Meta
Journal des traducteurs
Translators' Journal

ΜΕΤΑ

A Computer-assisted Approach to the Analysis of Translation Shifts

Jeremy Munday

See table of contents

Explore this journal

Cite this article

Article abstract

This article is an analysis of shifts in Seventeen Poisoned Englishmen, Edith Grossman's English translation, of a novel in Spanish by Gabriel García Márquez. It makes use of a variety of basic tools of corpus linguistics as aids to the inductive exploration of texts.

érudit

A COMPUTER-ASSISTED APPROACH TO THE ANALYSIS OF TRANSLATION SHIFTS

JEREMY MUNDAY
University of Bradford, Bradford, United Kingdom

Résumé

Cet article analyse le phénomène du transfert dans une traduction littéraire de l'espagnol vers l'anglais. L'analyse est faite à l'aide de différents outils de base de la linguistique de corpus.

Abstract

This article is an analysis of shifts in Seventeen Poisoned Englishmen, Edith Grossman's English translation, of a novel in Spanish by Gabriel García Márquez. It makes use of a variety of basic tools of corpus linguistics as aids to the inductive exploration of texts.

1. INTRODUCTION

Gideon Toury, in his *Descriptive Translation Studies and Beyond*, sets out a method for the comparison of a Source Text (ST) and Target Text (TT) pair: (1) situating the target text within the target culture system; (2) comparing the ST and TT, identifying the relationships between the coupled pairs of ST and TT segments and attempting generalizations about the underlying concept or *norm* of translation; and (3) building the results into other studies and drawing implications for future translating (Toury 1995: 36-39 and 102). A crucial area of Toury's methodology is the second step: deciding what to look at in order to see relationships between generally small ST and TT segments. Toury sees a mapping of the TT on to the ST that yields "a series of (ad hoc) coupled pairs" (p. 77), with an inevitably "partial" and "indirect" comparison that will undergo "continuous revision" during the analytical process itself (p. 80). Research techniques are therefore not explicit enough to be truly objective, perhaps because trying to apply a detailed comparison over a whole text is typically "arduous and tedious" (Holmes 1988: 80).

The present paper is an attempt to look at and go beyond, second stage of Toury's analysis. It begins from the concept of *translation shifts*, which has been used in the past (notably by van Leuven-Zwart 1989, 1990) as a way of getting to grips with the small changes ("shifts") that build up cumulatively over a whole text as a result of the choices taken by or imposed on the translator. This paper thus takes into account translation as a "decision-making process" (Levy 1967: 1171). Importantly, the study will use basic tools of corpus linguistics and lexicography to enable accurate and rapid access to surface features over a whole text, reducing the arduousness and tedium of what has previously been a manual task. In addition, the relating of the results to larger computerized control corpora (such as the now readily available British National Corpus) promises a systematic way for the analysis to break out of the confines of a single pair of texts to enable preliminary consideration of the influence of typical target-language patterns and of the translator's specific idiolect in the

creation of shifts. These are some of the advantages which a computer-assisted approach is able to provide. They can be added to those stressed in recent writing on corpora and translation studies: Baker (1993, 1995) sees corpora as playing a key role in future descriptive translation studies; Opas and Kujamäki (1995), in a small computer-assisted study of translated stream of consciousness techniques in Samuel Beckett's novels, are confident that "analysing large text corpora will also help rid translation criticism of isolated observations which have undermined normative translation theory" (p. 291).

As an illustration and investigation of the possibilities offered by corpora in descriptive translation studies, a translation shift/corpus linguistic approach will be applied to the analysis of the English translation of the short story *Diecisiete ingleses envenenados* by Gabriel García Márquez.

2. THE CONCEPT OF "SHIFT" IN TRANSLATION

Catford (1965) uses the term "translation shifts," which he defines as "departures from formal correspondence in the process of going from SL to TL" (1965: 73). His strictly linguistic definition, illustrated with decontextualized and idealized sentences, is much expanded in later writing on translation shifts. Miko (1970), for instance, concentrates on stylistic rather than syntactic or semantic changes, while Popovic (1970: 85) asserts the importance of the shift concept as a way of bringing to light "the general system of the translation" along the lines of Toury's later *norms* concept.

The most detailed model of shift analysis has been developed by van Leuven-Zwart, published originally as a doctoral thesis in Dutch and then, in abbreviated form, in two articles in *Target* (van Leuven-Zwart 1989, 1990). Her model is "intended for the description of integral translations of fictional texts" (1989: 154) and comprises two complementary models: (1) **a comparative model** which involves a detailed manual classification of microstructural shifts (semantic, stylistic and pragmatic, modulation, modification and mutation) between the ST and the TT, and (2) **a descriptive model** which attempts to calculate the effects of the microstructural shifts on the macrostructural level using the three functions of language from systemic linguistics (*interpersonal, ideational* and *textual* functions) and discourse concepts taken from Leech and Short (1981). In van Leuven-Zwart's analysis (1990: 178), segmentation (i.e. word order change) and cohesion are highlighted as two areas where the effects of microstructural shifts are visible on the textual and interpersonal functions of language and the discourse level. Segmentation and cohesion will be a main focus of examination in sections 4.2 and 4.3 of the present paper.

Van Leuven-Zwart's method has been applied by around seventy of her postgraduate students on Dutch translations of mainly Spanish language literary texts (1990: 86). The results show that "around 70% of the translations [...] show a percentage of shifts of approximately 100%" (1990: 88), with a preponderance of semantic shifts and with specification and explanation seen frequently. She compares these with Levy's hypotheses (1969: 122) and suggests that the trend to explanation in her results corresponds to Levy's *Intellektualisierung*, but that frequent specification is at odds with his contention that translations normally exhibit levelling or generalization. Her conclusions (1990: 92-93) are that the translation strategy of the works she analyzes is TT-oriented, towards what Toury (1995) calls the *norm of acceptability*. This kind of descriptive, norm-centred analysis seems to fit with Toury's own recipe for shift analysis:

the identification of shifts is part of the *discovery* procedures only, i.e. a step towards the formulation of *explanatory hypotheses*. The latter, in turn, necessitate the establishment of the overall CONCEPT OF TRANSLATION underlying whatever corpus one sets out to investigate. (Toury 1995: 85)

In this way shift analysis is not directed at exposing translation errors or flaws, but is a means of getting at the norms which govern the translation process.

However, there are serious drawbacks to van Leuven-Zwart's model. The comparative model is "very complex and difficult" (Gentzler 1993: 137), which van Leuven-Zwart herself partly recognizes (1989: 153-154). It is extremely difficult to keep track of all the different kinds of shift as there are 8 different categories and 37 subcategories, not all clearly differentiated. Perhaps because of this, the model tends to focus on decontextualized word-chunks in extracts from novels. This throws up the question of whether it was actually possible for the seventy or so students collaborating on van Leuven-Zwart's analysis to have all interpreted the categories in the same way. This could well be considered a crucial flaw in the model, since one of the main aims of descriptive translation studies is to provide a replicable analytical process (Toury 1995: 3).

The need, therefore, seems to be for a workable and manageable way of combining some of the shift approach's attention to detail with the flexibility afforded by Toury's "coupled pairs" method. It is here that the computer may well be able to provide some of the answers, being inherently suited to providing accurate and rapid access to data from a whole text. This could free the researcher to concentrate on close analysis. This paper will study a complete short text and its translation to see how the computer can be used to isolate trends of linguistic shifts. Areas that might be fruitful in future research will also be identified.

3. THE COMPUTER AS AN AID TO SHIFT ANALYSIS

The pair of texts to be analyzed are as follows: as ST, a short story by Gabriel García Márquez, *Diecisiete ingleses envenenados*, published in Spanish in 1992; and, as TT, *Seventeen Poisoned Englishmen*, translated by Edith Grossman and published in 1993. First of all, in order for a computer analysis to be carried out, a text needs to be in an electronically readable format. This can be done by scanning the text onto computer disk in text or ASCII format. The scanning procedure, while much quicker than keying in the text, does encounter difficulty distinguishing accented characters or differentiating between, for example, the letter *l* and the figure *1*. The accuracy of the scanner can be as low as 85%, so the text needs to be checked manually after scanning (see Irizarry 1988 for a detailed account of the process). The next step is to proceed to basic word statistics analysis to give an idea of the comparative texture of the two texts before contextualizing the findings by looking at specific instances. Both forms of analysis can be done using the readily available *Wordsmith* program devised by Mike Scott (Scott 1996).

3.1. Frequency Lists

Wordsmith can analyse orthographic word-forms (a group of typographical characters with a space on either side), count them, provide basic statistics (word length, sentence length, etc.), and order the words either alphabetically or in descending order of frequency. Care must be taken to instruct the program to discount extra characters such as hyphens (otherwise, for example, *good-bye* would be calculated as two separate

word-forms, *good* and *bye*), and then to manually check the frequency lists to remove any remaining errors.

A frequency calculation for the ST and TT of this study is as follows:

	distinct words (*types*)	running words (*tokens*)	type/token ratio
Spanish original	1452	4498	32.3%
English translation	1387	4561	30.4%

Table 1
Basic word statistics

The term "distinct words" known as *types* refers to the number of distinct word-forms in the text (*ship* occurs 11 times in the English but represents only one type). "Running words" known as *tokens* comprise the total number of word-forms in the text (*ship* would constitute 11 tokens). The English translation can thus be seen to be somewhat longer than the original, in accordance with the generally held but rarely demonstrated statement that a translation will be longer than the original because it sometimes has to make explicit what is implicit in the original (see, for example, Vinay and Darbelnet 1958: 185; Nida and Taber 1969: 163). However, systemic differences between the two languages may account for some of the differences: Spanish usually omits subject pronouns, indicating person by the inflection of the verb (e.g. *pensó/she thought*). When subject pronouns, including the Spanish *se*, are removed from the calculations, the total number of tokens for the story is 4406 in Spanish and 4348 in English, a reversal of the position in table 1. The comparative length of the ST and TT may depend on many variables, and seems to be an area far more complex than previously thought and worthy of careful future investigation on other texts.

The third column in table 1 shows the type-token ratio, calculated here by dividing the number of tokens by the number of types,[1] which is a simple measure of the superficial lexical complexity of a text. The higher the percentage, the more varied the vocabulary — Baker (1995: 236-237) suggests comparing the type-token ratio of originals and translations to help "capture global patterning that contributes to the identification of translations as translations," and to see if there is generally a process of lexical simplification during translation, as has been suggested by Blum-Kulka and Levenston (1983). The fact that the English translation is shown to have an initially lower percentage type-token ratio than the original suggests that some lexical simplification may have taken place, but again language differences blur direct comparisons: for example, the Spanish tokens *primero*, *primera* and *primeras* correspond to the single token *first* in the English list. This is what Malone (1988: 36) calls Convergence, inevitable because of the systemic differences in translation. It cannot be construed as deliberate lexical simplification.

One way of trying to tackle the inflection problem is by manually lemmatizing the word frequency lists — that is, by grouping inflections of verbs, nouns and adjectives according to their base form. To give a clearer idea of the range of tokens, in the example above, *primera*, *primero* and *primeras* would be grouped under the single

form *primer**. A calculation of lemmatized types and tokens (minus pronouns) is as follows:

	lemmatized types	tokens (- pronouns)	type/token ratio
Spanish original	1171	4406	26.6%
English translation	1216	4348	28.0%

Table 2
Adjusted word statistics

This calculation now gives a slightly higher type-token ratio for the TT, but in view of the rawness of the figures, caution and further investigation on other texts is advised.

One reason for the rawness of the data is the way the computer functions, as it is good at analyzing orthographic characters but not their sense. Key problem-areas are: (1) word-forms coming from the same root (e.g. *eat*, *eating* and *eaten* will appear separately, and the addition of the Saxon genitive to the name of Prudencia Linero will produce a new word-form, *Linero's*); (2) multi-word units (e.g. *all at once* appears as three word-forms — *all*, *at* and *once*); and (3) polysemic words (e.g. the computer cannot differentiate between the verb *like*, meaning "gustar," and the comparative form, meaning "como"). Human analysis of the statistics is thus vital.

The first step to contextualizing the data is to examine the word frequency lists. However, the most frequent ten words in the ST and TT (see Table 3 below) do not tell us very much: they are all function words and many will be the same for any text in the respective languages.

Spanish	**English**
de (304 instances)	the (385 instances)
la (177)	and (151)
el (165)	of (140)
y (158)	in (129)
en (142)	a (122)
que (139)	she (111)
a (99)	to (102)
los (83)	her (85)
un (77)	was (77)
se (65)	that (58)

Table 3
First ten words in the story in descending order of frequency

Examination of the first ten lexical words in the frequency lists (extracted by hand and set out in Table 4) shows a different trend:

Spanish	English
Prudencia (29)	Prudencia (30)
Señora (28)	Señora (29)
Linero (28)	Linero (28)
dijo (19)	only (19)
dos (14)	said (19)
barco (13)	first (14)
cura (11)	man (13)
hotel (10)	priest (11)
único (9)	ship (11)
vez (9)	time (11)

Table 4
First ten lexical words in the frequency list

Basic elements of the story show through here: the main character, Señora Prudencia Linero, leaves her native town for the first time and travels by ship from South America to Naples. There she stays on her own in a hotel and meets a priest. But clearly this is of limited use! When the list is analyzed more closely, certain individual lexical items seem to be more prominent in one of the two languages, but even this can be misleading: *two*, for instance, occurs only 9 times in the English, and 14 times in the Spanish. Closer examination shows this is because it occurs several times in the compound *seventy-two*, which, disregarding the hyphen, is counted as a single word-form. By contrast, the Spanish *setenta y dos* would be counted as three. *Dos* is also translated as *both* in the case of "los dos cerrojos" and "con las dos manos." On the other hand, *first* is more prominent in the English mainly because it covers, as we saw above, the several Spanish forms of the adjective *primero*.

Such frequency lists are also, therefore, very raw data — it would be wrong to think that these lists can be readily used to compare the consistency of all one-to-one equivalences in a single text of this length. Furthermore, Van Peer (1989) issues a strong, general warning against computational approaches to literature, claiming that they are "trapped in the levels of grammar and lexis" unable to deal with connotation, figurative use, style and the "transient nature" of language:

> When stylistic features of a text have been transformed into numerical form, they acquire a status that actually prevents them from being perceived as language-for-communication as such (1989: 302).

This is a crucial point but not a negation of the approach. Frequency lists, by their very nature, inevitably tend to focus attention on single decontextualized lexical items. Any analysis of the lists must take into consideration the original context of those items. But they also give an overall idea of the texture of a text and supply "a set of hints or clues to the nature of a text [...] one can get an idea of what further information would be worth acquiring [...] and so focus an investigation" (Sinclair 1986: 188). In this way the computer serves as an aid to, and not a substitute for, human analysis.

3.2. Concordances

In order to be better able to look at lexical items in context, two further tools will prove invaluable — the *concordance* and the *intercalated text*. The KWIC (keyword in

context) concordance, described by Baker (1995: 225) as "the corpus analyst's stock-in-trade," allows all examples of a particular search-term or -terms (*keywords*) to be called up on screen with a co-text of around forty characters to either side of the key-word. In-house lexicographers with big dictionary publishers use complex analytical packages and very large corpora, which will be looked at briefly in section 5 below. For the purposes of small-scale research, simple and effective concordances can be produced by *Wordsmith* (Scott 1996) or the earlier *MicroConcord* package (Scott and Johns 1993). The keyword which the analyst is examining appears in the centre of the screen, and all the concordance lines can be sorted alphabetically according to the words which occur either to the right or the left of the keyword. For example, in *Seventeen Poisoned Englishmen* the first three lines of the concordance for the keyword *her*, ordered alphabetically according to the first word to the right, would be:

1) When he received no reply, he showed **her** a postcard from a pack he took out of his
2) othing to eat. The porter accompanied **her**. A cool breeze had begun to blow in from
3) ats. That was where the officer found **her** again a little before two, drowning in swe

Several word-forms may be processed simultaneously. This is useful for examining grammatical categories: typing in the search term "eat*/ate" where "*" stands for any letter or letters will give all the instances of "eat/eats/eating/eaten" as well as "ate." All English participial endings can be accessed by entering "*ing." Of course, once again manual checking is required to remove word-forms that do not figure in the investigation; the search term "*ing" might call up examples of "thing."

Comparison of the ST and TT can be facilitated by holding on disk an additional "intercalated" version of the text. This *intercalated text*, an idea of Nelia Scott's (personal communication), can be produced by *Wordsmith*, which inserts the corresponding sentence of the TT after each sentence of the ST. This kind of alignment can be invaluable. For example the analyst, examining a concordance line in the ST, simply needs to press one key to widen the co-text of the search word, bringing up on screen the juxtaposed TT sentence of the intercalated text, which contains the translation corresponding to the search term.

To summarize, the basic word statistics, type-token ratio and frequency lists give an idea of the general texture of the texts and are of special value in spotting possible fruitful areas of investigation. The KWIC concordance, in conjunction with the intercalated text, can then be used to very quickly call up all the instances of a particular targeted item in order to check them against the corresponding ST or TT term. This is a potentially powerful tool to help analyze translation shifts.

4. **COMPARISON OF *DIECISIETE INGLESES ENVENENADOS* AND *SEVENTEEN POISONED ENGLISHMEN***

The previous section compared the general texture of the two texts. The first 800 words of ST and TT were then analyzed in detail manually for the type of translation shifts which van Leuven-Zwart had categorized. This analysis showed that Edith Grossman's translation adhered closely to the structure and vocabulary of the original, but that there were numerous shifts, some unavoidable: semantic shifts were slightly more common than stylistic ones and specification occurred more often than generalization. These findings are broadly in line with van Leuven-Zwart's (see section 2 above). The concordance was then used to examine these trends more deeply and to look at the accumulation of small translation shifts over the *whole* text. It is relatively easy to follow through lines of analysis as they suggest themselves and in this way the

precision of the computer can be harnessed flexibly. For instance, early on the consistency of the translation of *casa* as *house/home* was examined. The phrase *su casa* occurred a number of times, *su* not always being translated as *her*. It quickly became clear that an important shift was occurring within the possessive pronoun selections and, as will be seen in 4.2 below, this led to a more precise study of the pronoun choices in the two texts.

4.1. Consistency of Individual Lexical Items

Clearly, where a word in one language covers a different semantic field from the other language it cannot always be translated the same way — so *casa* can be translated as *house* or *home*. *Casa* occurs six times in the Spanish and is translated three times as *home* and three times as *house*.

In some cases, however, the repetition of a particular lexical item can be important for a theme and the lexical cohesion of the text. In the story, Prudencia's ship arrives in a port where the wreckage of other vessels destroyed in the recent war are all too visible. The word *escombros* is used twice within three sentences, first translated as *wreckage* and then as *wrecks*. A direct repetition is missed, but the word-forms are very close. However, a few pages later, the narrator uses the word *escombros* again when recalling Prudencia's nursing of her ill husband:

> 1a. "Se le acabó la mitad de la vida en el dormitorio frente a los *escombros* del único hombre que había amado..." (DIS: 168)
> 1b. "Half her life had been spent in the bedroom facing the *ruins* of the only man she ever loved..." (SPE: 124) (in these and later examples, all italicized emphases are mine).

Although the selection of *ruins* in the TT conveys a sense of decay, its selection does mean that the patterns of *lexical cohesion* (see Halliday and Hasan 1976) have been altered. Of course, as Baker points out (1992: 206), the ST networks of cohesion cannot usually be fully replicated in the TT; however, a direct repetition of *wreckage* at this point might have helped preserve a stronger unity of past and present within the theme. In cases of lexical cohesion, the concordance, by giving rapid access to *all* instances of an item, can be a major help for the analyst in checking for consistency.

4.2. Shifts in Cohesion: *su* and *her*

Modification of the cohesive links within the narrative may even, as van Leuven-Zwart (1990: 85) suggests, affect the narrative function of a text. In the following pages, we shall examine an example of such cohesive links through a concordance-assisted tracking of the consistency of the possessive pronoun *su/her* (with sense of *Prudencia's*), an area of investigation that arose from the initial examination of the translations of *su casa*. Halliday and Hasan (1976: 31) call this area of cohesion, *reference*.

In the Spanish original there are 21 instances of *su/sus* being used with the sense *her* (i.e. Prudencia's). The translations that appear in the English can be seen in table 5 below:

her	17
the	1
These	1
(omitted)	2

Table 5
Translations of the 21 instances of *su/sus* with the sense *her*

In the English translation there are 47 instances of *her* as a possessive pronoun in the sense of *Prudencia's*. These instances were examined to see what they corresponded to in the original Spanish text. Table 6 below shows the break-down:

Spanish definite article	23
su/sus	17
idiom	5
dative	2

Table 6
Spanish originals corresponding to the translation of *her*

The raw figures indicate that the word *her* is more prominent in the translation than was *su/sus* in the original. This can be explained in some cases where, because of the difference in structure of the language, *her* would be the only really reasonable translation ("*la* garganta/*her* throat," "*el* alma/*her* soul," etc.). Close analysis of other examples, however, shows that greater emphasis has been placed on cohesion in the English. The following extract, occurring towards the end of the story when Prudencia returns to the hotel after a walk around Naples, shows that the translation has made an explicit link that was not so explicit in the original.

2a. Le entregó la llave d*el* cuarto, sin prestarle más atención [...] la señora Prudencia Linero pasó los cerrojos de *la* habitación. Luego rodó contra la puerta la mesita de escribir y la poltrona, y puso por último *el* baúl... (my emphases) (DSI: 175)
2b. She [the landlady] handed her the key to *her* room, and paid no further attention to her [...] Señora Prudencia Linero bolted the locks in *her* room. After that she pushed the little writing table and the easy chair and *her* trunk against the door... (SPI: 131)

On three occasions in this extract the Spanish definite article is translated by the possessive pronoun in English: "*el* cuarto/*her* room," "*la* habitación/*her* room," "*el* baúl/*her* trunk"). The changes in many ways shift the focus of the narrative. In the Spanish text, the referent of the definite articles are what Halliday and Hasan (1976: 71) would term "identifiable in the specific situation." Character, narrator and reader are assumed to have shared knowledge about the specific room and trunk. In the English translation with possessive pronouns, the cohesion is tighter and the reader is not assumed to share as much knowledge. The insistence on telling the reader specifically that the room is "hers" in fact distances the reader from the main character, moving the focus of the story away from Prudencia's internal monologue and fears and towards the impersonal narrator. This is not an isolated instance: on two other similar occasions, "*la* habitación" and "*el* dormitorio" are translated as "*her* room," so that there is an accumulation of similar translation shifts involving the possessive pronoun.

The most interesting and perhaps controversial use of *her* occurs in a passage when the ship is arriving in Naples and Prudencia is thinking about her fellow passengers:

3a. La señora Prudencia Linero, que había hecho tantos amigos viejos a bordo, que había cuidado niños mientras sus padres bailaban y hasta le había cosido un botón de la guerrera al primer oficial, los encontró *de pronto* ajenos y distintos. (DSI: 160-161)
3b. *To her surprise*, Señora Prudencia Linero, who had made so many dear old friends on board, who had watched children while their parents danced, and even sewn a button on the first officer's tunic, found them all distant and changed. (SPE: 117-118)

"To her surprise" replaces the "de pronto" of the Spanish. There seem to be two clear shifts here: (1) by placing "to her surprise" at the beginning of the sentence the translator has caused it to lose its surprise value — we are not made to wait for this important phrase; (2) the translation "to her surprise" (rather than the semantically closer "suddenly" or "all at once") shifts the focus from the ST description of a sudden mental process to a TT explanation of the effect it has on Prudencia. *De pronto*, without a personal pronoun, allows the ST reader to share whit Prudencia the sudden realization that everyone is so distant. However, the use of *her* in the translation ("to *her* surprise") seems to indicate a distancing of the TT reader from Prudencia's thoughts and the incorporation of an explanation from an external narrator. Furthermore, the word *surprise* also explains, in the way a narrator would, how Prudencia reacted.

The analysis in this section arose from looking at the referential cohesion realized by one pronoun (*su/her*) and is the kind of shift which is not easy to monitor with van Leuven-Zwart's decontextualized categories, which analyze texts segment by segment. This is where real benefits of a computer-assisted approach can be seen. Even when the analysis proceeds on the basis of a single lexical-item or grammatical-word search term, the computer gives the analyst access to the immediate co-text of each instance and to all examples of that instance in the text as a whole. It is therefore possible to follow an individual shift to see if it becomes a trend over the whole text. This can also be done with search terms comprising larger segments of texts, such as word order and sentence beginnings. The following section examines more deeply the kind of shift involving sentence-initial and -final elements seen in example 3 above.

4.3 Shifts in Word Order/Segmentation

Segmentation, or word order change, was seen by van Leuven-Zwart (see 2 above) to have an effect on the textual and interpersonal functions at discourse level. By tagging sentence initial elements with a code such as "<s>" in the ST and "<se>" in the TT, it is then possible to use the code as a search term to give a concordance of all sentence-initial or sentence-final elements. These elements can be ordered alphabetically and the most frequent elements will often form a visual pattern on the concordance screen. Closer analysis can then look at how and why these patterns differ in the two texts. Applied to García Márquez's story, this process reveals a number of shifts. For instance, time elements are sometimes pushed to the end of the sentence in the English:

4a. *Desde el amanecer* se habían visto las luces de tierra. (DSI: 159)

4b. The lights on the land had been visible *since daybreak*. (SPE: 116)

5a. - Sucede mucho en verano *por estas aguas*. (DSI: 161)

5b. 'It happens pretty often *in these waters* during the summer.' (SPE: 119)

The stresses are slightly different in the TT: in 4b, for instance, the narrative structure of time in first position ("desde el amanecer") has been replaced by the grammatical Subject and Actor ("lights"). In 5b, the information flow of the sentence is focussed towards the time element ("during the summer"), whereas the ST stresses location ("por estas aguas"). Of course, there are many variables and influences here, such as differing marked and unmarked patterns between languages and the effect of the translator's own idiolect. Thus, example 5a may illustrate the Spanish preference for placing adverbials of time before those of place (see Butt and Benjamin 1988: 400) and the translation may simply be following a similarly unmarked English order. In section 5 below we look at how investigation using larger corpora may help in determining how marked the choices are.

Adjuncts of manner and place also tend to be fronted in the translation, as was seen in the *de pronto/to her surprise* example in 3. In 6b and 7b below the narrative chronology is again slightly altered. Interestingly, example 6 is another instance of *de pronto*:

6a. La señora Prudencia Linero se sintió *de pronto* en un instante alucinado, metida en una jaula de gallinas... (DSI: 165-166)
6b. *All at once*, in a kind of hallucination, Señora Prudencia Linero felt that she was in a chicken cage... (SPE: 122)
7a. Ella vio un adolescente lánguido detrás de un mostrador de madera con incrustaciones de vidrios de colores *en el vestíbulo*... (DSI: 166)
7b. *In the foyer* she saw a languid adolescent behind a wooden counter with insets of colored glass... (SPE: 123)

In the two Spanish sentences, 6a and 7a, the Actor (Prudencia) is in first position followed by a verb of perception ("se sintió," "vio") and then the adverbial phrase ("de pronto," "en el vestíbulo"). The order in Spanish means that the reader experiences the event in the same order as Prudencia and from her point of view. In the translation, the placing of the adverbial adjunct in first position disrupts this process: "all at once" begins sentence 6b with an explanation of the event, and "in the foyer" places the outside world in first position in 7b. This may again distance the reader from Prudencia's experience.

Similar changes occur to other participants as well. In 8, the people living near the hotel crowd round the building to watch as the dead bodies of the Englishmen who had suffered fatal food poisoning are carried out:

8a. Los vecinos asomados a los balcones, y los curiosos bloqueados en la calle, los iban contando a coro, como en un estadio, *a medida que los sacaban*. (DSI: 174)
8b. *As they were brought out*, the neighbors watching from the balconies, and the people held back on the street, counted them in chorus as if they were in a stadium. (SPE: 131)

The filmic sequence of the Spanish, flowing from the onlookers, through the noise of their chanting, to the sight of the men being carried out one by one, is broken in the translation.

Though sometimes quite subtle, the changes of word order in these examples are frequent. Admittedly, however, analyzing word order changes between languages is somewhat problematic, with Spanish word order known to be complex and more flexible than English, with "unquantifiable factors like rhythm, context, register and psychology" affecting constituent ordering (Butt and Benjamin 1988: 394). Nevertheless, the illustrative examples from just one short story have shown how a mainly visual and almost cinematic point of view in the ST may be shifted in translation. The translator has often placed a circumstantial adjunct in first position and thereby caused a blurring of the narratorial viewpoint in the TT.

5. BEYOND THE ST-TT PAIR: RELATING THE FINDINGS TO NORMS VIA A CONTROL CORPUS

By comparing of ST and TT, it has been possible to identify numerous shifts in cohesion and word order/narrative point of view. The next step, following Toury's method (see introduction) is to make a more general statement about the norms adopted by the translator. From the examples shown in the analysis, it is clear that the translator has not felt bound by the syntactic and stylistic structures of the ST; in other words, she has not adopted a norm of *adequacy* (see Toury 1995: 57). How far Grossman has adopted the other pole, the norm of *acceptability*, or adherence to conventional target language patterns, is more difficult to decide. Here we enter the realm of conventional patterns in the source and target languages, where the texts are also marked by the indi-

vidual styles of García Márquez and Edith Grossman. There are many variables here and style is not easily quantifiable (see van Peer's comments on stylistic features in 3.1 above). However, as Leech and Short emphasize (1988: 71), quantitative information on stylistic features is important in order to provide "concrete evidence" to support statements about style. Since style can be considered the sum of linguistic choices made by both author and translator (Snell-Hornby 1988: 124), any way to more systematically compare these choices to established convention may determine the basis for statements concerning norms adopted by the translator.

One way of doing this may be to relate the findings from the ST-TT comparison to a larger control corpus to gauge how far the findings represent typical patterns of language and how far they constitute the creative idiolect of the author or translator. There are practical problems in doing this. Ideally, the features in the *Dieicisiete ingleses envenenados* text would be checked against a corpus of Latin American literary texts (although influences of the journalistic genre are also often noted in García Márquez's writing). The same process would then be carried out on *Seventeen Poisoned Englishmen* using a suitable English language corpus. The results would be compared to see if the translation shifts noted in sections 3 and 4 above represent a normalization to TT patterns or are more the result of Edith Grossman's idiolect. At the moment, such a delicate comparison is very difficult as the detailed checking of García Márquez's idiolect is prevented by the apparent absence of a suitable control corpus of Spanish (though it should be noted that useful general texture findings for a sample of twentieth-century Spanish literary texts have been put together by Irizarry (1990)).

Nevertheless, a cautious attempt will be made here to check the influence of the *translator's idiolect* on the word order shifts. Again, selecting a suitable and readily available control database is not easy, given the different variables — Edith Grossman is from New York, the genre is the modern literary short story. Two databases were selected for the analysis: the approximately 10-million-word Associated Press Corpus (APC) which was the closest that could be found to the translator's North American idiolect, and the 110-million-word British National Corpus (BNC), which contains some examples of literary genres.[2] Neither fits perfectly with our requirements, since the APC has a journalistic skew and the BNC a British English distortion; nevertheless, if the purpose is a tentative examination of the possible incorporation of large corpora into small descriptive studies, the results may well be illuminating.

Four phrases were selected for examination. These were circumstantial adjuncts which had displayed sentence order translation shifts in sections 4.2 and 4.3 above. They are:

in the foyer (shifted to sentence-initial position in example 7);
since daybreak (shifted to sentence-final position in example 4);
all at once (shifted to sentence-initial in example 6); and
to her surprise (shifted to sentence-initial in example 3).

In the APC, there is only a small number of instances of *in the foyer*. They can be broken down as follows:

sentence-/clause-initial	clause-middle	sentence-/clause-final
1	4	5

Table 7
Location of adjunct *in the foyer* in concordance from APC

As can be seen from the table, the most common location for the adjunct in the corpus is either clause- or sentence-final, as in:

"He would disclose only that there was broken glass *in the foyer*."

Another location is in the middle of the sentence:

"Following the concert, the glitzy crowd gathered *in the foyer* for a buffet supper."

The only sentence-initial instance involves the use of *foyer* in a more complex place adjunct:

"*On the first and second floors of the main foyer* about 100 Chinese corporations set up booths..."

We might tentatively suggest that there is a marked tendency in North American journalistic writing to focus attention on a clause with the *foyer* adjunct in first position. Looking at examples from the different genres of the BNC, we see that there are 10 instances of *in the foyer* in sentence-initial position, and 68 in sentence-/clause-final. This seems to bear out the markedness of the first position and would suggest that the translation shift in example 7 above (changing Spanish clause-final to English sentence-initial) does not conform to typical target language patterns. It may thus be a characteristic of Edith Grossman's own individual idiolect or her perception of the narrative structure of the story.

On the other hand, findings for *daybreak* (shifted to sentence-final in example 4) suggest the TT displays a process of normalization to target language structures. In the APC, the 26 examples of time adjuncts with *daybreak* show the following breakdown:

sentence-/clause-initial	clause-middle	sentence-/clause-final
1	5	20

Table 8
Location of adjuncts with *daybreak* in concordance from APC

The typical pattern is "the fast began at daybreak," "they took turns sitting with her until daybreak," and so on. Grossman's placing of *since daybreak* at the end of the sentence therefore seems to fit in with a typical pattern in the target language.

In example 6 above, the Spanish *de pronto*, occurring in the middle of a clause, had been translated as *all at once* and shifted to the beginning of the sentence. Examples of *all at once* (in the sense of "suddenly") in the APC suggest this is slightly normalizing:

sentence-/clause-initial	after initial conjunction	clause-middle	sentence-/clause-final
6	2	3	5

Table 9
Location of adjunct *all at once* in concordance from APC

Sentence-middle placement (as in the original Spanish) is not unusual, as in:

"It happened all at once after a noise that sounded like thunder."

But the translation does shift the adjunct to first position. Interestingly enough, many of the sentence-initial examples are direct speech ("All at once I saw the train coming toward me..."), perhaps reflecting the influence of spoken language on Grossman's translation. However, this is a very tentative conclusion needing to be followed up by detailed research.

In example 3 above, "La señora Prudencia Linero [...] los encontró *de pronto* ajenos y distintos" had been translated as "*To her surprise*, Señora Prudencia Linero [...] found them all distant and changed." We shall only consider the sentence position of *de pronto* rather than whether it could have been rendered as "*all at once*." The problem with using the APC was that there were very few examples of *to her surprise*. This immediately suggests that the phrase is not typical of the journalistic genre. For this reason, it was decided to use a larger sample from the British National Corpus, with its wider variety of texts. Here the breakdown was:

sentence-/ clause-initial	after initial conjunction	clause-middle	sentence-/ clause-final
39	40	16	17

Table 10
Location of adjunct *to her surprise* from BNC

This table suggests that *to her surprise* is normally placed in first position or immediately after a conjunction in first position:

"But *to her surprise*, he hadn't seemed angry at all."

Grossman's choice of word order thus seems strongly oriented to typical target language norms. This is all the more the case since maintaining the middle-clause location of the original, though less common, is far from infrequent. For example,

"Another more formal card, *to her surprise*, came from Ernest and Emily."

Even more interesting is that the phrase often occurs in that position with the verb form *found*:

"She found *to her surprise* that she didn't mind the clothes."

Using this as a model for a possible translation of our ST segment we could get:

"Señora Prudencia Linero [...] found them, *to her surprise*, all distant and changed."

This would preserve both the order and narratorial viewpoint of the ST. In contrast, Grossman's actual translation seems, on the one hand, to conform to the typical target language positioning of *to her surprise*, but, on the other, not to conform to the typical collocation of "she found to her surprise." Perhaps the influence of typical syntactic patterning is sometimes stronger than that of some collocations. However, this would need to be further investigated and checked in future research.

From this brief attempt at moving beyond the ST-TT pairs, it seems that:

the *foyer* shift may be a reflection of the idiolect of the translator;
the *daybreak* and *to her surprise* shifts seem to conform to typical target language norms;
the *all at once* shift is partly normalized towards the target language and may show influences of the spoken language;

the general pattern of placing adjuncts in first position certainly reduces the variety of the TT. It may also be part of the translator's idiolect and/or display a concern to focus certain elements that are deemed important for the narrative structure of the translated story.

6. CONCLUSION

It should be re-emphasized here that the translation examined is not erronious; nor does it intentionally distort the original narrative. Indeed, comparison of the illustrative texts reveals that Edith Grossman's translation closely follows the original Spanish. Yet, even with such a translation, small and not-so-small shifts accumulate in the translation process over the whole text and affect the texture of the story. The analysis of shifts, particularly those involving cohesion and word order, has been useful in identifying changes in the narrative viewpoint; in future research, expanding the shift concept to cover areas such as Transitivity patterns, character development, and Modality and writer-reader relationships will hopefully lead to the development of a comprehensive and replicable framework for analyzing shifts in literary translation, supplementing the approach developed by Toury (see introduction).

This sort of descriptive study is greatly facilitated by the availability of tools of corpus linguistics. The *Wordsmith* program used in this investigation is user-friendly and has proved very practical as an aid to human analysis of a whole text. Word count and frequency statistics have given hints to the comparative texture of the ST and TT and have suggested fruitful areas of investigation (for example the consistency of *su/her*). The concordance program has also given rapid access to all instances of a given phenomenon and freed the researcher to concentrate on a close critical analysis of the results. In this way the computer may be able to revolutionize and encourage descriptive translation studies by reducing the arduousness of locating all the instances of a term, and by enabling small text segments to be quickly related to the text as a whole. In addition, the tentative attempt to relate the findings to the patterns of language in larger control corpora such as the APC and BNC (section 5) has suggested a more systematic way to establish the translation norms.

Future areas of development are many. Some preliminary tagging of the ST and TT for features that are likely to undergo translation shifts might be useful for the researcher; the tagging could be grammatical (to look more closely at clause beginnings or shifts from noun to verb, for example), functional (such as analysis of Transitivity patterns) or stylistic (the highlighting of the occurrence of particular lexical fields, an author's favourite constructions, words with positive and negative connotations, etc.). Refinement of the large control corpora according to genre or geographical area would also facilitate comparison of the results against a corpus of similar status. For literary analysis, it would also be useful to have a ready-analyzed database of individual writers since norms of usage in literature are often intentionally distorted by authors and it is unclear how far translators are permitted to go in distorting the norms of the target language. No doubt there will be growing use of corpora in the work of practising translators themselves as a means of attaining consistency and as a complement and counterbalance to the uncertainties of intuition.

Notes

1.   *Wordsmith* will normally calculate a type-token ratio per 1000 words, which is useful for comparing texts of different lengths. Since the ST and TT in this study were of similar length, the simpler ratio was used.
2.   Both the APC and British National Corpus are used for lexicographic purposes at Oxford University Press. My grateful thanks go to OUP for allowing me access to the databases for research purposes.

REFERENCES

BAKER, Mona (1992): *In Other Words: a Coursebook on Translation*, London, Routledge.

BAKER, Mona (1993): "Corpus Linguistics and Translation Studies: Implications and Applications", *Text and Technology: in Honour of John Sinclair*, Mona Baker, Gill Francis & Elena Tognini-Bonelli (Eds), Philadelphia and Amsterdam, John Benjamins, pp. 233-50.

BAKER, Mona (1995): "Corpora in Translation Studies: An Overview and Suggestions for Future Research", *Target*, 7 (2), pp. 223-243.

BLUM-KULKA, Shoshona & Eddie LEVENSTON (1983): "Universals of Lexical Simplification", *Strategies in Interlanguage Communication*, Claus Færch & Gabriele Kasper (Eds), London & New York, Longman, pp. 119-139.

BUTT, John & Carmen BENJAMIN (1988/1994): *A New Reference Grammar of Modern Spanish*, London, Edward Arnold.

CATFORD, J. C. (1965): *A Linguistic Theory of Translation*, London, Oxford University Press.

GENTZLER, Edwin (1993): *Contemporary Translation Theories*, London & New York, Routledge.

HALLIDAY, Michael A.K. & Ruqaiya HASAN (1976): *Cohesion in English*, London, Longman.

HOLMES, James S. (1988): *Translated!: Papers on Literary Translation and Translation Studies*, Amsterdam, Rodopi.

IRIZARRY, Estelle (1988): "Literary Analysis and the Microcomputer", *Hispania*, 71 (4), pp. 984-995.

IRIZARRY, Estelle (1990): "Stylistic Analysis of a Corpus of Twentieth-Century Spanish Narrative," *Computers and the Humanities*, 24 (4), pp. 265-274.

LEECH, Geoffrey & Michael SHORT (1988): *Style in Fiction*, London, Longman.

LEUVEN-ZWART, Kitty M. van (1989): "Translation and Original: Similarities and Dissimilarities, I", *Target*, 1 (2), pp. 151-181.

LEUVEN-ZWART, Kitty M. van (1990): "Translation and Original: Similarities and Dissimilarities, II", *Target*, 2 (1), pp. 69-95.

LEVY, Jirí (1967): "Translation as a Decision Process", *To Honor Roman Jakobson II*, (Series Maior 23), The Hague, Mouton de Gruyter, pp. 1171-1182.

LEVY, Jirí (1969): *Die Literarische (bersetzung). Theorie einer Kunstgattung*, Frankfurt, Athenäum Verlag (Translated from the Czech original by Walter Schamschula).

MIKO, Frantisek (1970): "La théorie de l'expression et la traduction", James S. Holmes (Ed.), *The Nature of Translation*, The Hague & Paris, Mouton de Gruyter, pp. 61-77.

MALONE, Joseph L. (1988): *The Science of Linguistics in the Art of Translation*, Albany, New York, State University of New York Press.

NIDA, Eugene A. & Charles R. TABER (1969/1974): *The Theory and Practice of Translation,* Leiden, Netherlands, E.J. Brill.

OPAS, Lisa Lena & Pekka KUJAMÄKI (1995): "A Cross-linguistic Study of Stream-of-Consciousness Techniques", *Literary and Linguistic Computing*, 10 (4), pp. 287-291.

PEER, Willie van (1989): "Quantitative Studies of Literature, A Critique and an Outlook", *Computers and the Humanities*, 23, pp. 301-307.

POPOVIC, Anton (1970): "The Concept "Shift of Expression" in Translation Analysis", *The Nature of Translation*, James S. Holmes (Ed.), The Hague & Paris, Mouton de Gruyter, pp. 78-87.

SCOTT, Mike (1996): *Wordsmith* (Program), Oxford, Oxford University Press.

SCOTT, Mike & Tim JOHNS (1993): *MicroConcord* (Program), Oxford, Oxford University Press.

SINCLAIR, John M. (1986): "Basic Computer Processing of Long Texts", Geoffrey Leech & Christopher N. Candlin (Eds), *Computers in English Language Teaching and Research*, Harlow & New York, Longman, pp. 184-203.

SNELL-HORNBY, Mary (1988/1995): *Translation Studies: An Integrated Approach,* Amsterdam & Philadelphia, John Benjamins.

TOURY, Gideon (1995): *Descriptive Translation Studies and Beyond*, Amsterdam & Philadelphia, John Benjamins.

VINAY, Jean-Paul & Jean DARBELNET (1958): *Stylistique comparée du français et de l'anglais: méthode de traduction*, Paris, Didier.

Illustrative examples taken from:

(DIS) Gabriel García MÁRQUEZ (1992): "Diecisiete ingleses envenenados", *Doce cuentos peregrinos*, Madrid, Mondadori, pp. 157-175.

(SPE) Gabriel García MÁRQUEZ (1994): "Seventeen Poisoned Englishmen", translated by Edith Grossman, *Strange Pilgrims*, Harmondsworth, Penguin UK, pp. 116-132.