Spelling out the optionals in translation: a corpus study

Maeve Olohan
Centre for Translation and Intercultural Studies, UMIST
PO Box 88, Manchester M60 1QD
maeve.olohan@umist.ac.uk

Abstract

While the use of translations in parallel corpora, mostly for the purposes of contrastive linguistic analysis, is relatively well established, the analysis of translated language as an object of study in its own right has only fairly recently been made possible through the development of corpus resources designed specifically for this purpose. The Translational English Corpus (TEC) at UMIST was the first corpus consisting exclusively of translations, in English, from a variety of source languages and text types. Much of the research carried out thus far using TEC (e.g. Laviosa-Braithwaite 1996, Kenny 1999 and 2000) has been interested in identifying and confirming features of translated language such as explicitation, normalisation, simplification and levelling out (Baker 1996). This kind of research is based on the assumption that, by retrieving and analysing data from TEC and a comparable corpus (e.g. the British National Corpus), it is possible to pinpoint consistent differences in syntactic or lexical patterning between translated English and original English. Some of these may arise from deliberate translation strategies on the part of the translator who wishes to make his/her text more explicit, to normalise or simplify etc. However, TEC can also be used as a means of identifying linguistic patterning which translators will not have been aware of producing, but which occurs as a result of the complex nature of the translation activity itself.

Against this background, this paper presents an investigation of explicitation in translation. Preliminary studies using TEC and a subcorpus of the BNC (Burnett 1999, Olohan and Baker 2000) have shown that patterns of use of the optional *that* with reporting verbs are rather different in translated English than in original English, with translated English very much favouring the use of *that*, even in contexts which do not warrant it, e.g. for purposes of disambiguation or for the signalling of more formal style. This paper will present further analysis of optional syntactic features in English and their occurrence in TEC and the BNC, test the hypothesis that translated English displays a higher incidence of a range of optional syntactic features than is observed in a comparable corpus of original English, and that this is direct evidence of subconscious processes of explicitation in translation.

1. Corpus-based translation studies

Corpus-based translation studies is a relatively new area of research within translation studies, motivated by an interest in the study of translated texts as instances of language use in their own right. This is in contrast to the not uncommon perception of translations as 'deviant' language use, a view which has generally led to the exclusion of translated texts from most 'standard' or 'national' corpora (Baker 1999). While translations have been seen as useful in parallel bilingual or multilingual corpora, this has usually been for contrastive linguistic analysis which has studied the relationship between source and target language systems or usage. Parallel corpora are naturally also of interest to the translation scholar as they facilitate investigation of the relationship between a translation and its source. Recent work using corpora in translation studies has, however, been more concerned with building corpora of translations so that the use of language in translations may be studied. The first corpus of this nature was the Translational English Corpus at UMIST (described below) which, since its inception, has provided the impetus and inspiration for a number of similar projects for other languages, including Italian, German, Spanish, Finnish, Catalan and Brazilian Portuguese.

One of the fundamental concepts in corpus-based translation studies has been the notion of comparable corpus, defined by Baker (1995: 234) as 'two separate collections of texts in the same language: one corpus consists of original texts in the language in question and the other consists of translations in that language from a give source language or languages…both corpora should cover a similar domain, variety of language and time span, and be of comparable length'. Baker's initial groundbreaking work posited a number of features of translation, or 'translation universals', which could be investigated using comparable corpora (Baker 1996). While the term universal in this context is somewhat controversial, not least because of the practical difficulties involved in testing whether something holds true across diverse languages (for many of which corpora of translations and/or original writing do not exist), it has been suggested, for example, that translations tend to be more explicit on a number of levels than original texts, and that they simplify and normalise or standardise in a number of ways. Much of the corpus-based work carried out to date has focused on syntactic or

lexical features of translated and original texts which may provide evidence of these 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 also allows us to investigate aspects of translators' use of language which are not the result of deliberate, controlled processes and of which translators may not be aware.

2. Corpus data

The Translational English Corpus is a corpus of translated English held at the Centre for Translation and Intercultural Studies at UMIST. It was designed specifically for the purpose of studying translated texts and it consists of contemporary written translations into English of texts from a range of source texts and languages. At the time of writing, it has over 6.4 million words. TEC consists of four text types – fiction, in-flight magazines, biography and newspaper articles – with fiction representing 82%, and biography and fiction together making up 96% of the corpus. The translations were published from 1983 onwards and were produced by translators, male and female, with English as their native language or language of habitual use.

The corpus of original English put together for this particular study is a subset of the BNC made up of texts from the imaginary domain. It is thus comparable in terms of genre and publication dates (from 1981 onwards). The texts have been produced by native speakers of English, both male and female. A minor difference between the two corpora which is not significant for current investigations is that TEC consists of full running texts whereas some of the BNC texts are extracts (some as long as 40,000 words). There is a little variation in size between the two corpora with TEC now slightly bigger than the BNC corpus. As TEC continues to grow, new texts will be added to the BNC subcorpus so that the corpora remain comparable in all respects.

The data discussed here was extracted from these two untagged corpora using Wordsmith Tools V.3.0.

3. Explicitation

The analyses reported on here arose from an interest in studying processes of explicitation in translation, where explicitation refers to the spelling out in target text of information which is only implicit in a source text. This has long been considered a feature of translation and has been investigated by a number of scholars (e.g. Vanderauwera 1985, Blum-Kulka 1986; Laviosa-Braithwaite 1996; Laviosa 1998; Baker 1995, 1996) who have identified different means or techniques by which translators make information explicit, e.g. using supplementary explanatory phrases, resolving source text ambiguities, making greater use of repetitions and other cohesive devices. This current research focuses, in so far as this is possible, on subconscious processes of explicitation and their realisation in linguistic forms in translated texts. Since the starting point is the linguistic form, we have concentrated on optional syntactic features, hypothesising that, if explicitation is genuinely an inherent feature of translation, translated text might manifest a higher frequency of the use of optional syntactic elements than original writing in the same language, i.e. translations may render grammatical relations more explicit more often – and perhaps in linguistic environments where there is no obvious justification for doing so – than original writers.

4. Analysis of optional syntactic features in English

Linguists may present the optional syntactic features of English in different ways, but we opted to base this study on Dixon's (1991: 68-71) omission conventions for English, presented in summary form as follows:

A. Omission of subject NP
B. Omission of complementiser *that*
C. Omission of relative pronoun *wh-/that*
D. Omission of *to be* from complement clause
E. Omission of predicate
F. Omission of modal *should* from a THAT complement
G. Omission of preposition before complementisers *that*, *for* and *to*
H. Omission of complementiser *to*
I. Omission of *after/while* in *(after) having* and *(while) *ing*
J. Omission of *in order*

These features span a range of linguistic phenomena, from frequently occurring relative pronouns to much less common constructions (e.g. *to be* in complement clause), and that they do not focus exclusively on optionality of omission. As will be obvious from the discussion below, they also vary considerably in terms of their identification and quantifiability in a corpus which is neither tagged nor parsed. In some instances, as can be seen in 4.3, 4.4, 4.9 and 4.10, omission is difficult to measure but occurrence, i.e. inclusion, can be traced and compared across corpora to give an indication of differences in usage of the longer surface form between corpora.

4.1. Omission of subject NP

This refers to omission of a subject NP in a number of circumstances, e.g. under coordination, in subordinate time clauses, from an ING complement clause or from a modal (FOR) TO complement clause. There is no obvious way of finding instances of these in a corpus which is not tagged for parts of speech.

4.2. Omission of complementiser *that*

Dixon states that 'the initial *that* may often be omitted from a complement clause when it immediately follows the main clause predicate (or predicate-plus-object-NP where the predicate head is *promise* or threaten' (1991: 70). An extensive analysis of the use of *that/zero*-connective with reporting verbs SAY and TELL, with reference to TEC and BNC, is presented in Olohan and Baker (2000). The results are summarised in Tables 1 and 2 below, which present both the absolute values (i.e. occurrences) and the percentages for each form:

Form	***say*** **(TEC)**	***say*** **(BNC)**	***said*** **(TEC)**	***said*** **(BNC)**	***says*** **(TEC)**	***says*** **(BNC)**	***saying*** **(TEC)**	***saying*** **(BNC)**
that	316	323	267	183	116	64	76	142
	55.5%	**26.5%**	**46.5%**	**19.2%**	**40.4%**	**12.8%**	**67.3%**	**43.0%**
zero	253	895	307	771	171	435	37	188
	44.5%	**73.5%**	**53.5%**	**80.8%**	**59.6%**	**87.2%**	**32.7%**	**57.0%**

Table 1: SAY + *that/zero* in BNC and TEC

Form	***tell*** **(TEC)**	***tell*** **(BNC)**	***told*** **(TEC)**	***told*** **(BNC)**	***tells*** **(TEC)**	***tells*** **(BNC)**	***telling*** **(TEC)**	***telling*** **(BNC)**
that	247	300	353	584	55	28	64	85
	62.8%	**38.2%**	**60%**	**43.6%**	**68.7%**	**37.5%**	**73.6%**	**42.3%**
zero	146	486	233	755	25	52	23	115
	37.2%	**61.8%**	**40%**	**56.4%**	**31.3%**	**62.5%**	**26.4%**	**57.7%**

Table 2: TELL + *that/zero* in BNC and TEC

It is immediately clear that the *that*-connective is far more frequent in TEC than in BNC. With the exception of *said* and *says*, *that* occurs more often than *zero* for all forms of SAY and TELL in TEC. By contrast, the *zero*-connective is more frequent for all forms of both verbs in the BNC corpus. These differences have been proven to be statistically significant. Furthermore, the results of the SAY and TELL study were consistent with findings by Burnett (1999) who reviewed use of the verbs SUGGEST, ADMIT, CLAIM, THINK, BELIEVE, HOPE and KNOW in TEC and BNC . While that study did not include all forms of these verbs, the data available shows that the *that*-connective is far more common than the *zero*-connective in translated than in original English for forms of all seven of the verbs investigated. The hypothesis that the optional *that* in reporting constructions occurs proportionately more frequently in translated texts than in original English texts is thus supported. Although Olohan and Baker (2000) highlight the relative vagueness with which omission and inclusion are accounted for in the linguistics literature, and the lack of guidance on this in reference works for users of English, there are clear patterns of usage in contemporary English writing as evidenced in the BNC corpus, and there is an equally clear contrast between these patterns and those perceived in translated English.

A brief analysis of one of the verbs suggested by Dixon, namely PROMISE, serves as further illustration and corroboration. Table 3 and Figure 1 below show that, although the number of instances of promise + *that/zero* were almost identical in the two corpora (135 in BNC and 131 in TEC), the relationship between *that* and *zero* in TEC (*that* = 67.9%, *zero* = 32.1%) is almost directly inverse to that in BNC (*that* = 34.1%, *zero* = 67.9%).

			That/zero		Total
			zero	**that**	
Corpus	**BNC**	Count	89	46	135
		% within Corpus	65.9%	34.1%	100.0%
		% within That/zero	67.9%	34.1%	50.8%
		% of Total	33.5%	17.3%	50.8%
Corpus	**TEC**	Count	42	89	131
		% within Corpus	32.1%	67.9%	100.0%
		% within That/zero	32.1%	65.9%	49.2%
		% of Total	15.8%	33.5%	49.2%

Table 3: PROMISE + *that/zero* in BNC and TEC

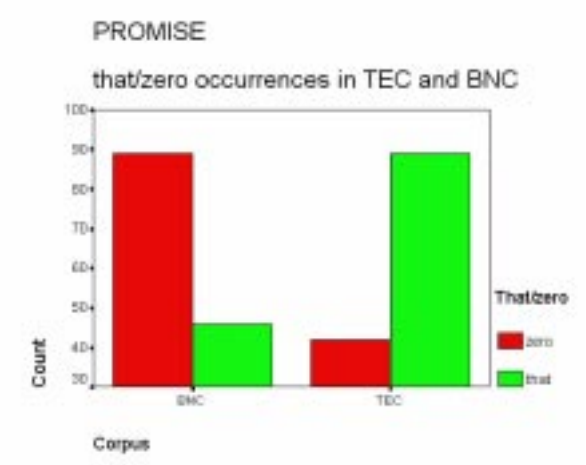

Figure 1: occurrences of PROMISE + *that/zero* in BNC and TEC

A breakdown of each lexical item (Table 4 and Figure 2) shows that this holds true for all forms of the verb, although some have low occurrences in general (e.g. *promises + that/zero* occurs only twice in TEC and not at all in BNC).

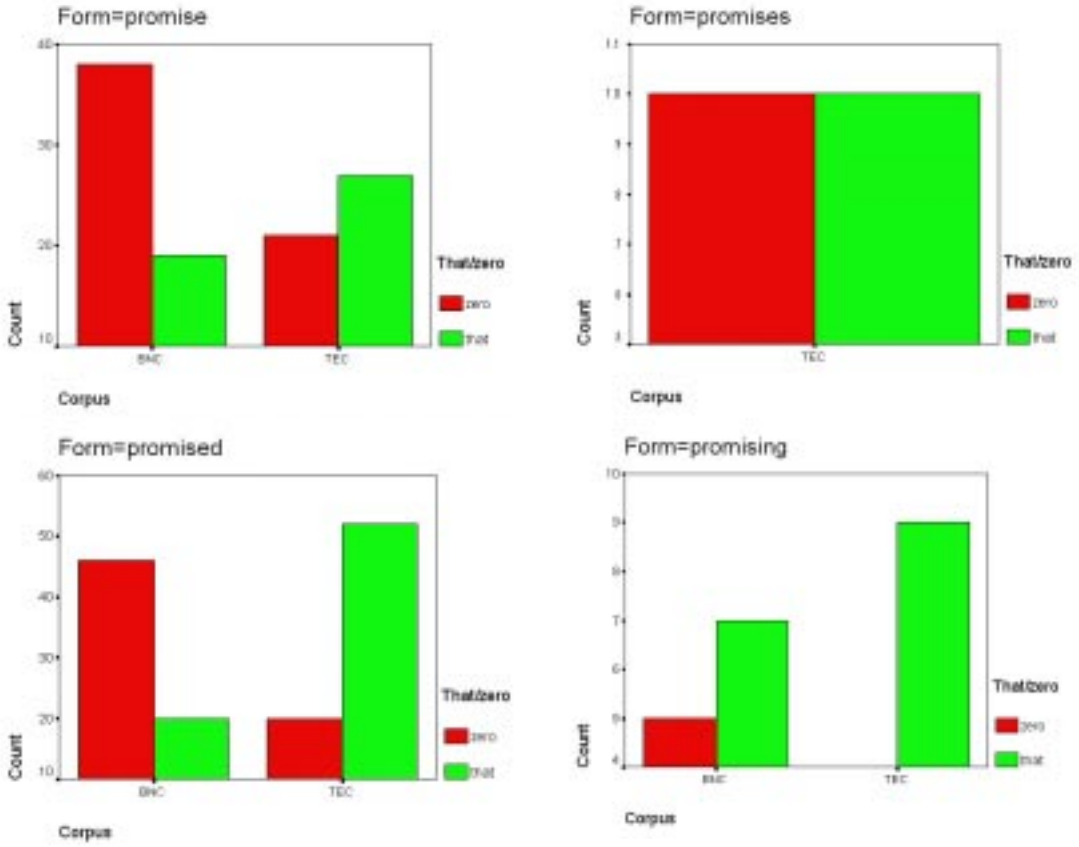

Figure 2: All forms of PROMISE + *that/zero* in BNC and TEC

			Form	That/zero zero	that	Total
promise	Corpus	BNC	Count	38	19	57
			% within Corpus	66.7%	33.3%	100.0%
			% within That/zero	64.4%	41.3%	54.3%
			% of Total	36.2%	18.1%	54.3%
		TEC	Count	21	27	48
			% within Corpus	43.8%	56.3%	100.0%
			% within That/zero	35.6%	58.7%	45.7%
			% of Total	20.0%	25.7%	45.7%
promises	Corpus	TEC	Count	1	1	2
			% within Corpus	50.0%	50.0%	100.0%
			% within That/zero	100.0%	100.0%	100.0%
			% of Total	50.0%	50.0%	100.0%
promised	Corpus	BNC	Count	46	20	66
			% within Corpus	69.7%	30.3%	100.0%
			% within That/zero	69.7%	27.8%	47.8%
			% of Total	33.3%	14.5%	47.8%
		TEC	Count	20	52	72
			% within Corpus	27.8%	72.2%	100.0%
			% within That/zero	30.3%	72.2%	52.2%
			% of Total	14.5%	37.7%	52.2%
promising	Corpus	BNC	Count	5	7	12
			% within Corpus	41.7%	58.3%	100.0%
			% within That/zero	100.0%	43.8%	57.1%
			% of Total	23.8%	33.3%	57.1%
		TEC	Count		9	9
			% within Corpus		100.0%	100.0%
			% within That/zero		56.3%	42.9%
			% of Total		42.9%	42.9%

Table 4: All forms of PROMISE + *that/zero* in BNC and TEC

4.3. Omission of relative pronoun *wh-/that*

This frequently occurring construction is difficult to measure in an untagged corpus. Thus far, only total counts of occurrence of *which* have been taken, with 11,201 in BNC and 23,607 in TEC. A first step in discarding irrelevant instances was to identify sentence-initial and sentence-final/clause-final *which*. Their removal leaves 10,457 concordance lines in BNC and 22,483 in TEC, indicating considerably higher usage of *which* in TEC. Obviously further detailed analysis of these instances is required to identify the occurrences in relative clauses where the coreferential NP is not in subject function in the relative clause, i.e. where omission could have taken place.

4.4. Omission of *to be* from complement clause

From a very frequent feature above, we come to a very infrequent structure. Dixon is referring here to the omission of *to be* with 'some verbs taking a Judgement TO complement clause, whose VP begins with *be*' (1991: 70), with an example of *thought* + *to be* + modifier. Both THINK + *to be* and FIND + *to be* were investigated in the corpora (see Table 5). The most common occurrence in both corpora was for the past tense forms (*thought* and *found*), and TEC exhibits a greater tendency overall to include *to be*, but the number of occurrences overall was very small in both corpora.

Form	BNC	TEC
THINK (+ *)(+ *) *to be*	2	6
FIND (+ *)(+ *) *to be*	4	7

Table 5: think + to be and find + to be in BNC and TEC

4.5. Omission of predicate

The omission of the predicate in coordinated clauses is difficult to capture in an untagged corpus and this has therefore not yet been investigated.

4.6. Omission of modal *should* from a THAT complement

This refers to the omission of modal *should* from a THAT complement with examples of verbs ORDER and SUGGEST. Neither is particularly common, and both occur predominantly in the past tense form (*ordered* and *suggested*). A greater proportion of omission is seen in TEC (see Table 6).

Form	BNC	TEC
ORDER + *that* + *should*	1	6
ORDER + *that* + *zero*	2	7
SUGGEST + *that* + *should*	19	19
SUGGEST + *that* + *zero*	43	58

Table 6: ORDER and SUGGEST + *that* + *should/zero* in BNC and TEC

4.7. Omission of preposition before complementisers *that, for* and *to*

Some transitive verbs with a preposition as last element in their lexical form which may take a complement clause in object function will omit the preposition before *that*, *for* and *to*, e.g. *he confessed to the crime, he confessed to strangling her*, but *he confessed that he had strangled her*. This is not an optional omission and is therefore not of interest in this study.

4.8. Omission of complementiser *to*

According to Dixon, the complementiser *to* is optional following HELP or KNOW. The form help was analysed, first discarding all uses of help as noun, as reflexive verb, verb + ING complement and verb + preposition, and then looking at occurrences of help (*) (*) to in detail (Table 7).

Form	BNC		TEC	
	Total occurrences	Relevant occurrences	Total occurrences	Relevant occurrences
Occurrences of *help*	2374	300	1792	365
help + *to*	62	26	72	38
help + * + *to*	67	50	98	80
help +* + * + *to*	19	3	35	19
Total help (+*) (+*) + *to*		79		137
help (+*) (+*) + zero		229		228

Table 7: *help* (+*) (+*) + *to* in BNC and TEC

This data tells us that although the word form *help* is more frequent in TEC, its verbal use in both corpora is quite similar with *help* (+*) (+*) + *to/zero* occurring slightly more often in TEC than in BNC, of which the complementiser *to* is used in 37.5% of TEC instances, compared with 26% of the BNC occurrences.

4.9. Omission of *after/while* in *(after) having + participle* and *(while) *ing*

As in 4.3 and 4.4 above and 4.10 below, we can more readily measure occurrence of these features rather than omission. Concordances of *while *ing* are pruned, discarding constructions such as *all the while *ing, after/in/for a while *ing, worth your while *ing*. The *while *ing* construction is much more frequent in TEC overall and in relation to the gerundial use (Table 8).

Form	BNC	TEC
Total *while *ing* concordances	150	360
Relevant concordances	138	330

Table 8: *while *ing* in BNC and TEC

A count of *after *ing *ed* (which obviously does not take irregularly formed past participles into account) also shows a tendency for TEC to use this construction more frequently than BNC (Table 9).

Form	BNC	TEC
*after *ing *ed*	11	65

Table 9: *after *ing *ed* in BNC and TEC

4.10. Omission of *in order*

According to Dixon, *in order* is usually omitted before *to* and may occasionally be omitted before *for* or *that*. While the investigation of every instance of the items *to, that* and *for* to see whether an *in order* has been omitted is not practical, we can easily measure usage of *in order to, in order for* and *in order that* and compare results from the two corpora. This investigation yields the following (Table 10):

Form	BNC	TEC
in order to	250	1225
in order for	1	14
in order that	12	18
Total	263	1257

Table 10: *in order to/for/that* in BNC and TEC

This does not conclusively prove that *in order* has been omitted more often in BNC but certainly indicates that the longer forms of the conjunctions appear with markedly higher frequency in TEC.

5. Correlations, contractions, co-occurrences

To return to the notion of explicitation then, it could be claimed, on the basis of these measures of inclusion and/or omission of optional syntactic elements above, that the language of TEC makes explicit grammatical and lexical relations which are less likely to be made explicit in original English. Furthermore, this tendency not to omit optional syntactic elements may be considered subliminal or subconscious rather than a result of deliberate decision-making of which the translator is aware – most translators do not have a conscious strategy for dealing with optional *that*, for example. It can be argued that it is the nature of the process of translation and the cognitive processing which it requires which produces the kind of patterning seen here. However, inclusion or omission of syntactic features do not reveal the whole story. Olohan and Baker (2000) pointed out that the optional *that* data discussed in that study revealed potentially different patterns in other features, such as use of modifiers, pronominal forms, modal constructions etc. in TEC compared with the BNC. Thus, although a specific syntactic or lexical structure can be investigated in terms of overall occurrence and of its usage within the narrow context of a concordance line, the wider issue of co-occurrence and interdependency of features must be considered. Research of this kind on the language of translation still has a long way to go; however a small example can be used to illustrate the possible significance of interdependencies and how they might be investigated further. We can take the data referred to earlier in relation to *promise* and re-examine it in relation to a number of linguists' suggestions that *that* is more likely to be omitted in informal usage (for example Storms 1966; Elsness 1984; Dixon 1991). If we also accept that the use of contracted forms constitutes evidence of informal style, then a search for contracted forms, within the *promise* concordance line only, reveals the following (Table 11):

Form	BNC	TEC
promise total	57	48
promise with contracted forms in concordance line	41 (72%)	21 (43.75%)
promise + that with contracted forms in concordance line	7 (17%)	4 (19%)
promise + zero with contracted forms in concordance line	34 (83%)	17 (81%)
promise with no contracted forms in concordance line	16 (28%)	27 (56.25%)
promise + that with no contracted forms in concordance line	12 (75%)	23 (85%)
promise + zero with no contracted forms in concordance line	4 (25%)	4 (15%)

Table 11: Co-occurrence of *promise +that/zero* and contracted forms in BNC and TEC

From this we can see that, although *that* occurs with much higher frequency in TEC than in BNC, promise co-occurs with contracted forms to a much higher degree in BNC than in TEC, and that, when the *that/zero* usage is correlated with contracted forms and then compared across corpora, there is actually little difference between the two corpora. Using contracted forms as a measure of informality, this would indicate, firstly, that there is a correlation between inclusion of *that* and level of formality, and, secondly, that the language of TEC may thus be judged more formal. A large-scale study of contracted forms based on production and pruning of word lists for both corpora yielded the following Table 12 and Figure 3):

Form	BNC forms	BNC totals	TEC forms	TEC totals
apostrophe	5,851		5,269	
*'s	4,818		4,623	
*'ll	212	9,651	43	4,799
*'d	111	10,645	29	5,349
*'t	48	40,782	30	20,316
*'ve	53	7,768	17	4,068
*'re	12	7,344	8	4,250
it's		9,554		5,046
that's		4,650		2,640
there's		2,655		1,424
he's		2,628		1,951
she's		2,266		1,154
what's		1,601		1,021
let's		913		654
who's		396		334
where's		241		117
how's		146		36
here's		132		89
e's		102		0
I'm		8,773		4,256
d' = do	3	418	3	84
t' = the	99	126	0	0
y' = you	22	53	7	7

Table 12: Contracted forms in BNC and TEC

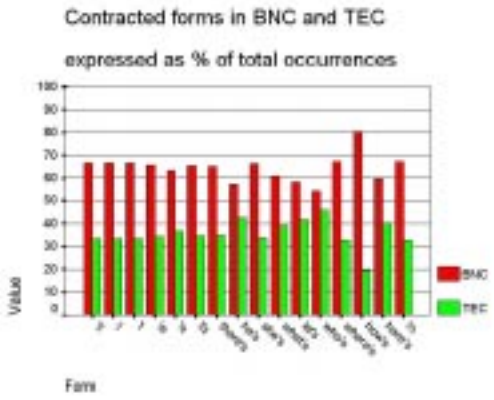

Figure 3: Contracted forms in BNC and TEC as percentage of total occurrence across corpora

The most frequent form with apostrophe is *'*s*, which in the vast majority of cases is a possessive marker rather than a contraction of *is* or *was*; many of the *'*s* occurrences are with names, and many occur only once or a couple of times in the corpus. For this reason, individual occurrences of *'*s* have not been counted, apart from the most common *'*s* contractions in BNC (*it's, that's, there's, he's, she's, what's let's, who's, where's, how's, here's, e's*). Without looking at data for individual occurrences for *'*s* forms, we can see from the figures above that the total number of *'*s* forms is similar for both corpora. This is in stark contrast with all other categories, which represent true contractions rather than grammatical markers. For all other contracted forms counted, a very clear and consistent pattern emerges; they are much more frequently used in BNC than in TEC.

As mentioned above, one of the conclusions of the linguistics literature in relation to the optional *that* is that omission is more likely in informal usage. This may also be the case for omission of the relative pronoun *that* or *which* and the other optional features discussed above. The only exception is perhaps the modal *should* following verbs such as *suggest* and *order*; if the modal is omitted, the subjunctive is used, which arguably constitutes more formal style than the *should* construction. Interestingly this is the only feature above for which TEC seems to favour omission rather than inclusion. On all other optional forms, TEC is considerably more likely to use the optional item and longer surface form.

According to the co-occurrence patterns which Biber (1988) and Biber et al. (1998) suggest as underlying the five major dimensions of English, *that*-deletion and contractions are in the top three

features at the positive end of one scale (Dimension 1); this is indicative of their tendency to co-occur in texts of shared function. These and the other features grouped with them are associated with 'involved, non-informational focus, related to a primarily interactive or affective purpose and on-line production circumstances' (Biber et al. 1998: 149). Biber et al. continue to describe certain of these positive features, including the two we have dealt with here – *that*-deletions and contractions – as constituting a reduced surface form which results in a 'more generalized, less explicit content' (ibid.). They talk of two separate communicative parameters, i.e. purpose of the writer (informational vs. involved) and production circumstances (allowing careful editing vs. constraints of real-time production). Dimension 1 is therefore labelled 'involved versus informational production' (ibid.). Relating this to the findings above, it would appear than the BNC writing is more involved, more generalised, less explicit, less edited than the writing in TEC; the original writer's purpose is more involved, the translator's less so. The translator's surface form is not reduced to the same extent as the original writer's, the translator is thus more explicit, less generalised in both form and content. The translation is perhaps more carefully edited; are original writers more concerned with the creative content and translators with explicitation of linguistic relations?

6. Conclusion

In terms of concrete findings of this kind in corpus-based translation studies, there is considerable scope for further studies, particularly in the area of co-occurrence. Mauranen's (2000) study of research on comparison of co-selectional restrictions in Finnish translation and original English is one of the few which tackles collocational and colligational patterning in translated language using comparable corpora, and much more research of this nature needs to be done. In addition, the other co-occurrence features proposed by Biber for this and the other four dimensions of English could be investigated and compared across the corpora. Ongoing work in Saarbrücken using a tagged version of TEC is likely to yield interesting results in this respect, and research is continuing at UMIST to identify and investigate relationships between linguistic features of translated language and the cognitive and social factors which may give rise to them.

References

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. In Somers H (ed) *Terminology, LSP and translation: studies in language engineering, in honour of Juan C. Sager,* Amsterdam and Philadelphia, John Benjamins, pp 175-186.

Baker M 1999 The role of corpora in investigating the linguistic behaviour of translators. *International Journal of Corpus Linguistics* 4(2): 281-298.

Biber D 1988 *Variation across speech and writing*. Cambridge, CUP.

Biber D, Conrad S and Reppen R 1998 *Corpus linguistics: investigating language structure and use*. Cambridge, CUP.

Blum-Kulka S 1986 Shifts of cohesion and coherence in translation. In House J and Blum-Kulka S (eds) *Interlingual and intercultural communication: discourse and cognition in translation and second language acquisition studies*. Tübingen, Gunter Narr. pp 17-35.

Burnett S 1999 *A corpus-based study of translational English*, Unpublished MSc dissertation, UMIST.

Dixon R M W 1991 *A new approach to English grammar, on semantic principles*. Oxford, Clarendon Press.

Elsness J 1984 That or Zero? A Look at the Choice of Object Clause Connective in a Corpus of American English. *English Studies* 65: 519-533.

Kenny D 1999 *Norms and creativity: lexis in translated text.* Unpublished PhD thesis, UMIST.

Kenny D 2000 Lexical hide-and-seek: looking for creativity in a parallel corpus. In Olohan M (ed) *Intercultural faultlines: research models in translation studies 1*, Manchester, St. Jerome. pp 93-104.

Laviosa-Braithwaite S 1996 *The English Comparable Corpus (ECC): a resource and a methodology for the empirical study of translation*. Unpublished PhD thesis, UMIST.

Laviosa S 1998 The English Comparable Corpus: a resource and a methodology. In Bowker L, Cronin M, Kenny D. & Pearson J. (eds) *Unity in diversity: current trends in translation studies.* Manchester, St. Jerome Publishing. pp 101-112.

Olohan M, Baker M 2000 Reporting *that* in translated English: evidence for subconscious processes of explicitation? *Across Languages and Cultures* 1(2): 141-158.

Mauranen A 2000 Strange strings in translated language: a study on corpora. In Olohan M (ed) *Intercultural faultlines: research models in translation studies 1*, Manchester, St. Jerome. pp. 119-141ю

Storms G 1966 *That*-clauses in Modern English. *English Studies* 47: 249-270.

Vanderauwera R 1985 *Dutch Novels Translated into English: The Transformation of a 'Minority' Literature*. Amsterdam, Rodopi.