodology has its limitations. In the following lines, we will focus on two of these: one more practical and hopefully only temporary, the limited availability of corpora for cross-linguistic research, and the other more fundamental, the corpus predilection for form-based research.

#### 3.1 Availability

A survey of current cross-linguistic research shows that quite a number of CL and TS researchers still carry out their research entirely manually. This may be partly blamed on a computer phobia, but in the majority of cases, the main reason is simply the limited availability of computerised corpora.

The most easily accessible corpora for cross-linguistic research are undoubtedly comparable corpora of original languages. English is particularly well equipped with large balanced corpora such as the *British National Corpus* or the *Bank of English*. For other languages, such as Spanish, Dutch, Italian, Greek, Portuguese, Swedish, Czech and Chinese, there are electronic text collections that are regularly used for cross-linguistic research, but on the whole these tend to be less representative than the English mega corpora. Less widespread languages, however, may not have any corpus resources at all or access to them may be severely limited.<sup>6</sup>

For the other types of corpus, however, electronic resources are scarce. Comparable corpora of translated language only exist for a small number of languages (English and Finnish, see Baker 1999<sup>7</sup> and Puurtinen this volume respectively), and while we are witnessing a rapid growth in the number of bilingual and multilingual resources, some of which can even be explored online (see section 4.1), many high quality resources remain inaccessible to the academic community. This is the case, for instance, of the excellent English-Norwegian and English-Swedish corpora, which are only available to a limited group of researchers because of copyright restrictions. Available parallel corpora tend to include older, copyright free texts or alternatively, highly specialised texts such as documents from the European Union or the World Health Organization, the disadvantage of which is that it is often impossible to determine the source and target languages, a major variable for both CL and TS studies.

The dearth of parallel corpora accounts for the relatively large number of cross-linguistic studies which are corpus-based in the sense that they rely on authentic texts rather than introspective methods, but not in the more usual sense of computer corpus-based. Hantson and Chuquet (both this volume) belong to this category. Another way of circumventing the lack of parallel corpora is to use comparable corpora of original texts and to rely on native speakers' intuitions to get at the translation in the other language. As shown by Davidse & Heyvaert's and Charteris-Black's studies in this volume, this approach, though clearly a stopgap measure in the face of a lack of electronic resources, can lead to excellent results.

### 3.2 Form-based approach

Apart from the question of corpus availability, there is another factor that may lead researchers to opt for manual work, namely the very object of investigation. If the research focus is a particular form – the preposition *in*, the adverb *actually* or all adverbs ending in *-ly* –, it is highly advantageous to have an electronic corpus and a monolingual text retrieval software tool such as *WordSmith Tools*,<sup>8</sup> which makes it possible to extract all the occurrences of a given form quickly and reliably and to visualise them in context, or a bilingual concordancer (see section 4.1). However, if the research focus is on a semantic category such as agency or causality, or a syntactic structure, such as complex noun phrases, automatic retrieval becomes more difficult, if not simply impossible. In some cases, attempts at automatic retrieval bring up so much irrelevant material that researchers may prefer to opt for a purely manual investigation (see Puurtinen's study of non-finite clauses in this volume). In addition, automatic retrieval and subsequent concordance display is not ideally suited for phenomena such as tense usage, the analysis of which requires a much larger context (see e.g. Chuquet's study of the 'imparfait', this volume).

Even if use of corpora should be encouraged, it should not be felt as a necessity. What matters is the use of solid empirical data, whether electronic or not. The last thing we want is for researchers to organise their investigations around the existence of electronic corpora or the capabilities of the computer.

## 4. Applications and tools

### 4.1 Multilingual software tools

Exploitation of large multilingual corpora necessitates the development of new multilingual software tools. The most important such tool is the bilingual or multilingual concordancer, a software tool that allows researchers to retrieve all occurrences of any given form – morpheme, word, phrase, sentence – in one language and to align them with their translations in another language. Two such programs are available: *MultiConcord* (Woolls 1997, Corness 2002 & King this volume) and *Paraconc* (Barlow 1995 & 1999). Both tools are extremely useful for cross-linguistic research, but they differ in several important respects. *Multiconcord* has an inbuilt sentence aligner and a language test creation tool, but only limited sort and search facilities; *Paraconc* on the other hand, whilst it has the disadvantage of requiring sentence-aligned texts, has more advanced sort and search facilities.<sup>9</sup> Both tools require some pre-editing of the texts (paragraph alignment for *Multiconcord* and sentence alignment for *ParaConc*), which undoubtedly impedes more widespread use of these tools. A commercially available, fully automatic multilingual tool combining the functions of sentence aligner and corpus explorer has yet to be produced. Fortunately, there are also some web-based concordancers that allow researchers to explore bilingual corpora online.<sup>10</sup>

### 4.2 Teaching

Corpus-based analyses have led to a new inductive teaching methodology, called data-driven learning (DDL), which Johns and King (1991: iii) describe as “the use in the classroom of computer-generated concordances to get students to explore regularities of patterning in the target language, and the development of activities and exercises based on concordance output”. DDL can involve monolingual corpora and/or bilingual corpora and be integrated into foreign language teaching – both general and for specific purposes (LSP) – and translation teaching.

In general foreign language teaching, aligned corpora are used in consciousness-raising exercises. Learners are presented with examples of language features such as modals, prepositions, conjuncts or pronouns in the source language and their aligned translations in the target language and asked to reflect on the interlingual similarities and differences. This inductive stage is usually followed by a series of activities, which take the form of corpus-based fill-in exercises where either the search item or the aligned translation has been blanked out.<sup>11</sup> A good example of a DDL approach to grammar teaching is the *Online Chemnitz Internet Grammar*, which makes extensive use of the English-German Translation Corpus.<sup>12</sup> Although DDL is often presented as more ideally suited for advanced proficiency levels, St. John’s (2001) experiment suggests that it can also be of great benefit for beginners.

Bilingual corpora are also an extremely valuable pedagogical resource in LSP. Kübler and Foucou (this volume) demonstrate the important contribution of technical English corpora – both monolingual and bilingual (English-French) – in describing the use of verbs in Computer Science and preparing pedagogical material tailor-made for French-speaking ESP students.

According to Bernardini (1997), “one of the reasons why translation teaching as it is generally understood (exercise and correction) is often perceived as ineffective and tentative, is that it still lacks a solid pedagogic background”. In her view, traditional exercises in translation should go hand-in-hand with corpus-based learning activities which develop the skills that are immediately relevant for the education of translators, in terms of awareness, reflectiveness and resourcefulness. Corpus-based classroom activities for translator trainees may involve comparable and parallel corpora of general or specialised language. Zanettin (1998: 618-620) demonstrates how small comparable corpora of English and Italian help learners to compare the behaviour of similar discourse units in the two languages and select the translations which best adhere to the linguistic and genre conventions of the receiving culture. In order to test the effectiveness of corpus-based methods in translation training, Bowker (1998 and this volume) assigns the same translation task to two groups of translator trainees – one using a specialised monolingual corpus, the other conventional reference tools – and finds that the former outperformed the latter in several major respects, notably subject-field understanding, correct term choice and idiomatic expression. Admittedly, there is as yet little empirical evidence of this sort. In spite of this, Aston (1999) is optimistic about the future of corpora as translation and learning tools: “It is our experience at Forli that few trainee-translators who have used corpora would wish to be without them, notwithstanding (or because of?) the investment in time and effort required to compile corpora and to learn how to use them, and we expect that, as the number of available corpora and the quantity of suitable software increases, the use of corpora for translation and translator-training will gather further momentum, with a growth in its cost-effectiveness”.

## **5. Conclusion**

By virtue of their object of study, the fields of Contrastive Linguistics and Translation Studies share a great deal of common ground: they “are interested in seeing how ‘the same thing’ can be said in other ways, although each field uses this information for different ends” (Chesterman 1998: 39). The corpus has the potential to bring the two fields even closer together as both CL and TS researchers now rely on the same type of data, use the same software tools and are partly interested in the same corpus-based applications, notably reference materials – dictionaries, grammars – and teaching methods. Unfortunately, as rightly stated by Chesterman (1998: 6), “Although these are neighbouring disciplines, it nevertheless often appears that theoretical developments in one field are overlooked in the other, and that both would benefit from each other’s insights”. In particular, lack of familiarity with TS findings may lead CL

researchers to interpret their data in terms of differences between language systems when they result from translation norms or strategies, while TS researchers may similarly misinterpret their data because of a lack of awareness of a systematic difference between the two language systems established by CL. Another more practical reason that should lead researchers in the two fields to cooperate is the shortage of corpora, which considerably hinders cross-linguistic research. We need more and better corpora for cross-linguistic research and as data collection is very time-consuming, there is a great deal to be gained from joining forces. If CL and TS pool knowledge and resources, one can safely predict a bright future for corpus-based cross-linguistic research and applications.

### Notes

- 1 Bibliographies such as the Contrastive Linguistics bibliography (<http://bank.rug.ac.be/contragram/biblio.html>) and the Bibliography of Translation Studies (<http://aix1.uottawa.ca/~lbowker/bibtsweb/bibts.htm>), for instance, include references to publications in both fields and the newly set up scientific journal *Languages in Contrast* explicitly states its aim to promote interdisciplinary studies, 'particularly those that make links between contrastive linguistics and translation'. The same holds true for the special issue of the translators' journal *META* devoted to the corpus-based approach (Laviosa 1998) and the recent Rodopi volume on corpus-based translation studies (Laviosa 2002).
- 2 The Hansard Corpus can be obtained from the Linguistic Data Consortium: <http://www ldc.upenn.edu/Catalog/LDC95T20.html>.
- 3 This is obviously not entirely true. French newspaper texts for example have often been found to contain traces of the (usually) English texts on which the journalists have based their articles.
- 4 If, following TS terminology, the term 'translational corpus' is retained to refer to corpora of translated texts, it would be better to use the term 'parallel corpus' to refer to corpora of source texts and their translations in order to avoid confusion with the term 'translation corpus'. One additional reason that might tip the scales in favour of the term 'parallel corpus' is that this has become the standard term to refer to collections of texts aligned with their translations in cross-linguistic NLP research.
- 5 For a description of this special type of contrastive research called *Contrastive Interlanguage Analysis*, see Granger 1996.
- 6 One such example is the Malay corpus owned by the Malaysian Language Planning Agency which can only be consulted on site (see Charteris-Black this volume).
- 7 Baker's Translational English Corpus is freely available and can be explored on line: <http://tec.ccl.umist.ac.uk/tec/>

- 8 A demo version can be downloaded from Mike Scott's website: <http://www.lexically.net/wordsmith/version3/index.htm>
- 9 For more information on Multiconcord, see <http://web.bham.ac.uk/johnstf/paracon.htm>; on ParaConc: <http://www.ruf.rice.edu/~barlow/pc.html>
- 10 The following bilingual corpora can be explored online: English-Slovene corpus: <http://nl2.ijs.si/corpus/index-bi.html>; English-Portuguese corpus: <http://www.portugues.mct.pt/COMPARA/>; English-French corpus: <http://www-rali.iro.umontreal.ca/ProjetTransSearch.fr.html>
- 11 Sample handouts can be downloaded from the DDL websites of Tim Johns ([http://web.bham.ac.uk/johnstf/ddl\\_lib.htm](http://web.bham.ac.uk/johnstf/ddl_lib.htm)) and Joseph Rézeau (<http://www.uhb.fr/campus/joseph.rezeau/concord.htm>).
- 12 <http://www.tu-chemnitz.de/phil/InternetGrammar/>. See also Schmied (forthcoming).

### References

- Aijmer, K., B. Altenberg, and M. Johansson (eds). 1996. *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies*. Lund Studies in English 88. Lund: Lund University Press.
- Aijmer, K., B. Altenberg, and M. Johansson. 1996. Text-based contrastive studies in English. Presentation of a project. In Aijmer et al (eds) *Languages in Contrast: 73-85*
- Altenberg, B., and S. Granger (eds). 2002. *Lexis in Contrast. Corpus-based approaches*. Studies in Corpus Linguistics 7. Amsterdam & Philadelphia: Benjamins.
- Anthone, A. 1996. *Patterns of French/English Conjunct Use in Original and Translated Journalese*. Unpublished MA Dissertation. Louvain-la-Neuve: Université catholique de Louvain.
- Aston, G. 1999. Corpus use and learning to translate. *Textus* 12: 289-314. Also available online: <http://www.sslmit.unibo.it/guy/textus.htm>
- Baker, M. 1993. Corpus Linguistics and Translation Studies. Implications and Applications. In M. Baker, G. Francis, and E. Tognini-Bonelli (eds) *Text and Technology*, 233-250. Amsterdam & Philadelphia: Benjamins.
- Baker, M. 1995. Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target* 7(2): 223-243.
- Baker, M. 1999. The Role of Corpora in Investigating the Linguistic Behaviour of Professional Translators. *International Journal of Corpus Linguistics* 4(2): 281-298.
- Barlow, M. 1995. ParaConc: A concordancer for parallel texts. *Computers and Texts* 10 (CTI Textual Studies). Also available from <http://www.ruf.rice.edu/~barlow/para-ox.html>
- Barlow, M. 1999. MonoConc 1.5 and ParaConc, *International Journal of Corpus Linguistics*, 4(1): 319-327.

- Bernardini, S. 1997. A 'trainee' translator's perspective on corpora. Corpus use and learning to translate, 14-15 November. Available online <http://www.sslmit.unibo.it/cultpaps/trainee.htm>
- Bowker, L. 1998. Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study. In S. Laviosa (ed.) *L'approche basée sur corpus/The Corpus-based Approach*, 631-651.
- Chesterman, A. 1998. *Contrastive Functional Analysis*. Benjamins: Amsterdam & Philadelphia.
- Corness, P. 2002. Multiconcord: A computer tool for cross-linguistic research. In Altenberg & Granger (eds) *Lexis in Contrast. Corpus-based approaches*, 307-326.
- Granger, S. 1996. From CA to CIA and back: An integrated approach to computerized bilingual corpora and learner corpora. In Aijmer et al (eds) *Languages in Contrast*, 37-51.
- Granger, S., L. Beheydt, and J.P. Colson (eds). 1999. *Contrastive Linguistics and Translation*. Special issue of *Le Langage et l'Homme* 34(1). Leuven: Peeters.
- Hartmann, R. 1980. *Contrastive Textology*. Heidelberg: Julius Groos Verlag.
- Hervey, S., and J. Higgins. 1992. *Thinking Translation*. London: Routledge.
- Holmes, J. 1988. The Name and Nature of Translation Studies. In J. Holmes (ed.) *Translated. Papers on Literary Translation and Translation Studies*, 67-80. Amsterdam: Rodopi.
- James, C. 1998. *Errors in Language Learning and Use. Exploring error analysis*. London & New York: Longman.
- Johansson, S., J. Ebeling, and K. Hofland. 1996. Coding and aligning the English-Norwegian Parallel Corpus. In Aijmer et al (eds) *Languages in Contrast*, 87-112.
- Johansson S., and H. Hasselgård. 1999. Corpora and cross-linguistic research in the Nordic countries. In Granger et al (eds) *Contrastive Linguistics and Translation*, 145-162.
- Johns, T., and P. King (eds). 1991. *Classroom Concordancing. ELR Journal* (New Series) 4.
- King P., and D. Woolls. Creating and using a multilingual parallel concordancer. Available online: <http://web.bham.ac.uk/johnstf/paracon.htm>
- Laviosa, S. 1998. L'approche basée sur corpus/The Corpus-based Approach. Special issue of *META. Journal des Traducteurs* 43(4): 473-659.
- Laviosa, S. 2002. *Corpus-based Translation Studies. Theory, Findings, Applications*. Amsterdam & New York: Rodopi.
- Malblanc, A. 1968. *Stylistique comparée du français et de l'allemand*. Paris: Didier.
- Malmkjaer, K. 1998. Love thy Neighbour: Will Parallel Corpora Endear Linguists to Translators? In S. Laviosa (ed.) *L'approche basée sur corpus/The Corpus-based Approach*: 534-541.
- Newmark, P. 1988. *A Textbook of Translation*. Englewood Cliffs: Prentice-Hall.
- Odlin, T. 1989. *Language Transfer. Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.

- Schmied, J. Forthcoming. Learning English Prepositions in the Chemnitz Internet Grammar. In Granger, S. & Petch-Tyson, S. (eds) *Extending the Scope of Corpus-based Research: New Applications, New Challenges*. Amsterdam & Atlanta: Rodopi.
- Schmied, J., and H. Schäffler. 1996. Approaching translationese through parallel and translation corpora. In C. Percy, C. Meyer & I. Lancashire (eds) *Synchronic Corpus Linguistics*, 41-56. Amsterdam & Atlanta: Rodopi.
- Selinker, L. 1992. *Rediscovering Interlanguage*. London & New York: Longman.
- St. John, E. 2001. A case for using a parallel corpus and concordancer for beginners of a foreign language. *Language Learning and Technology* 5(3): 185-203. Available online: <http://llt.msu.edu/vol5num3/stjohn/default.html>
- Teubert, W. 1996. Comparable or parallel corpora? *International Journal of Lexicography* 9: 238-264.
- Toury, G. 1980. *In search of a theory of translation*. Tel Aviv: Porter Institute.
- Toury, G. 1995. *Descriptive Translation Studies and beyond*. Amsterdam & Philadelphia: Benjamins.
- Venuti, L. (ed.). 2000. *The Translation Studies Reader*. London & New York: Routledge.
- Vinay, J.P., and J. Darbelnet. 1977. *Stylistique comparée du français et de l'anglais*. Paris: Didier.
- Woolls, D. 1997. MultiConc (version 1.0), CFL Software Development, Birmingham.
- Zanettin, F. 1998. Bilingual Comparable Corpora and the Training of Translators. In S. Laviosa (ed.) *L'approche basée sur corpus/The Corpus-based Approach*, 616-630.