

Jarmo Jantunen

Comparable Corpora in Translation Studies: Strengths and Limitations

Abstract

This article deals with the advantages and limitations of so-called comparable corpora in Translation Studies. Comparable corpora are text collections in two separate subsets in the same language: one subcorpus consists of non-translated (original) material in language A, and the other consists of translations produced into the same language. This article will focus on phenomena such as representativeness, objectivity, applications in translation training and translators' work as well as linguistic approaches in translation studies. Two translational corpora, namely the Translational English Corpus (TEC) and the Corpus of Translated Finnish (CTF), are studied as examples of corpora that have already been used in analyses in the field of Corpus-based Translation Studies.

Keywords: Corpus-based Translation Studies, Translational English Corpus, Corpus of Translated Finnish, comparable corpus, representativeness

1. Introduction

In the last few decades, large electronic databases of real language have provided new insights into several research questions in linguistics. The discipline known as corpus linguistics has been a "source of evidence for improving descriptions of the structure and use of languages, and for various applications, including the processing of natural language by machine and understanding how to learn or teach a language" (Kennedy 1998: 1). The approach is descriptive and focuses on actual language use by studying corpora consisting of real texts that have occurred naturally in written or spoken form. In translation studies, a descriptive approach emerged in the 1970's (Hermans 1999: 7) and has been one of the major branches ever since. The adoption of corpus linguistic methods has tended to bring about a growth of descriptive and empirical studies in the field, developing an approach of its own, namely Corpus-based Translation Studies (CTS) (see e.g. Kenny 2001).

The use of computer corpora in CTS began in the middle of the 1980's (e.g. Canadian Hansard Corpus), but it was in the 1990's that interest in translational corpora grew. One of the earliest widely known translational corpora is the English-Norwegian Parallel Corpus (Johansson 1998), which has been used in contrastive studies and in analyses of translations between English and Norwegian. Another, already very well-known corpus is the Translational English Corpus (TEC) (see e.g. Laviosa 1997; Olohan and Baker 2000) held at UMIST in Manchester, which has mainly been used in studies focusing on differences between translated and non-translated language. Finally, I would like to mention a Finnish translational corpus, namely the Corpus of Translated Finnish (CTF) (Mauranen 2000), compiled at Savonlinna School of Translation Studies at the University of Joensuu. This article discusses these two latter corpora and also their advantages and restrictions. The aim is not, however, to compare the two corpora in terms of their suitability for analyses of translated language. Rather, it aims to discuss the strengths and limitations of comparable corpora by using TEC and CTF as examples of existing and already utilized corpora. The discussion will concentrate on the corpora and the methodology, not, for example, on the software being used in analyses or on any particular analysis.

2. Corpora in Translation Studies

In linguistics, the need to study authentic texts to describe the essence of language used in real communicative events began after several decades of theoretical analysis, which had dominated the field. In translation studies, a similar need rose after a period of theoretical and especially prescriptive research, as the interest of scholars shifted toward analyses of translated texts as such. In both fields, the demand for descriptive analyses created a need to compile electronic corpora that could be analysed with computer software. The demand for corpus-based research was launched by Mona Baker (1993, 1995), who stresses not only the significance of access to a large amount of running text but also the importance of corpora in developing research on translations and translating. According to Baker (1993: 243), "the most important task that awaits the application of corpus techniques in translation studies... is the elucidation of the nature of translated text as a mediated communicative event." This can be done, for example, by identifying and analysing hypotheses about wide-spread linguistic tendencies in translations with the help of corpora of translated and non-translated texts in the same language.

Baker (1995) proposes three types of corpora that are of specific interest to scholars in translation studies: parallel, multilingual and comparable corpora. A *parallel* corpus consists of source texts written in language A and their translations in language B, thus it is usually a bilingual corpus¹. Owing to their inherent structure, parallel corpora can be seen as 'text-bound' corpora: both (or all) subsets contain different versions of the same texts, original texts in the one subset and their translations in the other. Parallel corpora, like the Hansard Corpus mentioned above and the German-English Parallel Corpus of Literary Texts (GEPCOLT, Kenny 2001) are suitable, for example, for studying translational behaviour between particular language pairs and the equivalence of linguistic features between SL and TL texts (Kenny 1998: 51). *Multilingual* corpora, in turn, consist of monolingual subcorpora in different languages and can be used, for example, to analyse the natural patterns of the target language in question and hence to train translators. Finally, *comparable* corpora, which Baker takes a special interest in, are corpora that consist of "two separate collections of texts in the same language" (Baker 1995: 234). In other words, a comparable translational corpus contains two subcorpora: one subcorpus which is compiled of original, spontaneously produced texts in the language in question and another subcorpus that consists of translations in that language.

In the following sections, I will concentrate on comparable corpora. But first, I will briefly describe the structure of two existing translational corpora that have already been used in comparable analyses of translation, namely the Translational English Corpus (TEC) compiled at UMIST in England and the Corpus of Translated Finnish (CTF) compiled in the project "Translated Finnish and Translation Universals: A Corpus Study" at Savonlinna School of Translation Studies in Finland (for CTF see Mauranen 2000). At the time of writing, the TEC is a 7-million-word corpus² of texts translated into English from several languages, both European (French, German, Spanish, Portuguese, Italian, Welsh, Polish) and non-European languages (Arabic, Hebrew, Thai). It consists of subcorpora of fiction, biography, inflight magazines and news. Texts included in the corpus were published mainly in the 1990's: according to the headers in the corpus, about 80% of the texts were published in the 1990's and about 18% in the 1980's. The Finnish CTF

¹ Parallel corpora can also be multilingual if they contain translations from texts of one source-language into several target languages (Kenny 2001: 62).

² The TEC should reach twenty to thirty million words by the end of 2003 (Olohan and Baker 2000: 151).

consists of four main text categories: fiction, academic prose, popular science and children's literature. Furthermore, it contains a few publications that belong to text categories such as biography, detective story and light reading. The texts were published in 1995–2000. The completed size of the CTF is 9.6 million words: 5.8 million words in the subcorpus of translated Finnish and 3.8 million words in the subcorpus of non-translated Finnish. The source languages are: Indo-European languages like English, Russian, German, French, Spanish, Dutch, Norwegian and Swedish as well as Finno-Ugric languages like Hungarian and Estonian.

3. Advantages and limitations of comparable corpora

3.1. Comparable corpora in general

According to Baker (1996: 175; 1999: 282), corpus linguistics has excluded translations from monolingual corpora on the basis that translations are not representative of the language being studied. This means that the number of corpora including translated texts has been very limited. Comparable corpora are thus an important addition to the small group of translational corpora – which have merely been parallel or multilingual until the 1990's. The most important advantage that comparable corpora have over the other types of translational corpora is that they can be used to study the specific characteristics of translated language in comparison with non-translated language. The characteristics of translated language manifest themselves as so-called translation universals (Baker 1993; 1996; Laviosa-Braithwaite 1995) or wide-spread tendencies or typicalities in translations (Chesterman 2001), meaning features that typically occur in translated language and that exist in translations regardless of the source or target languages involved (Baker 1995: 234). Tendencies such as explicitation, textual conventionality and simplification, to name but a few, are classified as universals in this sense. Earlier these phenomena have been studied without computer corpora, and it was the results of the manual studies that generated the hypotheses for the corpus-based investigations that are currently being carried out.

In addition to the applications described above, there are also other, more theory-related, reasons to compile monolingual comparable corpora: first of all, comparison of translations and non-translations indicates that translations are worth studying as such. In compiling comparable (and other translational) corpora, it is implied that translated texts contain interesting linguistic features to be investigated. In linguistics, the attitude towards translations has been

paradoxical: by excluding them from large general corpora, linguists imply that translations are special in one way or another, but at the same time, they have typically had no interest in investigating translations. However, the situation is gradually changing: for example, in Finland, linguists studying the Finnish language are aware of the value of studying translations and translated language. This is because it has increasingly been understood that a) translations have had an important role in the development of written Finnish language throughout history, b) translations are published in ample measure and widely consumed in Finland, and c) as they are part of everyday life, translations may affect not only the language itself but also our way of thinking and seeing the world.

One of the principles of (corpus) linguistic analyses is the attempt to reach maximum objectivity. Objectivity can be reached in empirical studies either by using a large amount of text material (corpora) or by utilizing software programs that are capable of analysing the material automatically or semi-automatically and that perhaps use statistical procedures. People involved with translations (such as scholars, publishers and ordinary readers) usually have an intuition-based idea of what translations are and should be like. These subjective interpretations are often based on observations of how texts are translated – and very often these interpretations focus on cases where some aspects of translations seem to be odd or wrong. Here the value of translational corpora is obvious: they enable us to gain more objective information on what translations and translated language are really like. But, although corpora and statistics reduce the element of intuition in analyses, we must keep in mind that the interpretations of the findings still remain subjective.

The problem of objectivity may also arise when a corpus is compiled. According to Tymoczko (1998: 654), “corpora in translation studies are products of human minds, of actual human beings, and, thus, inevitably reflect the views, presuppositions, and limitations of those human beings” (see also Kenny 2001). This is linked to the question of representativeness, which is one of the most important methodological issues in corpus compilation. The actual texts included in the corpus inevitably affect the perception of the language or language variant that the corpus represents. This is due to the basic notion of corpus studies: corpora are investigated in order to make generalizations that concern a language or language variant as a whole. In translation studies this is related to questions like: What are the modes of translations (written translations or interpreting, human or machine translations)? Which genres or text types are included in corpora? What is the

range of the source languages? Are the translations made into the mother tongue or into a foreign language, and by professionals or non-professionals? In the case of comparable corpora, an essential question is whether the subcorpora (translational and non-translational) are really comparable. Do they meet the requirements, for example, of a similar time span of publishing, similar text types and genres and length of texts? In the context of studies on wide-spread tendencies in translations, the question of representativeness is crucial. We must consider very carefully which genres and text types are included in the corpora and which source languages are included in the subcorpora of translations when the aim is to discuss so-called universal tendencies. If we attempt to find support for tendencies that exist widely in translations, should the text types or genres be similar in all comparable corpora that exist so that generalizations could be possible? A vast number of methodological questions still arise in corpus-based translation studies, and especially in studies that focus on universal tendencies.

Another limitation concerning the structure of comparable corpora is introduced by Kenny (1998: 53), who suggests that in the literature of the target culture, a genre that would be comparable to a new genre introduced through translation might not exist. Then, non-translations are totally lacking and compilation of a corpus is impossible. This situation could appear when a literary system in a given culture is young and has not yet taken shape. There is also another reason, related to Kenny's notion, which hinders the building of a comparable corpus, namely the lack of translations belonging to a genre or text type which is part of the literary system of a culture. This type of situation could be present in cultures where texts are published and consumed both as non-translations (originals) and in foreign languages, but not as translations. An example could be special hobby magazines or lifestyle magazines for marginal groups of people that are read in a foreign language and also (usually less) in the mother tongue but that are not translated because of the limited consumption. The range of text types in a comparable corpus is thus necessarily restricted to those that exist in the literary system of the target language and to those that are translated.

Bowker (2000: 19) states that the value of monolingual comparable corpora is mainly theoretical and that corpora of this kind do not provide help for translators in their work. However, the knowledge of specific features in translations is crucial, especially in translation training; it is important to keep in mind that teaching is always based on a theoretical background. Furthermore, if a translator wishes to avoid certain translational features in her/his work, for one reason or another, s/he must first be aware of these

features. How, then, can s/he find them? The answer is: by consulting comparable corpora or studies based on such corpora. (However, it is important to keep in mind that the aim of descriptive translation studies is not to say what kind of solutions a translator should use.) Finally, comparable corpora can be used along with grammars and dictionaries in the translation process. It is a well-known fact that dictionaries and grammars tend to offer too few examples of the use of lexemes and grammatical constructions. Unlike grammars and dictionaries, corpora give examples of language use in real contexts. Thus, a translator can check the use of words, collocational and colligational limitations, for instance, and test the existence of grammatical patterns by consulting the non-translational part of a comparable corpus, which consists of the "natural" language data. This can be essential especially when one is translating into a foreign language. (For an integrated model of the use of dictionary, grammar and corpus, see Johansson 1998.) Comparable corpora are thus useful directly in translator training and, at least indirectly, in actual translating when the translator has knowledge of the specific features of translated language.

Finally, let us consider the relation between translation studies and linguistics. It seems that the interaction between research in linguistics and research in translation studies has been profitable for the latter. As Fawcett (1998: 120) puts it, "it is clearly fair to say that linguistics does have something to offer translation studies". This is a widely accepted point of view and also true, of course. (For linguistic approaches in translation studies, see Baker 2000; Fawcett 1998.) However, Malmkjær (1997) claims that translation studies also has something to offer linguistics and to prove this she introduces studies in which linguistic categories (such as particles) have been investigated by using translational data. In very recent studies, comparable corpora have been used to illuminate the metatext in Finnish academic texts (Mauranen 2000) as well as the system and range of Finnish adverbs of degree and non-finite syntactic constructions (Jantunen 2001, Jantunen and Eskola 2002). The findings of these studies are useful both in linguistics and in translation studies. In the context of corpus-based translation studies, Baker (1993) suggests that translation should be taken seriously by linguistics, and by other related disciplines, because linguistics can benefit from the results of investigations done in translation studies. This is obvious especially in studies related to comparable corpora; studies that focus on both translated and non-translated language inevitably produce information about both language variants. If the results extracted from the translational subcorpus do not interest linguists because of the "poor research material", at least the results

based on non-translated material should be valid. The studies of Laviosa (e.g. Laviosa-Braithwaite 1996, 1997; Laviosa 1998), for example, have produced information about lexical density, sentence length and the use of vocabulary across different genres in English literature. A similar tendency can also be seen in Finland where the ongoing studies using CTF produce not only information concerning differences between original and translated language but also interesting information about the Finnish language itself.

3.2. TEC and CTF in particular

In this section, I will concentrate on TEC and CTF; but it is worth mentioning that my aim is not to rank the corpora. Rather, I will focus on the questions discussed earlier in more detail. To begin with, I will return to the issue of representativeness. In his article on this subject, Biber (1993: 243) states that “a thorough definition of the target population and decisions concerning the method of sampling are prior considerations.” It is also clear that before the process of defining the target population, the methods and the aim of a study, which determine the whole research process, should be considered carefully. Biber (*ibid.*) suggests that the definition of the target population has at least two aspects, firstly, what texts are included and excluded from the target population and secondly, what text categories are included in – or excluded from – that population. Although both criteria are equally important, in this paper I will discuss only the latter aspect.

In translation studies, the range of genres in corpora is limited to the text categories that include translations in that culture. The range of categories might, however, be wide, especially in “minor” language cultures where translation plays an important role. Which categories should then be represented in a corpus? Should translations be chosen based on criteria such as text status (see Mauranen 2000), or should the number of translations carried out in a specific genre be the major criterion? The CTF contains four main text categories as described above. The text categories are selected mainly according to the text status: fiction and academic prose, in particular, have a high prestige value in the Finnish literary system. Nonetheless, if the criterion were the number of translations produced or consumed, the range of categories in the corpus might be different. Obviously, in addition to the texts included in CTF, many other types of translations are produced and widely consumed in Finland. Such categories are, for example, help menus for software programs, user manuals for different kinds of apparatus, subtitles of television programs and films, and everyday texts such as lists of foodstuff

ingredients and other consumer goods. Subtitling, in particular, is a field which produces many translations every day, but which is not highly regarded in (corpus) studies. Should these categories be included in corpora, if the aim is to gain information about translations and language in translations in general? If not, why not? In many cases, the answer is the difficult access to the texts or, for lists of ingredients, the huge amount of work that would be needed to compile a reasonably large corpus. Since the corpora cannot include all existing, interesting or even widely consumed translations, the researcher must be aware of what kind of generalizations s/he makes about translations and translated language; the relationship between the sample and the target population is crucial and must be taken into account very carefully.

Another topic worth discussing here is the range of source languages of the texts in translational subcorpora. Both of the corpora discussed here contain texts from two language groups: TEC's languages belong mostly to the Indo-European or Semitic language group, CTF's languages belong to the Indo-European and Finno-Ugric language groups. If the aim is to study potentially universal features of translations, the range of source languages should be as wide as possible. Consequently, these corpora should also contain languages that belong to other language groups, for example, to African, Asian, Indian and North American Native language groups. This is important for ascertaining the impact of source languages that represent different cultural, political and historical backgrounds, to name but a few aspects. Still, we have at least two problems here. One, and the most obvious, problem is that the number of texts translated from these languages may be very small. That is the case at least in Finland. Another reason, which again concerns the "minor" languages, is that translations from "exotic" languages have often been translated via a "major" language, in Finland mainly via English. This makes analysis of the effect of the original source language, if not impossible, at least very unreliable. Obviously, the range of possible source languages and genres in corpora varies across cultures. However, it is possible and even desirable to combine and compare the results of studies based on TEC and CTF (and other existing or forthcoming comparable corpora) and in that way obtain a broader picture of translated language.

Although the information gathered from different studies based on different comparable corpora is essential in order to understand the very nature of translated language, there are, however, difficulties related to the comparability of different comparable corpora. To illustrate this, the structure of TEC and CTF are compared. First of all, the compositions of TEC and CTF are decisively different from each other. While CTF consists of two main

subcorpora (corpora of translated and non-translated language), TEC contains only translated texts. In the project of compiling CTF, one important criterion has been the comparability of the two main subcorpora. In practice this means that texts should represent similar properties with respect to time span, domain and sample size. Contrary to CTF, TEC is a source of translated texts only. The researcher whose aim is to compare translated and non-translated texts needs to compile a comparable corpus of her/his own, in which the one part is from TEC and the other part is from another corpus, such as the British National Corpus (BNC) (see Laviosa-Braithwaite 1996). A situation like this might cause problems for the researcher: the time span of the texts in the corpora may be different, the corpora may contain different domains and the sample sizes may also differ. For example, most of the texts in TEC were published in the 1990's (43 % after 1994), while in BNC most of the texts are from 1975–1993 (see BNC web page). In addition, TEC is a full-text corpus while BNC consists of text samples up to 45 000 words. Whether this is a problem depends on the aim of the study. In studies focusing on syntax, the difference between the length of samples may not be crucial; but in analyses of lexicon and cohesion, the length of the samples may play an important role. The problem of the comparability of TEC and BNC (NON-TEC) is discussed in more detail in Laviosa-Braithwaite (1996: 75-81).

The final point I shall take up is the lack of interpreting corpora in the field of translation studies. Shlesinger (1998: 486) argues that "corpus-based interpreting studies offer a tool which is both viable and revelatory not only for the study of interpreting, *per se*, but for translation studies as a whole." A corpus-based methodology can provide material and tools for analysing not only the specific nature of interpreting compared to non-interpreted spoken language but also the characteristics of interpreting compared to (written) translation. To study these different discourses, Shlesinger (*ibid.*) suggests that a three-part corpus (interpreted texts, original oral discourses and written translations) would be ideal for corpus-based interpreting studies. The compilation of an oral corpus has problems of its own; one of the major problems is transcription, which is very time-consuming. However, attempts to fulfil the demands have been made: at Savonlinna School of Translation Studies in the Training Project for Interpreters in Russian, an interpreting corpus was compiled, which consists of interpreting from Finnish into Russian and vice versa (see Jänis, forthcoming). The subcorpus of Russian-Finnish interpreting can then provide material suitable for studies that compare either interpreting and translation (with CTF as a control corpus) or interpreting and other oral discourse (Finnish corpora of oral discourse).

4. Conclusion

This paper has discussed some of the advantages and limitations of comparable corpora in translation studies. Like every other corpus, comparable corpora have limitations of their own. They have been judged to be suitable only for theoretical studies, but it is easy to find applications where comparable corpora can be used in practice. The limitations of comparable corpora are often external, related, for instance, to text types that are translated or that exist in the original language in the target culture. The paper also discussed two existing translational corpora, namely TEC and CTF, and it appeared that the structure of these corpora is clearly different. This suggests, naturally, that the two corpora also demand a different approach from the researcher, with respect to the whole research project and especially to the interpretation of the results.

References

- Baker, Mona (1993) Corpus Linguistics and Translation Studies: Implications and Applications. In Mona Baker, Gill Francis and Elena Tognini-Bonelli (eds.), *Text and Technology: in Honour of John Sinclair*, pp. 233-250. Amsterdam and Philadelphia: John Benjamins.
- Baker, Mona (1995) Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target* 7 (2): 223-243.
- Baker, Mona (1996) Corpus-based translation studies: The challenges that lie ahead. In Harold Somers (ed.), *Terminology, LSP and Translation. Studies in Language Engineering in Honour of Juan C. Sager*, pp. 175-186. Amsterdam and Philadelphia: John Benjamins.
- Baker, Mona (1999) The Role of Corpora in Investigating the Linguistic Behaviour of Professional Translators. *International Journal of Corpus Linguistics* 4 (2): 281-298.
- Baker, Mona (2000) Linguistic Perspectives on Translation. In Peter France (ed.), *The Oxford Guide to Literature in English Translation*, pp. 20-26. Oxford: Oxford University Press.
- Biber, Douglas (1993) Representativeness in Corpus Design. *Literary and Linguistic Computing* 8 (4): 243-257.
- Bowker, Lynne (2000) Towards a methodology for Exploiting Specialized Target Language Corpora as Translation Resources. *International Journal of Corpus Linguistics* 5 (1): 17-52.
- BNC Web page, <http://info.ox.ac.uk/bnc/what/basic.html>, last updated 29.2.2000, quoted 10.10.2000.

- Chesterman, Andrew 2001: Beyond the Particular. Paper presented at the Translation Universals - Do They Exist? Conference, Savonlinna, Finland, 19-20 October 2001.
- Fawcett, Peter (1998) Linguistic approaches. In Mona Baker (ed.), *Routledge Encyclopedia of Translation Studies*, pp. 120-125. London and New York: Routledge.
- Hermans, Theo (1999) *Translation in Systems. Descriptive and System-oriented Approaches Explained*. Manchester and Northampton: St. Jerome Publishing.
- Jantunen, Jarmo H. (2001) Synonymity and Lexical Simplification in Translations: A Corpus-based Approach. *Across Languages and Cultures* 2 (1): 97-112.
- Jantunen, Jarmo H. and Sari Eskola (2002) Käännös suomi kielivarianttina: syntaktisia ja leksikaalisia erityispiirteitä. *Virittäjä* 2/2002: 184-207.
- Johansson, Stig (1998) On the Role of Corpora in Cross-linguistic Research. In Stig Johansson and Signe Oksefjell (eds.), *Corpora and Cross-linguistic Research. Theory, Method, and Case Studies*, pp. 3-24. Amsterdam and Atlanta: Rodopi.
- Jänis, Marja (forthcoming) From the A language to the B language - what is the difference?. In *Interpreting in the 21st century. Challenges and opportunities*.
- Kennedy, Graeme (1998) *An Introduction to Corpus Linguistics*. London and New York: Longman.
- Kenny, Dorothy (1998) Corpora in translation studies. In Mona Baker (ed.), *Routledge Encyclopedia of Translation Studies*, pp. 50-53 London and New York: Routledge.
- Kenny, Dorothy (2001) *Lexis and Creativity in Translation. A Corpus-based Study*. Manchester and Northampton: St. Jerome Publishing.
- Laviosa, Sara (1997) How Comparable Can 'Comparable Corpora' Be?. *Target* 9 (2): 289-319.
- Laviosa, Sara (1998) Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose. *Meta* 43 (4): 557-570.
- Laviosa-Braithwaite, Sara (1995) Comparable Corpora: Towards a Corpus Linguistic Methodology for the Empirical Study of Translation. In M. Thelen and B. Lewandowska-Tomaszczyk (eds.) *Translation and Meaning Part 3*, Proceedings of the Maastricht Session of the 2nd International Maastricht-Lodz Duo Colloquium on "Translation and Meaning", Maastricht, The Netherlands, 19-22 April, 1995, pp. 153-163. Maastricht: Hogeschool Maastricht.
- Laviosa-Braithwaite, Sara (1996) *The English Comparable Corpus (ECC): A Resource and a Methodology for the Empirical Study of Translation*, PhD Thesis, UMIST, Manchester, UK.
- Laviosa-Braithwaite, Sara (1997) Investigating Simplification in an English Comparable Corpus of Newspaper Articles. In Kinga Klaudy and János Kohn (eds.), *Transfere Necesse Est*, Proceedings of the Second International Conference on Current Trends in Studies of Translation and Interpreting, Budapest, Hungary, 5-7 September, 1996, pp. 531-540. Budapest: Scholastica.
- Malmkjær, Kirsten (1997) Interaction Between Research in Linguistics and Research in Translation. In Anna Mauranen and Tiina Puurtinen (eds.), *Translation - Acquisition - Use*, AFinLA Yearbook 1997, pp. 9-22. Jyväskylä.
- Mauranen, Anna (2000) Strange Strings in Translated Language: A Study on Corpora. In Maeve Olohan (ed.), *Intercultural Faultlines. Research Models in Translation*

- Studies 1: Textual and Cognitive Aspects*, pp. 119-141. Manchester and Northampton: St. Jerome Publishing
- Olohan, Maeve and Mona Baker (2000) Reporting *that* in Translated English. Evidence for Subconscious Processes of Explicitation?. *Across Languages and Cultures* 1 (2): 141-158.
- Shlesinger, Miriam (1998) Corpus-based Interpreting Studies as an Offshoot of Corpus-based Translation Studies. *Meta* 43 (4): 481-493.
- Tymoczko, Maria (1998) Computerized Corpora and the Future of Translation Studies. *Meta* 43 (4): 652-659.

Contact address:

Jarmo Harri Jantunen
University of Joensuu
Savonlinna School of Translation Studies
P.O. Box 48
FIN-57101 SAVONLINNA
FINLAND
jarmo.jantunen@joensuu.fi