

Corpus Linguistics and Translation Studies: Interaction and Reaction

Maeve Olohan

Centre for Translation and Intercultural Studies - UMIST

This paper focuses on one particular parallel development in linguistics and translation studies, namely corpus-based analysis of language use. Recent years have seen the compilation of corpora of translations, designed specifically to investigate the language and features of translation, usually by comparing translations with non-translations. Some of the interactions between corpus linguistics and corpus-based translation studies are traced in terms of perceptions of translated texts and underlying assumptions of corpus-based studies. Corpus-based translation studies is placed in the context of current theoretical trends in translation studies and, through brief reference to research which has aimed to investigate potential features of translation, attention is drawn to the importance of contextualising translation by combining corpus-based investigations with other kinds of methodologies and analyses.

1. Translation in corpus linguistics

From the perspective of a translation scholar interested in corpus-based translation studies, it is immediately striking that the range of areas of language studies dealt with in general introductions to corpus linguistics (e.g. Biber, Conrad & Reppen 1998; McEnery & Wilson 2001; Kennedy 1998) does not include translation. Tony McEnery and Andrew Wilson (2001), for example, cover numerous topics within linguistics: lexical studies, grammar, semantics, pragmatics and discourse analysis, sociolinguistics, stylistics and text linguistics, historical linguistics, dialectology and variation studies and psycholinguistics. In addition, they deal with related fields: the teaching of languages and linguistics, cultural studies and social psychology. Teaching translation, but not translation studies, is covered in one paragraph in the language teaching section. The lack of attention to translation studies may be because the use of corpora in translation studies is relatively new, or perhaps because the exchange of knowledge between linguistics and translation studies has tended to be rather mono-directional. Moreover, the perception of translations has traditionally not been particularly favourable in linguistics; their exclusion from so-called language reference corpora (such as the British National Corpus) would indicate that they are not considered as representing language use, in English-speaking contexts at least. Often, the way in which they are used in parallel corpora indicates that translations are not seen as texts which exist and function in their own right in the target language system, nor as being subject to a range of constraints which differ from other text production situations. One conventional view taken of trans-

lation in corpus linguistics is revealed by the following definition: “a bilingual parallel corpus is a corpus that contains the same text samples in each of two languages, in the sense that the sample are translations of one another” (Oakes & McEnergy 2000:1). Thus, source texts and translations are the “same” text, that is, unless there are “discrepancies” between them: “In real life, discrepancies between a source text and its translation, such as differences in layout, omissions, inversions etc., are quite common” (Simard et al. 2000:42). Elsewhere “discrepancies” are also described as “idiosyncrasies”, resulting from various factors, not least the translator’s alcohol consumption:

Of course, any particular translation will contain a number of idiosyncrasies and the translator in trying to get the best overall translation may have to make compromises [...] in order to get the best overall result. The translator has to strive for an optimal solution for a translation in the face of competing pressures. The way in which a work is translated in a particular instance will depend on a number of factors, including the form of the previous discourse and other contextual influences, including perhaps how much wine the translator had at lunch time (Barlow 2000:110-111).

Few translators have the luxury of the leisurely lunch conjured up here, but many translation scholars will be familiar with views of translation, held within neighbouring disciplines, which do not necessarily take account of advances and current concerns in translation theory and translation research. However, with an increase in interaction between translation scholars and corpus linguists comes greater understanding of translation; Stig Johansson’s acknowledgement of the difficulties inherent in using a corpus of texts and their translations for cross-linguistic study, while still viewing translation very much in terms of its ‘equivalence’ to a source text, also reflects some awareness of the cotextual, contextual and extratextual influences on translators and translation, and an interest in studying features of translation:

it is well-known that linguistic choices often differ depending upon the individual translator, or there may be outright mistakes in translation. To what extent can we then make generalizations based on translated texts? And can we really be sure that the same meanings are expressed in the source and the target text? Or should we rather think in terms of degrees or types of equivalence? [...] Most seriously, to what extent can we take translated texts to be representative of ordinary language use? Translated texts may differ from original texts because of source language influence [...] Moreover, there may be general features which characterize translated texts (Johansson 1998:6).

2. Translation in translation studies

Theo Hermans (1999:7-16) outlines succinctly the development of the descriptive paradigm in translation studies from the 1960s to the present day as an approach which is interested in translation “as it actually occurs, now and in the past, as part of cultural history” (ibid.:7). However, the discussion

of how to identify or label something as a translation pervades large sections of translation studies literature. Many scholars have adopted Gideon Toury's (1995:31-35) suggestion that we focus our research on anything which is assumed to be a translation. Toury proposes the "assumed translation" as a way of accounting for the "variability" of the object of translation, which is characterised by "difference across cultures, variation within a culture and change over time" (Toury 1995:31). Hermans (1995) argues that it is not that simple. Reviewing *In Search of a Theory of Translation*, the precursor to Toury's (1995) *Descriptive Translation Studies and beyond*, Hermans sees a conflict between translation as perceived among a group of people, and translation as a label attached by a researcher to a certain behaviour, i.e. "a kind of universal human activity, the common denominator extrapolated from all occurrences of translational action through space and time" (Hermans 1995:220). While the former may be adequate for many research purposes, Hermans feels that the "notion of translation" cannot be universal since it has been extrapolated from only a limited set of occurrences of translational action. Researchers rely on their own cultural understanding of the concept of translation, represented by some or other label in whatever language, when studying concepts which may be more or less similar or different to their own understanding of the concept labelled 'translation'. Maria Tymoczko (2002:17fn) appears to agree with this perspective, arguing that Toury's definition operates on the level of theory and permits "any culture's definition of translation to be treated as equally valid". On the level of research hypotheses, according to Tymoczko, "any research may and usually even must limit the scope of inquiry for practical reasons" (ibid.) while nonetheless making this delimitation explicit.

Prototype theory provides a useful alternative viewpoint (see Halverson 1999 and Olohan forthcoming a) by removing the question of what constitutes a translation and the need for clear boundaries between categories. Emphasis is placed on 'best examples', so the issue becomes one of centrality or gradience of membership to the 'translation' category and other categories. Thus, prototype theory allows us to think about the relationship between those different objects considered to be translations in some context on the one hand and a prototypical translation on the other. This means looking at features which are likely to be shared by prototypical translations but which less prototypical translations may not exhibit; the latter are not invalidated as objects of research by virtue of not having all of these features. Norms tell us about the expectations on particular translations in particular contexts;¹ prototype effects are category judgements from subjects on a concept. Thus, in a sense, the extent to which a translation displays prototype effects is measured on the basis of normative expectations. The norms and the prototype effects are cognitively, socially and culturally determined and vary across time and space.

Thus, the commonalities between a norms-based perspective on translation and a prototype approach can be seen in Hermans (2000a) in which the author says of certain forms of translation (e.g. homophonic translation) that they:

bring the boundaries of translation into view. By challenging what is permissible, they probe the conceptual, socially recognized perimeter of what counts as ‘translation’ – not, that is, ‘translation’ in an absolute sense, but as it is understood in the world in which these texts are deployed (Hermans 2000a:263).

Hermans is not interested here in the notion of “translation” “in an absolute sense”, or “translation per se” or “some universal, de-historicized idea of translation” (ibid.:272). He stresses the concept of ‘translation’ “as it is understood in the world in which these texts are deployed”, and “translation as we have come to think of it” (ibid.:271). He talks in this particular paper about the self-reflexivity of translation, and discusses an example in which the translator has intervened explicitly in the text to highlight the difficulty of translating in a manner which conforms to a “prevailing concept of translation tied to particular sets of cognitive and normative expectations” (ibid.:272). It is this norms-based conceptualisation of translation which embodies the experiential, cognitive and socio-cultural nature of translation, and, in this respect, appears to be consonant with the prototype theory approach.

This discussion leads to the conclusion that we cannot talk about universals of translation or universal laws of translation because we cannot account for all translation, all variables etc. and the approach does not accommodate the existence of a decontextualised concept of translation. However, as with all other abstract and complex notions, we often use more concrete ones in a metaphorical way to help us to understand translation. These more concrete concepts are usually grounded in basic human experience and there may therefore be commonalities in how different cultures, societies and language communities conceive of translation over time and through space, although there will certainly also be differences.

3. Translation in corpus-based translation studies

Corpus-based methodology clearly has some applicability within the broad theoretical framework of descriptive translation studies, since it appears to provide a method for the description of language use in translation. Unlike much multilingual corpus linguistics research, corpus-based translation studies focuses on the translation, not in terms of its relationship to a source text but instead foregrounding it as an instance of text production and communication in its own right. The discussion of what we understand translation to be is therefore important for a number of reasons. Firstly, researchers’ viewpoints on the concept of translation form an important basis for the application of corpus-based methodology to the study of translation, since they will underpin the choice of object of study, i.e. what kind of translation, produced when, by whom, for what purpose. They thus form the basis of decisions in corpus design and issues of representativeness, i.e. decisions as to which particular texts might be included in a corpus to be used to study that particular kind of translation. They are crucial in the analysis and interpretation of data too, since this requires clarity on the issue of what concept of ‘translation’ is

being described by these data. And, since corpus analysis usually places emphasis not only on what is observable but also on what is regular, typical and frequent, it relates directly to norms as discussed by descriptive translation scholars.

Against this backdrop, Mona Baker's (1995:234) initial suggestions for research using a comparable corpus (i.e. a corpus of translations in a language and a comparable corpus of non-translations in that same language) were to capture "patterns which are either restricted to translated text or which occur with a significantly higher or lower frequency in translated text" (ibid.:235).² She points out that these may be related to a specific linguistic feature in a specific language, but that we may also find out about "the nature of translated text in general and the nature of the process of translation itself" (ibid.:236). From this comes Baker's focus on what she termed "universals of translation" at that time – in the light of the problematic nature of the notion 'universal', these are now more commonly referred to as features of translation. She posited a number of features of translation which could be investigated using comparable corpora (Baker 1996), for example, that translations may be more explicit on a number of levels than non-translated texts, and that they may simplify and normalise or standardise in certain ways.

Much of the comparable corpus research carried out to date in translation studies has focused on syntactic or lexical features of translated and non-translated texts which might provide evidence of such processes of explicitation, simplification or normalisation. It should be stressed that, while translators may at times consciously strive to produce translations which are more explicit or simplified or normalised in some way, the use of comparable corpora is also seen as a way of investigating aspects of translators' use of language which are not the result of deliberate, controlled processes. Translators may not be aware of these processes but the translation product may provide indirect evidence of cognitive processing inherent to translation. An example of one such aspect is the use of the optional *that* with reporting verbs SAY and TELL, studied in Olohan and Baker (2000); the use of the optional *that* was found to be considerably higher in the Translational English Corpus than in a comparable corpus comprising texts from the British National Corpus. This was posited as being a reflection of explicitation, based on the hypothesis that explicitation will usually involve the use of a longer surface form in preference to a shorter one, leaving less room for ambiguity. This study drew on Günter Rohdenburg's (1996) work on cognitive complexity and grammatical explicitness in English. He examines formal contrasts involving the deletion or addition and the substitution of grammatical or closed-class elements, providing evidence for the complexity principle: "in the case of more or less explicit grammatical options the more explicit one(s) will tend to be favored in cognitively more complex environments" (ibid.:151). Thus, the higher incidence of reporting *that* in translated English could be considered to be part of a more general pattern of grammatical explicitness, and explanation for this explicitation may be linked to the cognitive complexity of the translation task.

4. Contextualisation of language use in corpus linguistics

The Firthian and neo-Firthian approach to linguistics focuses on language in its social context and has often been used to provide a general framework for corpus linguistic investigations.³ Thus, according to corpus linguist Jan Aarts (1999:3), corpus linguistics is interested in describing ‘language use’ and he formulates a number of requirements for a descriptive model of language use (*ibid.*:6-7):

- (1) the model should allow the combination of a quantitative and a qualitative description of the data
- (2) the model must establish a relation between phenomena that are external to the language system and system-internal phenomena
- (3) the model should allow the description of the full range of varieties, from spontaneous, non-edited language use (usually spoken), to non-spontaneous edited language use (usually written or printed).
- (4) the model should allow an integrated description of syntactic, lexical and discourse features (Aarts 1999:6-7).

These requirements on the description of language use are reflected in other scholars’ description of what corpus linguistics aims to do. Graeme Kennedy (1998:1), for example, introduces corpus linguistics as “one source of evidence for improving descriptions of the structure and use of languages”. He stresses that advances in computer technology have made it easier to work with larger quantities of text but have not drastically changed the nature of text-based linguistic study: “corpus linguistics is not a mindless process of automatic language description” (*ibid.*:2). Rather, corpora are used by linguists who seek to answer questions, and “some of the most revealing insights on language and language use have come from a blend of manual and computer analysis” (*ibid.*:2-3). Thus, quantitative and qualitative analyses are combined for the description of language as it is actually used, setting this in opposition to the theoretical possibilities offered by the language system. Kennedy (*ibid.*:7-10) sees the computerized corpus as enabling generalisations to be made about language use, stressing that interest is typically not just in what occurs but in what is probable and what is likely to occur. He makes it clear that, while theories may be derived from corpus studies, corpus linguistics is not a linguistic theory. Corpus data may be combined with other sources of linguistic evidence, and may be used within various frameworks of linguistic description, focusing on a range of aspects of language use.

5. Contextualisation of language use in corpus-based translation studies

The focus of descriptive translation studies as formulated above was translation “as it actually occurs, now and in the past, as part of cultural history” (Hermans 1999:7). The contextualisation of translation plays a crucial role here and, as Hermans (*ibid.*:155-156) points out, the current trend is towards

foregrounding social, political and ideological contexts and effects, as exemplified by postcolonial or feminist approaches to translation. The study of translation in the context of power imbalance does not have to be confined to the specific contexts of postcolonialism, gender relations, ethnography etc. Referring to norms of translation, Hermans (2000b:12-13) asserts that they involve “different and often competing positions and possibilities, they point up various interests and stakes being pursued, defended, coveted, and claimed – and the desires and strategies of both individuals and collectives to further their own ends”.

A not uncommon view of corpus-based translation studies is that it is a methodology which allows linguistic and cultural-studies approaches to translation to be combined or integrated, and the effect of ideology on translation to be studied (Tymoczko 1998:657; Kohn 1996:47). These aims are clearly related to a tracing of the link between text and context, between “regularities of actual behaviour” (Toury 1995:265) and the aforementioned “interests and stakes being pursued”. Hermans is not as optimistic about the potential of corpus-based translation studies in this respect, asserting that “text-crunching” will, for example, tell us something about the linguistic make-up of texts, but nothing about their status, i.e. the extent to which translations are peripheral, or not, at a particular moment in a given culture (1999:93-94). Although acknowledging the potential usefulness of corpus-based studies in translation, Ian Mason (2001:71) also cautions against ignoring the rhetorical purposes which govern language production. Concentration on concordance data can lead to a lack of consideration of contextual and co-textual factors; he stresses the importance of looking at the influence of genre, discourse and textual purpose on choices made by source writer and translator, as well as other motivations, such as the communication goals of both text producers and the translator’s orientation or *skopos*. This cannot be done through vague generalisations based on quantitative data but requires a combination of quantitative and qualitative analyses to explore these pragmatic factors related to discourse, genres, and text designs (*ibid.*:78).

This is without doubt the most challenging area of corpus-based analyses, whether of translated or non-translated texts, and there is much scope for improvement in methodological approaches to these kinds of investigations. While quantitative “text-crunching” is relatively straightforward, it must be acknowledged that corpora provide certain kinds of data (e.g. frequency lists, concordances) which need to be integrated into an appropriate theoretical framework and combined with other data from other sources if studies are to transcend the trivial or the obvious. Thus, to return to Hermans’ example, the extent to which translations are peripheral at a given time in a given culture can best be discerned, not through corpus analysis but perhaps through evidence of the nature of their reception (reviews, distribution figures etc.). However, study of a corpus of translations deemed peripheral or non-peripheral could provide valuable evidence of the degree of normalisation or creativity in the linguistic make-up of these texts and thus go some way to establishing links between text and reception.

A study of contemporary literary translations in English (Olohan forthcoming b) shows a tendency for formality in fictional dialogue (e.g. through absence of direct speech and of contracted forms) which differs markedly from the representation of dialogue in a comparable corpus of literary texts from the British National Corpus. Similarly, a study of the use of moderators (*quite, fairly, pretty, rather*) in this literary translation corpus of the Translational English Corpus (Olohan forthcoming a) suggests a tendency for translators not to use these lexical items to the same extent as authors of English fictional texts. It is hypothesised that translated literary text exhibits a different kind of speaker-hearer or writer-reader interaction than comparable texts; not only are these texts characterised by less explicit interaction between characters or narrators in the form of direct speech and dialogue, but these data also suggest that the interaction or involvement is played down by less use of this set of degree-modifying adverbs, the function of which is primarily to signal the writer's or speaker's perception of the propositional content of an utterance. Reduced use of these signals may result in less successful "mediation of role", to use the term from Halliday's (1973:58) description of the interpersonal function of language. These data could be combined with a study of reviews and critiques, of the kind carried out by Peter Fawcett (2000), to ascertain to what extent the reception of these texts and their relatively peripheral status may have been influenced directly by these and other features of translation.

Comparative studies of texts produced by individual translators (Baker 2000 and forthcoming; Olohan forthcoming a) show that it is possible to develop concepts such as 'translator's style' but they also provide evidence for a wide range of variation of linguistic behaviour across translators and texts. Corpus analyses of the type referred to here have thus far done little more than pinpoint some of these features and choices, which may be linked to literary genre (e.g. fiction vs. biography), to use of specific narrative structures, to a translator's individual 'style' or even to editorial intervention. It is clear that in order to proceed beyond the observation of certain patterns or innovations, studies must be supplemented by information about the translation process⁴, but also more elusive information about the editorial process, and analysis at additional linguistic and textual levels which may rely little on corpus techniques. Thus Mason's view that linguistic features and translators' choices need to be considered in the context of genre, text purpose, discourse structure etc. is re-affirmed.

The cause-and-effect relations in these cases are far from straightforward and indeed may be impossible to establish. However, corpus-based studies of this type can provide us with hypotheses for further testing. For example, in the case of the modifiers above, it is reasonable to assume that the source languages for the literary translations have a diverse range of linguistic elements with which to convey the kinds of meanings rendered by *fairly, pretty, quite* and *rather* in English. The fact that these four items are used less in this corpus of translations than in a comparable collection of non-translated texts could have a number of causes, including:

- (a) the source languages do not ‘moderate’ as much as might be typical for English fiction texts and translators convey this lack of moderation by using fewer moderators in their translations
- (b) translators remove or downplay elements of ‘moderation’, perhaps as part of a (non-deliberate) process of disambiguation or explicitation

Obviously, a corpus-based study of translations and comparable texts alone cannot test (a) above. In order to do this, similar studies of moderation in other languages in both translations and non-translated texts would need to be carried out, as would further studies of other English texts and translations. As far as (b) is concerned, lower incidence of moderators in translation may correlate with other evidence to suggest that translators remove ambiguity and make explicit on a number of levels. One way of investigating this posited aspect of the translation process further would be to investigate other scalar modifiers in a similar way, and to contrast these with non-scalar maximisers (e.g. *completely*, *absolutely*, *totally*, *utterly*). Since the latter are generally unambiguous, they may be less likely than moderators to undergo disambiguation or explicitation processes in translation.

6. Conclusions

This paper arose out of a belief that corpus-linguistic methodologies are indeed useful in the study of aspects of translation, just as technological advances have rendered these methods increasingly useful for the study of language more generally (McEnery & Wilson 2001:195; Kennedy 1998:294). However, this brief discussion has highlighted the complexity of some of the central concerns of translation studies and their extension far beyond the analysis of linguistic data and lexical, syntactic and semantic studies. Corpus-based translation studies is confronted with issues related to the concept of ‘translation’ itself, the universality of translation as an activity, of features and norms of translation, before it can proceed to corpus compilation and data gathering. The study of data on the translation product cannot be separated from study of the translation process – the ‘features of translation’ discussed here and throughout much of the corpus-based translation studies literature would perhaps be more aptly labelled ‘features of the translation process’, since they refer to processes (simplification, explicitation etc.) of a cognitive nature which may be constrained and influenced by social, cultural and other factors. Yet, our means of empirically investigating these aspects of the translation process are far from perfect. The need to contextualise and contextualise translation and our study of it means that predominantly quantitative studies of corpus data are limited in their usefulness; we also need qualitative analysis and studies of data other than those extracted from corpora if we are to ‘go beyond the words on the page’ (Mason 2002). Describing aspects of the linguistic make-up of translations may be difficult, but establishing the causes for and effects of this make-up is almost

certainly more problematic. Thus, while translation studies benefits from the adoption of corpus-linguistic methods for the former, the latter requires a combination of corpus-based studies with studies of literary, social, historical, ideological and cognitive contexts.

Bibliography

- Aarts, J. (1999). "The description of language use." H. Hasselgård & S. Oksefjell (eds) (1999). *Out of Corpora: Studies in honour of Stig Johansson*. Amsterdam and Atlanta: Rodopi, 3-20.
- Baker, M. (1995). "Corpora in translation studies: An overview and some suggestions for future research." *Target* 7(2), 223-243.
- Baker, M. (1996). "Corpus-based translation studies: The challenges that lie ahead." H. Somers (ed.) (1996). *Terminology, LSP and Translation: Studies in language engineering, in honour of Juan C. Sager*. Amsterdam and Philadelphia: John Benjamins, 175-186.
- Baker, M. (2000). "Towards a methodology for investigating the style of a literary translator." *Target* 12(2), 241-266.
- Baker, M. (forthcoming). "A corpus-based view of similarity and difference in translation." Submitted to *Language and Literature*.
- Barlow, M. (2000). "Parallel texts in language teaching." S. P. Botley, A. M. McEnery & A. Wilson (eds) (2000). *Multilingual Corpora in Teaching and Research*. Amsterdam and Atlanta: Rodopi, 106-115.
- Biber, D., Conrad, S. & R. Reppen (1998). *Corpus Linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Chesterman, A. (1997). *Memes of Translation*. Amsterdam and Philadelphia: John Benjamins.
- Fawcett, P. (2000). "Translation in the broadsheets." *The Translator* 6(2), 295-307.
- Halliday, M.A.K. (1973). *Explorations in the Functions of Language*. New York: Elsevier.
- Halverson, S. (1999). "Conceptual work and the 'translation' concept." *Target* 11(1), 1-31.
- Hermans, T. (1995). "Toury's empiricism version one." *The Translator* 1(2), 215-223.
- Hermans, T. (1999). *Translation in Systems: Descriptive and system-oriented approaches explained*. Manchester: St. Jerome.
- Hermans, T. (2000a). "Self-reference, self-reflection and re-entering translation." D. de Geest et al. (eds) (2000). *Under Construction: Links for the site of literary theory*. Leuven: Leuven University Press, 259-274.
- Hermans, T. (2000b). "Norms of translation." P. France (ed.) (2000). *The Oxford Guide to Literature in English Translation*. Oxford: OUP, 10-15.
- Johansson, S. (1998). "On the role of corpora in cross-linguistic research." S. Johansson & S. Oksefjell (eds) (1998). *Corpora and Cross-Linguistic Research*. Amsterdam and Atlanta: Rodopi, 3-24.
- Johansson, S. (forthcoming). "Reflections on corpora and their uses in cross-linguistic research." F. Zanettin, S. Bernardini & D. Stewart (eds) (forthcoming). *Corpora in Translator Education*. Manchester: St. Jerome.
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London and New York: Rodopi.
- Kenny, D. (2001). *Lexis and Creativity in Translation: A corpus-based study*. Manchester: St. Jerome.

- Kohn, J. (1996). "What can (corpus) linguistics do for translation?" K. Klaudy, J. Lambert & A. Sohár (eds) (1996). *Translation Studies in Hungary*. Budapest: Scholastica, 39-52.
- Mason, I. (2001). "Translator behaviour and language usage." *Hermes* 26, 65-80.
- Mason, I. (2002). "The Analyst's Toolbox." Lecture given at *Translation Research Summer School*, Manchester, 28 June 2002.
- McEnery, T. & A. Wilson (2001). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Oakes, M. & T. McEnery (2000). "Bilingual text alignment – an overview." S. P. Botley, A. M. McEnery & A. Wilson (eds) (2000). *Multilingual Corpora in Teaching and Research*. Amsterdam and Atlanta: Rodopi, 1-37.
- Olohan, M. & M. Baker (2000). "Reporting *that* in translated English: Evidence for subconscious processes of explicitation?" *Across Languages and Cultures* 1(2), 141-158.
- Olohan, M. (forthcoming a). *Corpus-based Translation Studies*. London and New York: Routledge.
- Olohan, M. (forthcoming b). "How frequent are the contractions? A study of contracted forms in the Translational English Corpus." To appear in *Target* 15.
- Rohdenburg, G. (1996). "Cognitive complexity and increased grammatical explicitness in English." *Cognitive Linguistics* 7(2), 149-182.
- Simard, M et al. (2000). "Bilingual text alignment: Where do we draw the line?" S. P. Botley, A. M. McEnery & A. Wilson (eds) (2000). *Multilingual Corpora in Teaching and Research*. Amsterdam and Atlanta: Rodopi, 38-64.
- Toury, G. (1995). *Descriptive Translation Studies and beyond*. Amsterdam and Philadelphia: John Benjamins.
- Tymoczko, M. (1998). "Computerized corpora and the future of translation studies." *Meta* 43(4), 652-659.
- Tymoczko, M. (2002). "Connecting the two infinite orders: Research methods in translation studies." T. Hermans (ed.) (2002). *Crosscultural Transgressions. Research models in translation Studies II: Historical and ideological issues*. Manchester: St. Jerome, 9-25.

¹ See Chesterman 1997, Hermans 1999 and Hermans 2000 for more detailed discussion of norms in translation studies.

² It should be noted that there is a difference between the use of the term 'comparable corpus' here and in many corpus linguistics contexts, as can be seen for example in Johansson (forthcoming). The first corpus of the kind discussed here was the Translational English Corpus, designed and constructed under the direction of Mona Baker (see the CTIS website at <http://www.umist.ac.uk/ctis>). It consists of a corpus of translations in English (from a range of different source languages) and a comparable corpus element which is a set of English-language texts from the British National Corpus.

³ Dorothy Kenny (2001:8-16 and 30-33) provides a useful summary of the approach to linguistics introduced by J. R. Firth and now followed, either implicitly or explicitly by neo-Firthian linguists such as M.A.K. Halliday and John Sinclair, as well as other corpus linguists and translation theorists.

⁴ As part of the Translational English Corpus, data on the translation process are collected and stored with the corpus texts, thus enabling researchers to integrate information about the translator, source author and publishing contexts etc. into the analysis.