ENGLISH PARALLEL CORPORA
AND APPLICATIONS

Jos Hallebeek
Vakgroep Spaans
Universiteit van Nijmegen

ABSTRACT

In recent years there has been u revival of contrastive linguistics. This development follows the widespread interest in corpus-based approaches to descriptive linguistics. A special kind of corporu. the so-colled 'parallel' corporu, is used especially as u tool for contrastive investigation. These corpora ure now becoming readily available for a variety of languages. In this contribution we will first define the character of a parallel corpus and the purposes for which such o corpus can he used. Next we will present a series of parallel corpora with texts in English and some other language(s). After that we will discuss the methodology used in the compilation and exploitation of parallel corpora. This will he followed by a survey of unumber of linguistic studies realised using parallel corpora. The article ends with u comprehensive bibliography concerning studies about and with parallel corpora.

KEY WORDS: *corpus linguistics, parallel corpus and English language*

RESUMEN

En los últimos años asistimos u un resurgimiento de la lingüística contrastiva. Este desarrollo sigue el interés generalizado por el enfoque basado en análisis de corpus de parte de la lingüística descriptiva. Una clase especial de corpora, los llamados 'paralelos', se utiliza preferentemente como recurso en la investigación contrastiva. Hoy se va disponiendo poco a poco de estos corpora para una gran variedad de lenguas. En esta contribución, primero, definimos el carácter especial que tiene un corpus paralelo, así como los objetivos paru los que re puede usar. Luego, presentamos una serie de corpora paralelos que contienen textos en inglés y otra(s) lengua(s). Después, hablaremos sobre la metodología utilizada en la compilación y la explotación de corpora paralelos. Sigue una relación de estudios lingüísticos realizados con la ayuda de corpora paralelos. El artículo concluye con una extensa bibliografía de estudios sobre y con corpora paralelos.

PALABRAS CLAVE: *lingüística del corpus, corpus paralelo y lengua inglesa*

1. WHAT IS A PARALLEL CORPUS'?

Adhering to the terminology used by Baker (1995: 230) and McEnery (1996b: 58) a parallel corpus is a collection of texts in one or more languages with their translation into another language or languages that have been stored in a machine-readable format. It concems original texts and their translation. i.e. the same texts in different languages. Because the corpus consists of original texts and their translations it has also been referred to as a translation corpus by Johansson & Hofland (1994: 25) and Schniied & Schäffler (1996: 41). among others.

A parallel corpus should not be confused with a comparable corpus. The latter usually refers to a corpus consisting of original texts and translation in one and the same language (Baker 1995: 234). It does not concem the same texts. Such a corpus is used to ascertain differences in language use in original and translated texts. In Aijmer (1996) the term parallel corpus is used to refer to a comparable corpus and the term translation corpus is used for a parallel corpus. A comparable corpus can also contain original texts in different languages concerned with the same topics in both languages. Those texts are not translations of each other (Leech 1997: 21).

When a corpus contains original texts in two or more languages and the translations of those texts into different languages it is simultaneously a parallel corpus and a comparable corpus. because the corpus comprises original texts and translations thereof into another language as well as original and translated texts in the same language.

11. USE OF PARALLEL CORPORA

A parallel corpus is mostly used for the automatic construction of lexicons and for research into translation problems in two or more languages contrastively. Furthermore a systematic study of the differences between languages through a comparison of original and translation has proven its value in ascertaining the idiosyncratic aspects of those separate languages. A further research theme is the phenomenon known as translationese (Baker 1993: 243). It concerns deviations in the normal use of a language in translated texts as a consequence of the influence of the language of the source text. Lauridsen (1996) questions the use of translated texts from parallel corpora for the analysis of the structure of a language or a text. Do we actually test the language. or is it more the translator's achievement being tested? Gellerstam (1996) concurs that translations are not uselul for studies about language typology or linguistic universals. because of the fact that the translation is influenced by the source language. He thinks that a parallel corpus can be a useful tool for language learning. Parallel corpora. for example. supply material for training courses and a translator's education.

Automatic translation systems have started using parallel corpora increasingly to supply desired translations. Systems based on the syntactic and/or semantic analysis of complete sentences are obscured ever more. Using an aligned bilingual corpus statistical methods are now deployed to find a translation for new texts that have not been translated. This works as follows. When a sentence is offered up for translation. the programme first checks if the corpus already contains it in a similar form. If so. the translation can be copied instantly. If not, the sentence is split into segments using syntactical grammars that need not be complete and a translation for those segments is found. The translation of groups of words is sooner found in the corpus than that of complete sentences. (Brown et al: 1990. 1993: Tsujii et al: 1991). Such systems are not entirely operational yet.

Furthermore. corpora comprising only translated texts are also used for research into the properties of translations. Some linguists (Baker 1993: 242-246) are of the opinion that pattems can be found in translated texts. independent of the language. that cannot be explained as being caused by interference by the source or target text. They call these universal features of translation: i.e. features that are typical of translated texts in general. They are. for example. the following occurrences. A simpliiication of word usage. a greater complexity than is usual in original texts. a tendency towards solving and avoiding ambiguity. the prevention of ungrammatical sentences and expressions even if they occur in spoken language. the completion of elliptic sentences.

III. EXAMPLES OF PARALLEL CORPORA

Because technological advancement has increased the possibilities for creating corpora and the consultation thereof an increasing number of parallel corpora are being created nowadays. This does not mean they will be available for large numbers of researchers. Especially problems concerning copyright causes these corpora to only be available to a small number of people. A number of corpora that have been developed are mentioned below in which one of the languages. either the source or target language of the translation is English.

The English-German Translation Corpus of the University of Chemnitz (Schmied & Schaffler 1996). The corpus is being compiled in co-operation with the Research Centre for English and Applied Linguistics from the University of Cambridge. It is a parallel project with Lund and Oslo/Bergen when it comes to the methodology (see below). As text types the corpus contains speeches of ministers. tourist brochures. documents from the EU. scientific texts and literature: English and German original texts with the translations. The corpus is not yet completed and will have a size of about one million words.

The English-Norwegian Parallel Corpus (ENPC) of the Universities of Oslo/Bergen (Johansson 1997). In fragments of 10.000 to 15.000 words. 100 Original texts and

translations thereof. about 2.5 million words in total. The proportion fictional : non-tictional texts = 3:2. In 1996 the corpus was expanded with other languages with the original English texts as a starting point. The other languages are Dutch. German. Portuguese. and possibly Swedish and Finnish.

The English-Swedish Parallel Corpus of the University of Lund (Aijmer et al. 1996: 73-85). Originally an English-Swedish corpus with forty original English and forty original Swedish texts. Later Finnish and Norwegian were added as target languages: all source texts in English must be available in translation in Finnish. Norwegian and Swedish. Fragments of 10.000 to 15.000 words. always taken from the start of a book. The corpus has a wide range of text types. authors and translators. although only written texts. The texts are in conteniporary language: since 1980. The proportion fiction : non-fiction is 1:1.

The European Corpus Initiative Multilingual Corpus I (ECI/MCI). A multilingual corpus on CD-ROM. Besides texts from German. French and Dutch newspapers the CD-ROM also contains a parallel corpus of about five niillion words in English. French and Spanish. The texts thereofwere taken from the Official Bulletin. B Series 1984-1989 of the International Labour Organisation (ILO).

The Hansard Corpus. It contains a collection of proceedings of the Canadian Parliament in English and Canadian French. Includes spontaneous discussion and written correspondence. besides legislative propositions and prepared speeches. It covers a time span from the niid-1970's through 1988. The material is available on CD-ROM from the IBM T.J. Watson Research Center and from Bell Communications Research Inc. and was released in January 1999 from the LDC (Linguistic Data Consortium) (Hansard 1999).

The International Telecommunications Union Corpus. Used in the European CRATER-project (Corpus Resources and Terminology Extraction). The project was a co-operation between Lancaster University (England) and Universidad Autónoma de Madrid (Spain). The corpus is in three languages: English. French and Spanish. It consists entirely of technical texts from the Intemational Telecommunications Union. The corpus is completed and consists of 5.5 million words (McEnery & Oakes 1996a).

The Intersect Parallel Corpus. University of Brighton. The project started in 1994. The aim is to construct and analyse a parallel bilingual corpus of French and English written texts. There is no information about the current state of affairs and obtained results except for Salkie's study (1997) about 'but' and 'mais'.

MULTEXT (Multilingual Text Tools and Corpora) Project. supported among others by the European Commission. Strictly speaking it is a series of projects whose goals are to develop standards and specifications for the encoding and processing of linguistic corpora. and to develop corpora for a great variety of languages. including Bulgarian. Catalan. Dutch. English. Estonian. French. German. Hungarian. among many others. The project co-ordinator is Jean Véronis from the University of Aix en Provence. The work on the encoding and processing has been done in close collaboration with the EAGLES and TEI

projects. There is a corpus available that consists of a set of segrnents from the Official Journal of European Cornrnunity in five languages: English. Gerrnan. Italian. Spanish and French.

   The Multilingual Parallel Corpus. Started in 1993 as a European Union funded project. under the leadership of Mrne F. Roussel. University of Nancy. The aim of the project was the developrnent of a parallel concordancer. named Multiconcord. for direct autornatic coniparison of an indexed corpus of texts and their translations. The programme has to achieve paragraph-based alignrnent with any pair of bilingual texts. To which degree this prograrnrne will be ready to use is unclear. Originally six languages were involved: Danish. English. French. German. Greek and Italian. In 1997 Finnish. Portuguese. Spanish and Swedish were added. The corpus contains a nurnber of texts in at least three languages. All languages have been represented by at least one source text. All texts were taken frorn Stephen Hawking's. A *Brief History of Time* (a selection). the *Paris Eurodisney Brochure.* the *Opel Car manual* and sorne fiction texts. (King 1997)

   The Parallel Corpus of German and English Fiction. It's being compiled at Dublin City University and UMIST. The corpus will contain about one million running words in German. complete novels. with their translations in English. (Kenny 1997)

   More extensive data concerning the corpora that do not include a bibliographical reference can be found on the Internet by searching for the title of the corpus or the term parallel corpora.

IV. METHODOLOGY IN COMPILATION AND EXPLOITATION

In order to use a parallel corpus properly the original texts and their translations need to be aligned. which rneans that the sentences or parts thereof. with their translation are placed next to one another. It would be ideal if this alignrnent were to take place to the level of the word. but this is currently still impossible. Several automated rnethods have been developed. Firstly there is the Gale & Church-rnethod (Gale & Church 1993). This rnethod has been applied to the *Canadian Hansard Corpus.* It is purely statistical based on two starting points. The Iirst one is that longer sentences in the one language are translated into longer sentences in the other. The second one is that it is more obvious for one sentence in the one language to be translated into one sentence in the other. rather than two or three sentences. However. translators know all too well that the latter is not always the case.

   In the *CRATER project* the Gale & Church rnethod has been cornbined with a probabilistic approach. The latter rnakes use of penalties to come to a correct alignrnent of translation and original. In short this means that a coniparison is rnade between the number of lener syrnbols of the sentences and the nurnber of sentences in the original and the translation. When a cornparison of the nurnber of letter symbols shows that one sentence

was translated into more than one sentence. then this latter combination is awarded a number of penalties. The more penalties the smaller the probability that this is indeed a translation (McEnery & Oakes 1996).

A further refinenient of the method was established by the recognitioii of so-called cognates in the original and the translation. They are words and especially symbols that the two languages have in common. They are. for example. proper names. numbers. question marks. exclamation marks. quotation marks. colons, etcetera. They are also known as anchor points; the name is self-explanatory. In the *English-Norwegian Parallel Corpus* (Johansson & Hofland 1994: 29-32) they even use a bilingual lexicon of 850 words. which contains mainly function words. but also content words. such as the days of the week. the months and the most common adjectives and nouns. But also word stems. such as 'open' in English. which also stands for 'opens'. 'opened'. 'openly'. 'openness'. Based on the presence of anchor words it is ascertained whether or not a combination of sentences is an original and its translation. A comparison of the number of'words and letter symbols is also taken into account. A sentence in a language with few compounds contains more words. but probably about the same amount of letter symbols as the same sentence in a language that knows many compounds.   The way in which the *English-Norwegian Parallel Corpus* of the Universities of Oslo/Bergen is structured is a clear example of the application possibilities of a parallel corpus. It contains original texts in English and Norwegian. The translations of those texts into another language are also included. This corpus is therelore useful for both contrastive analyses and the study of translations; a contrastive study of the original texts in both languages and a contrastive study of the original texts in the one language and the translation thereof into another. Translation studies can be conducted on the concordance between English and Norwegian and vice versa. or of original and translated texts in the same language. or of translated texts in each of the languages. (S. Johansson et al. 1996. 1997)

V. STUDIES PERFORMED USING PARALLEL CORPORA

In recent years an increasing nuniber of studies into linguistic phenomena are heing conducted using parallel corpora. Some of these are reasonably complete. well worked out and justified. others seem to be only tentative and meant to test the use of a corpus. Because it is interesting to check into the possibilities for research themes a number of published studies are briefly discussed below by name of the author and a reference to the publication in question.

V.1. Studies on the basis of the *Chemnitz English-German Translation Corpus*

The *English-German Translation Corpus* of the University of Chemnitz has provided all the material needed for the study of the translationese phenomenon (Schmied & Schäffler 1996). Original texts were used in English and German. as well as translations of English texts into German. The concrete study topic was the question whether or not English is less direct than German when it comes to language use and if this leaves traces in translations of English into German. Both authors tried to discover quantitative and qualitative differences within the elements and structures that signify tentative and indirect usage. This concems mostly modal adverbials and verbs. impersonal constructions. commentatory interjections. etcetera. It is a very complex analysis that bears the risk of formulating interpretations that are too subjective.

The same English-German corpus provided the material to verify the following question for the same authors. Is a translated text more explicit than the corresponding source text? There is structural explicitness. i.e. depending on the differences between two languages. and on the non-structural explicitness. i.e. language independent. related to the translation process in general. Both kinds of explicitness are expressed on a lexical as well as a grammatical level. The reverse process is condensation. also structural and non-structural and on a lexical and grammatical level. Condensation occurs when elements that are present in the source text disappear in the target language (the translation). The authors of the study are not explicit in the explanation of the quantity of material they have analysed. They only give a few examples of the phenomena discussed.

V.2. Studies on the basis of the *English-Norwegian Parallel Corpus*

Using the *English-Norwegian Parallel Corpus*. expanded with German texts Johansson (1997) draws a comparison in the use of the generic person in three languages: English. Norwegian and German. The varieties of English include: 'one'. 'you'. 'we'. 'people'. 'they' and other constructions such as the passive voice. English 'one' can be taken as a starting point as well as German and Norwegian 'man' and they can then be compared to the English original. The investigation of the relation between the three languages in this grammatical aspect is very interesting. The contrastive approach brings out features of a language that would not have come to light in a monolingual analysis.

Hasselgård (1997) compared the sentence openings in the original and the translation in eight texts; four of them English and four of them Norwegian with their translation. 150 Sentences were randomly selected from each text. so in total it concemed 1200 sentences. As a working hypothesis she assumed that Englisli and Norwegian are closely related languages and that therefore the order of the elements in both languages will be similar.

Furthermore. the verb in Norwegian is always in second position and fronting of objects and complements occurs more often in Norwegian than in English. A difference in the placement of adverbials is the second most frequent change in word order between both languages. Finally there must be an influence of the source language on the target language in translations. As the conclusion of the investigation she formulates that the translations in both texts approach the source language. especially in word order. and that Norwegian is more flexible than English when it comes to word order. The conclusions are all but sensational as could be expected. As far as the structure and elaboration of the analysis are concerned the article contains a number of interesting suggestions.

V.3. Studies on the basis of the *English-Swedish Parallel Corpus*

Wikberg's study (1996) contains a comparison between interrogative sentences in English and Swedish. To this effect he used original and translated segments of novels in both languages. Interrogative sentences can express multiple linguistic actions and are not restricted to requests for information. They can also be other kinds of requests. suggestions. etcetera. It is interesting to see how mutual differences are manifested in both languages.

The topic of the study of M. Johansson (1996) is the occurrence of fronting in Swedish and English in three original English and three original Swedish texts with their translations. Johansson considers the placement of a random element in front of the subject in an English declarative sentence fronting. This does not include conjuncts. relative pronouns or interrogatives. and interjections like well. In Swedish the standard order is that an element is placed in front of the verb and subject. It uould appear that fronting is less common in English than in Swedish. which results in the direct object or adverbial being placed in front of the verb and subject.

V.4. A study on the basis of the *Intersect Parallel Corpus*

In his article Salkie (1997) reports on an investigation of English 'but' and French 'mais'. These words do not always correspond in translations. He talks extensively about translation equivalence. natural language and the kind of theoretical framework that might be helpful to structure the work in the field of contrastive linguistics.

V.5. Studies on the basis of the *Multilingual Parallel Corpus*

King (1997) used the *Multilingual Parallel Corpus* for research into translator behaviour. to

check translator behaviour against a bilingual dictionary and to verify claims made in translation theory. In the article only a few examples of the analysis of these three aspects can be found.

Resides a good introduction on the state of the art in translation studies and the contribution of computerised corpora. Ulrich (1997) gives a fine example of an Italian to English translation of the Italian 'ossia'. from texts of Italo Calvino. These texts are part of the Multilingual Parallel Corpus.

V.6. A study on the basis of the *Parallel Corpus of German and English Fiction*

Kenny (1997) intends to use the *Parallel Corpus of German and English Fiction* to investigate the following question: "Are translated texts lexically more conventional than their source texts or original texts in the target language?" The idea is to identify unusual collocations in original texts in English and German. and then to see if the number of this type ofcollocations is more limited in translation. There is no information about the results of the investigation yet.

V.7. A study on the basis of the Bible

Vraukó (1997) works out an original idea: use the Bible in various language editions for contrastive studies. The translator is usually not a single person; the influence of'the source text is minimal because of the history of the different national editions. The Bible contains many different kinds of text sorts: dialogue. spoken and written language. informative. instructional. persuasive and imaginative language. He has used a bilingual American-Hungarian edition of the New Testament to make a comparison of the use of the passive voice in (American) English and in Hungarian. It's a pity that Vraukó gives only a feu examples of the construction in both languages and that the analysis is not worked out. The article really does not contain more than the suggestion that the Bible is suitable for such research. The few examples mean nothing uithout further analysis. Furthermore the text shows that both versions of the Bible are not available in an electronic format.

VI. CONCLUSION

It is clear there is an ever-growing interest in the composition and use of parallel corpora. Especially in the northern countries a lot of time and energy goes into this. The EU also appears to be willing to financially support such activities. The applications of these kinds of corpora, however. are still developing. Without exaggeration it may be said that this new shoot on the corpora tree has a bright future to look forward to. The applications are many. after all: translation programmes. translation education. second language education. grarnmatical studies. the composition of bilingual lexicons.

Besides the necessary development of the required software for the alignment of original texts and their translation. the availability of texts is a major problem. A general corpus. which is useful for many purposes. requires the presence of a great variety in kinds of text. The technical manuals. government documents. parlianiental minutes or tourist leaflets in original and translation found in many corpora represent but limited aspect of the possibilities of language use. Novels. plays. travel descriptions. diaries. popular science manuals. and. not to be forgotten. spoken language are textual varieties that most certainly deserve a place in a general corpus.

Another problem is that translations are more common from one language than from the other. This goes especially for English. which is a source for innumerable translations. but is used less as a target language. The quality of'the translations should also be taken into account. A bad translation can never be useful for contrastive studies or as an aid for translation training. Translations of literary works are relatively frequent and on top of that of an ever better quality. A disadvantage is that they only contain one type of'text.

From the brief descriptions of the studies carried out concerning parallel corpora it is obvious that there are many interesting angles for conducting research in the Iield of grammar. contrastive linguistics. translation and lexicology. And finally. a parallel corpus with appealing texts. such as the ones found in periodicals can be very useful in language education.

REFERENCES

Aijmer. Karin. Bengt Altenberp & Mats Johansson (eds.)(1996): *Languages in Contrast. Papers from o Symposium on Text-based Cross-linguistic Studies. Lund 4-5 March 1994.* Lund University Press/Chartwell-Bratt Ltd. Bromley. Lund

Baker. Mona (1993): "Corpus linguistics and translation studies. Implications and applications". In Baker *et al.* (1993: 233-250)

Baker. Mona (1995): "Corpora in translation studies. An overview and some suggestions for future research". In *Target.* 7.2. 223-243

Baker. Mona. Gill Francis & Elena Tognini-Bonelli (eds.)(1993): *Text and technology. In Honour of John Sincltrir.* John Benjamins Publishing Company. Philadelphia/Amsterdam

Brown. P.. J. Cocke. S. della Pietra. V. della Pietra. F. Jelinek. J. Laffarty. R. Mercer & P. Roosin (1990): "A statistical approach to machine translation". In *Computational Linguistics.* 16(2). 79-85

Arown. P.. V. della Pietra. S. della Pietra & R. Mercer (1993): "The mathematics of statistical machine translation: parameter estimation". In *Computational Linguistics.* 19(2). 263-301

Fries. Udo. Gunnel Tottie & Peter Schneider (eds.)(1994): *Creating und using English Language Corpora.* Rodopi. Amsterdam - Atlanta. GA

Gale. W.A. & K.W. Church (1993): "A program for aligning sentences in bilingual corpora". In *Computational Linguistics.* 19(1). 75-102

Gellerstam. Martin (1996): "Translations as a source for cross-linguistic studies". In Aijmer *et al.* (1996: 53-62)

Hansard Corpus of Parallel English and French Handset TIMIT Speech Corpus Lincoln Laboratory... 1999

Hasselgård.H. (1997): "Sentence openings in English and Norwegian". In Lewandowska & Melia (1997: 3-20)

Johansson. Mats (1996): "Fronting in English and Swedish: A text-based contrastive analysis". In Percy *et al.* (1996: 29-39)

Johansson. Stig (1997): "Using the English-Norwegian parallel corpus - a corpus for contrastive analysis and translation studies". In Lewandowska & Melia (1997: 282-296)

Johansson. Stig & Jarle Ebeling (1996): "Exploring the English-Norwegian

Parallel Corpus". In Percy *et al.* (1996: 3-15)

Johansson. Stig. Jarle Ebeling & Knut Hofland (1996): "Coding and aligning the English-Norwegian Parallel Corpus". In Aijmer *et al.* (1996: 87-112)

Johansson. Stig & Knut Hofland (1994): "Towards an English-Norwegian parallel corpus". In Fries *et al.* (1994: 25-37)

Kenny. Dorothy (1997): "(Ab)normal translations: a German-English parallel corpus for investigating normalization in translation". In Lewandowska & Melia (1997: 389-392)

King. Philip (1997): "Parallel corpora for translator training". In Lewandowska & Melia (1997: 393-402)

Lauridsen. Karen M. (1996): "Text corpora and contrastive linguistics: which type of corpus for which type of analyses". In Aijmer *et al.* (1996: 63-71)

Leech. Geoffrey (1997): "Teaching an Language Corpora: a Convergence". In Wichmann *et al.* (1997: 1-23)

Lewandowska-Tamaszczyk & Patrick James Melia (eds.) (1997): *Palc'97. Practical Applications in Language Corpora.* Lódz University Press

Ljung. Magnus (ed.)(1997): *Corpus-based studies in English.* Rodopi. Amsterdam-Atlanta. GA

McEnery. Tony & Michael Oakes (1996a): "Sentence and word alignment in the CRATER Project". In Thomas *et al.* (1996: 211-231)

McEnery. Tony & Andrew Wilson (1996b): *Corpus Linguistics.* Edinburgh University Press

Percy. Carol E.. Charles F. Meyer & Ian Lancashire (eds.) (1996): *Synchronic corpus linguistics.* Rodopi. Amsterdam - Atlanta. GA

Sadler. V. (1989): *Working with Analogical Semantics.* Foris. Dordrecht

Salkie. Raphael (1997): "Naturalness and contrastive linguistics". In Lewandowska & Melia (1997: 296-312)

Schniied. Josef & Hildegard Schäffler (1996): "Approaching translationese through parallel and translation corpora". In Percy *et al* (1996: 41-56)

Schmied. Josef & Hildegard Schaffler (1997): "Explicitness as a universal feature of

translation". In Ljung (1997: 21-34)

Thomas. Jenny & Mick Short (eds.) (1996): *Using Corpora for Language Research.* Studies in Honour of Geoffrey Leech

Tsujii. J.. S. Ananiadou. J. Carroll & S. Sekine (1991): *Methodologies for the Development of Sublanguage MT System II.* CCL UMIST Report No. 91/11

Ulrych. Margherita (1997): "The impact of multilingual parallel concordancing on translation". In Lewandoska & Melia (1997: 421-435)

Vraukó. Tamas (1997): "The text of the Bible as a corpus for contrastive analyses". In Lewandowska & Melia (1997: 313-320)

Wichmann. Anne, Steven Fligelstone. Tony McEnery & Gerry Knowles (eds.)(1997): *Teaching and language corpora.* Longman. London and New York

Wikberg. Kay (1996): "Questions in English and Norwegian: Evidence from the English-Norwegian Parallel Corpus". In Percy *et al.* (1996: 17-38)

Wilson. A. & A. McEnery (eds.) (1994): *Corpora in Language Education and Research: A Selection of Papers from Talc94.* Unit for Computer Research on the English Language. Technical Papers 4 (special issue) Lancaster University

Zanettin. F. (1994): "Parallel words: designing a bilinpual database for translator activities". In Wilson *et al.* (1994: 99-111)