Meta

Journal des traducteurs
Translators' Journal

MƐTA

Normalisation and the Translation of Phraseology in the COVALT Corpus

Josep Marco

Citer cet article

Marco, J. (2009). Normalisation and the Translation of Phraseology in the COVALT Corpus. *Meta*, *54*(4), 842–856. https://doi.org/10.7202/038907ar

Résumé de l'article

Le présent article traite de l'hypothèse selon laquelle l'emploi d'unités phraséologiques dans les textes traduits peut être considéré comme un indicateur d'une tendance vers la normalisation. En effet, les unités phraséologiques sont des formes conventionnelles de la langue cible appartenant au répertoire lexical de cette dernière. Des données puisées dans le sous-corpus anglais-catalan de COVALT (corpus valencien de textes littéraires traduits) indiquent que les textes traduits en catalan sont moins *phraséologiques* que les textes sources anglais. Toutefois, cette différence est faible, ce qui semble témoigner d'un effort, de la part des traducteurs, pour préserver ou recréer une phraséologie significative dans les textes cibles. Cependant, il faudra mener d'autres études pour identifier les motivations sous-jacentes à cette pratique.

érudit

TERMINOLOGIE ET LINGUISTIQUE

Normalisation and the Translation of Phraseology in the COVALT Corpus[1]

JOSEP MARCO
Universitat Jaume I, Castelló, Spain
jmarco@trad.uji.es

RÉSUMÉ

Le présent article traite de l'hypothèse selon laquelle l'emploi d'unités phraséologiques dans les textes traduits peut être considéré comme un indicateur d'une tendance vers la normalisation. En effet, les unités phraséologiques sont des formes conventionnelles de la langue cible appartenant au répertoire lexical de cette dernière. Des données puisées dans le sous-corpus anglais-catalan de COVALT (corpus valencien de textes littéraires traduits) indiquent que les textes traduits en catalan sont moins *phraséologiques* que les textes sources anglais. Toutefois, cette différence est faible, ce qui semble témoigner d'un effort, de la part des traducteurs, pour préserver ou recréer une phraséologie significative dans les textes cibles. Cependant, il faudra mener d'autres études pour identifier les motivations sous-jacentes à cette pratique.

ABSTRACT

In this article, it is assumed that phraseological usage can be regarded as an indicator of normalisation in translated texts, as phraseological units are target-language standardised forms belonging to its lexical repertoire. Drawing on data yielded by the English-Catalan subcorpus of COVALT (Valencian Corpus of Translated Literature), it was found that Catalan translated texts are less *phraseological* than their corresponding English source texts, though only by a narrow margin. The narrowness of the margin seems to bear witness to some effort on the translators' part to retain or recreate a noticeable degree of phraseological activity in translated texts. However, further research is needed into the motives underlying translator behaviour in this respect.

MOTS-CLÉS/KEYWORDS

phraséologie, normalisation, corpus COVALT, difficultés de traduction, catalan
phraseology, normalisation, COVALT corpus, translation difficulties, Catalan

1. Introduction

Corpus-based Translation Studies (CBTS), as a distinct approach, has been around for over a decade now. It was launched in the mid-1990s by a series of seminal articles by Mona Baker (1993, 1995, 1996) which in many respects have guided its course up to the present. However, the initial seed has germinated into a variety of interests, or research lines, not at all incompatible with one another but with quite distinctive flavours.

1) What might be referred to as the canonical line is the one initiated by Baker herself, which focuses on the main features of translated language – vis-à-vis non-translated language. It is strongly indebted to Descriptive Translation Studies (Kenny 2001: 48; Olohan 2004: 17). Source texts do not come into the picture at all, research is typically based on comparable corpora and what scholars ultimately search for is translation universals.

2) But there are alternative lines. Bernardini (2005), for instance, argues that corpus-based translation research has been biased in favour of comparable corpora and the balance needs to be redressed. That kind of research has thrown light on a number of interesting aspects of translation behaviour – and hopefully will continue to do so – but "it is the very nature of translation as a mediated communicative event (Baker 1993) that makes an exclusively target-oriented approach to translation analysis methodologically questionable" (Bernardini 2005: 6). It is argued that parallel and reference corpora need to be used to complement the data yielded by comparable corpora (as in Teich 2003).

Different growths are well documented in Laviosa (2002) and Olohan (2004). In fact, the existence of such handbooks – it might be argued – shows that CBTS is well established as a discernible approach within our discipline.

However, not enough attention has been paid to the fact that corpora and corpus analysis tools represent a qualitative leap as far as research methods are concerned. True, this has been repeatedly remarked (see Laviosa 2002: 27, for instance), but it is a point that can hardly be overstated. Translation Studies research – just like research in many other language-centred disciplines – used to be anecdotic until very recently, and remains so in many cases. The reason for this is because the amount of data an individual scholar, or even a research group, was able to handle was very limited and, as a result, they (i.e., *we*) felt obliged to end many of their contributions on an apologetic note, along these familiar lines: *our conclusions are such and such, but further research should be carried out in order for them to be generalisable.* This difficulty is now partly overcome, since results from such large amounts of data as corpus-based translation scholars can often handle are more generally valid. In fact, the amount of data that can be analysed by electronic means is virtually limitless. That does not mean that the output of such research is *the truth*, in any philosophical sense, but it is certainly less (fatally) limited than the output of manual analysis. The kind of analysis performed by the computer is not comparable to human analysis, in terms of subtlety; but even so, if selectively applied, automated or semi-automated analysis can throw light on new areas of research by virtue of its sheer bulk.

All this can be illustrated by reference to the pervasive phenomenon of phraseology, which, under such various terminological guises as idioms, fixed expressions, clichés, etc., has attracted translation scholars' attention for several decades now. Vinay and Darbelnet (1958), for instance, illustrate the technical procedure they call *equivalence* by reference to the translation of phraseological units. Also, phraseological units – including collocations – are part and parcel of such textually oriented translation works as Baker (1992) and Neubert and Shreve (1992). More recently, they have been presented (Molina Martínez 2006) as items at the interface between language and culture. Only in the Spanish context several monographs (e.g., Corpas 2003; Van Lawick 2006) have broached the subject of the translation of phraseology. All this bears witness to the interest aroused by it; but the studies mentioned are seldom empirical, and when they are, they move within the narrow limits of manual analysis.

This article focuses on the study of a number of phraseological units extracted from the English-Catalan section of COVALT (Valencian Corpus of Translated Literature), a multilingual corpus – still under construction – made up of the translations into Catalan of narrative works originally written in English, French, and German published in the autonomous region of Valencia from 1990 to 2000. The English-Catalan sub-corpus currently includes 23 pairs of source text + target text which amount to 1,161,359 words (571,909 English, 589,450 Catalan). Corpus analysis is carried out by means of AlfraCOVALT, a bilingual concordancing programme developed within the COVALT research group by Josep Guzman (Guzman and Serrano 2006; Guzman 2007).

2. Phraseology and the normalisation hypothesis

The main research question addressed in this article could be formulated as follows: is there any discernible relationship between the translation of phraseology in the COVALT corpus and any of the so-called *translation universals* on which the bulk of corpus-based translation research is focused? An initial analysis of data provided by COVALT pointed to interference as the most likely candidate, since about 50% of the phraseological units identified had been translated as what might be regarded as similar phraseological units, i.e., segments that were phraseological in character in the target language and did not depart significantly from the source text segments. Be it noted that Baker and other scholars who base their research on comparable corpora do not even mention interference as a translation universal, as source texts are not taken into account. But interference was already put forward by Toury even before electronic corpora became mainstream as a research tool, and has since been taken up by other scholars such as Mauranen (2004, 2005), Eskola (2004) or Pym (2007). Going back to my point, the COVALT data I have just referred to was flawed in a fundamental way: it was unidirectional, in that it only included source text phraseological units and their corresponding translated segments, without ever proceeding the other way around, i.e., starting from the target texts. And it might well be hypothesised that translated texts can contain many phraseological units not taking their cue from source text phraseological material. If that were the case, the balance might be tipped in favour of normalisation, or the (alleged) tendency of translated text to closely adhere to target language conventions. In fact, conventionality stands out as a common feature of both phraseology and the tendency to normalise.

There's no space here even for a brief overview of the field of phraseology, which would scarcely do justice to the multiplicity of approaches to its object of study. Moreover, the study of phraseology, after years of being a minor concern for linguists, has become the aim of a linguistic discipline in its own right and arouses interest all over the world. A translation scholar who wishes to examine how phraseological units fare in a particular parallel corpus does not need to delve too deeply into theoretical details and pitfalls, but must take a general survey of the field because they will be forced to make decisions concerning practical matters.

Firstly, our scholar will have to determine the scope of their study. Leaving aside the fact that it is not always clear what exactly qualifies as a phraseological unit, as witnessed by the variety of definitions we come across in the literature (see Van

Lawick 2006: 43-46), there are different kinds of phrasemes and the analyst will have to decide whether they include them all in their study or narrow their scope to just some of them. Again, there are almost as many attempts at typology as authors. In one of them, which enjoys some currency in Spain, Corpas (1997) provides a threefold classification into utterances (i.e., phraseological units which constitute full utterances or sentences), idioms (which typically operate at phrase level) and collocations. For the purposes of this particular study, and also in the more general work carried out within the COVALT group, we decided to concentrate on the first two types (utterances and idioms) and leave out collocations, for two reasons. First, collocations show a lower degree of fixedness than other phraseological units and, as a result, a higher frequency of occurrence, which would make the number of concordance lines to be dealt with virtually unmanageable. And second, collocations are generally regarded (Van Lawick 2006: 71-72) as less prototypical representatives of the *phraseological unit* class than the other types. Since the division between collocations and other kinds of phrasemes is not clear-cut either, it has been necessary to resort to several dictionaries, repertoires and even general corpora, both in English and Catalan.

Secondly, in order to analyse the parallel concordances and especially the relationship between the *replacing and replaced segments* (to put it in Toury's terms), the scholar will need a list of techniques or procedures (I would rather not use the term *shift*, as it only covers dissimilarities between the source and the target). Within that list, they may wish to distinguish between phraseological units that can be regarded as similar, perhaps even identical, to those encountered in source texts, from phraseological units that are clearly different. But that distinction will prove problematical, as the dichotomy sameness / difference is largely a matter of degree and, again, constitutes a cline rather than a clear-cut division. So criteria will be needed to account for classifications.

On the other hand, normalisation has featured among so-called universals since the beginning, so to speak, i.e., since Baker (1993, 1996) drew her list of likely candidates to that consideration. However, it had already been identified as a trait of translated text – not universally but in specific environments – even before the advent of CBTS. Thus, Toury's law of growing standardisation is but another way to designate the same phenomenon. According to Toury (1995: 267-268), what such a law implies is that "in translation, source-text textemes tend to be converted into target-language (or target-culture) repertoremes." Textemes are textual elements and relations which characterise the source text and often give it its peculiar flavour; repertoremes, for their part, are elements and relations belonging to the habitual repertoire of the target language or culture; and the alleged replacement of ST textemes with TL repertoremes is a way of saying that translated texts tend to be more conventional than originals, to leave aside (often because it cannot be helped: see, for instance, Parks 1998: 12-13) much of what is distinctive of those originals, either through translator's lack of awareness or because the distinctive features in question will not travel well.

In her empirical study of 50 Dutch novels and their English translations, Vanderauwera (1985: 93) finds evidence of a "tendency towards textual conventionality" in translated texts at all language levels. This tendency she ascribes to translations being "target-accommodating," i.e., to a deliberate wish (on the part of those agents

involved: translators, publishers, editors) to make translations conform to target-reader expectations when reading fiction translated from a smaller culture.

In the present era of electronic corpora, normalisation remains one of the translation features under scrutiny. A significant study in this respect is Kenny's (2001), who set out to examine some items involving creativity and how they fared in translation. To that end, she built a corpus of German fictional texts and their English translations. She focused on three kinds of phenomena: creative *hapax legomena* (or word forms occurring only once in the corpus), forms peculiar to a particular author and unusual collocations. Kenny's findings are contradictory, insofar as normalising techniques rank high in the first phenomenon but not so high in the other two. It is concluded that translators do often normalise, but they also deploy their skills in a creative way.

Other studies, rather than being ambiguous, go against the assumptions underlying the normalisation hypothesis, as is the case of Mauranen (2000). Indeed she later claims that "[t]he results on patterns of lexical combination, mainly collocations, seem to point towards untypical combinatory tendencies in translations" (Mauranen 2005: 79). Her study focuses on a particular lexical item in Finnish (haluta) and its combinations, and Mauranen's conclusion is that not only are lexical frequencies untypical in translations, but the collocational range of the lexical item under scrutiny is wider in translated than in original Finnish, and its collocational patterns are also divergent. In an attempt to account for such divergence, the author claims that "it seems that translators utilize the resources of the target language by making relatively more use of what you *can do* than what you *typically* do" (Mauranen 2005: 80).

However, Bernardini (2007), in a very recent study, reaches a somewhat different conclusion. Drawing on data provided by a small bi-directional corpus made up of "extracts from novels and short stories in original and translated English (source language: Italian)" and "similar extracts in original/translated Italian (source language: English)" (Bernardini 2007: 4), she first identifies word combination *types* in that corpus and then obtains *their frequency and relatedness* in two large reference corpora – the *British National Corpus* for English, and the *Repubblica* corpus (340 million words from the eponymous newspaper) for Italian. Bernardini's aim is two-fold: a) to determine whether translated texts are more *collocational* than original texts in the same language, and b) to elucidate whether any differences found can be attributed to the translation process. On the basis of the monolingual comparison, the author hypothesises for one of the patterns chosen (*N prep|conj N*, i.e., a two-noun collocation with either a preposition or a conjunction in between) that "Italian translators tend to make use of *N prep|conj N* established sequences (potential collocations) more than Italian authors do" (Bernardini 2007: 8). In order to ascertain that this finding can indeed be attributed to the translation process, 1,061 parallel concordances are browsed and 127 shifts "leading to increased institutionalisation" identified. It is worth noting that in the analysis of a small sample of those shifts the term *normalisation* occurs more than once. The author concludes by saying that "we can tentatively suggest that translated texts would seem to be more *collocational* than original texts in the same language, and that there is some evidence that this is a consequence of the translation process," even though "[i]t is difficult to tell what was the principal driving force behind these shifts [i.e., the shifts "leading to increased

institutionalisation" referred to above]" (Bernardini 2007: 14). The issue of explanation will be taken up again towards the end of this article.

Findings, then, are contradictory so far, as it is not at all clear that translated text tends to normalise in all cases. It is precisely at this juncture that the present study may prove relevant, since it can lend weight (or otherwise) to the normalisation hypothesis. In this respect, it must be emphasised that the present study strongly relies on the assumption that descriptive research is, at its best, cumulative in nature, in the sense that a considerable amount of data from different origins must be analysed before generalisations can be held as valid, or as valid in specific environments. In other words, the only valid generalisations are those which draw on a wealth of case studies.

3. Methodology

The kind of work presented here is framed in a project aiming to study phraseology in the COVALT corpus on a larger scale. It must therefore be considered as work in progress, as a sort of pilot study. Since one of the main problems encountered in empirical work of this kind is that of scope, it has been necessary – for practical reasons – to restrict the scope of the present work in the following ways:

1) it focuses on the English-Catalan sub-corpus of COVALT; the French- and German-Catalan sub-corpora are thus not considered. But even the English-Catalan sub-corpus is made up of more texts than are here taken into account. Some of them had to be provisionally left out as the bilingual concordancer did not work smoothly on them and they need re-editing. Table 1 provides a list of the texts on which the present study is based;

2) as mentioned above, of the three kinds of phraseological units identified by Corpas (1997) only two have been included in the study, collocations being left aside (again, for practical reasons);

3) one of the crucial decisions to be made is that of which search words, or nodes, are selected to be entered as queries. In the larger project, as far as the English-Catalan sub-corpus is concerned, the selected search words were the following lemmas with their morphological variants (as the corpus is not lemmatised): *arm*, *body*, *ear*, *eye*, *face*, *foot*, *hair*, *hand*, *mouth*, *neck*, *nose*, *tongue*, *blood*, together with their Catalan equivalents. In the present work, however, the number of search words has been restricted to 3, although they are the most productive three: *eye*, *foot*, *hand*. Their Catalan equivalents are respectively **ull**, **peu** and **mà**.

TABLE 1

Texts from the English-Catalan sub-corpus of COVALT included in this study

Title (Author)	Number of words in the ST	Number of words in the TT
The Wonderful Wizard of Oz (Frank L. Baum)	39.504	38.253
The Secret Garden and Other Stories (Gilbert K. Chesterton)	21.427	24.022
Typhoon (Joseph Conrad)	30.001	29.714
The Grizzly King (James Oliver Curwood)	45.744	43.722
The Adventure of the Bruce-Partington Plans (Arthur Conan Doyle)	10.722	10.653
The Adventure of the Norwood Builder (Arthur Conan Doyle)	9.278	9.160
The Keeper (Barry Faville)	46.722	49.166

The Ghostly Rental (Henry James)	13.414	12.534
The Dead (and other stories from *Dubliners*) (James Joyce)	26.932	28.087
Worlds of Exile and Illusion (Ursula K. Le Guin)	42.504	46.453
The Cruise of the Dazzler (Jack London)	35.556	36.756
The Dunwich Horror (Howard P. Lovecraft)	17.405	18.592
Billy Budd, Sailor (Herman Melville)	30.370	33.963
Bartleby, the Scrivener (Herman Melville)	14.485	14.714
The Gold Bug (Edgar Allan Poe)	13.681	12.409
The Mysteries of Paris (Edgar Allan Poe)	40.486	41.694
The Pit and the Pendulum (Edgar Allan Poe)	6.227	7.409
Tobermory (Saki)	22.075	22.851
The Suicide Club (Robert Louis Stevenson)	27.601	28.098
The Strange Case of Dr Jekyll and Mr Hyde (Robert Louis Stevenson)	25.670	25.911
The Bottle Imp (Robert Louis Stevenson)	12.256	12.752
The Man that Corrupted Hadleyburg (Mark Twain)	18.167	18.503
Vancouver Nightmare (Eric Wilson)	21.682	24.034
TOTAL	**571.909**	**589.450**

The query matches yielded by AlfraCOVALT are not yet the data to be analysed. They are just raw material insofar as they include both phraseological and non-phraseological occurrences. The analyst must then proceed manually to tell the former apart from the latter, a task which, as already pointed out above, is far from straightforward, on account of frequent borderline cases. The phraseological units thus identified are copied onto an Excel file – as the Excel filter utility facilitates groupings within a field – and assigned a technique label; i.e., the relationship between ST and TT segments is analysed and described in terms of the technique used.

As to the classification of techniques employed in the translation of phraseological units, our point of departure was Delabastita's (1996) list of techniques used in the translation of wordplay. The adequacy of such a list was an interpretive hypothesis, in Chesterman's sense (2004: 2), but it was only partly borne out by its confrontation with empirical data. Once modified and refined, the list of techniques for our present purpose looks as follows (PU stands for *phraseological unit*):

1) PU → Similar PU: the translated segment is a target-language phraseological unit and is similar in both overall meaning and metaphorical base to the ST phraseological unit;
2) PU → Different PU: the translated segment is a target-language phraseological unit but it is different from the ST phraseological unit in either overall meaning or metaphorical base, or in both, or in some other relevant respect;
3) PU → Collocation: the translated segment is not an idiom or a phraseological utterance in the target language, but a collocation;
4) PU → No PU: the translated segment is not phraseological;
5) Omission: the ST segment including the phraseological unit has been omitted in the translation;
6) Direct copy: the ST segment has been translated more or less literally, but the result is not a phraseological unit in the target text. It is a calquing technique which gives rise to a certain degree of incoherence in the translation;
7) Collocation → PU: a ST collocation is translated as a phraseological unit in the TT;
8) No PU → PU: a non-phraseological segment in the ST is translated as a phraseological unit in the TT.

As might be expected, all sorts of difficulties arise in the process of technique classification, on account of borderline cases, but the main ones could be summarised as follows:

First, it is not always straightforward whether a given segment is a phraseological unit (of the kinds dealt with here) or a collocation. Collocation is often a matter of frequent co-occurrence, whereas phraseology, as a category, relies on a cluster of features which may (or may not) be actualised, in a concrete phraseological unit, to a greater or lesser extent. Some of these features are, for instance, figurativeness and non-compositionality (the meaning of the whole is more than the sum of its parts, i.e., than the sum of the meanings of the individual words), and they may prove helpful as distinguishing criteria. *To open one's eyes*, for instance, is regarded as a collocation when it is meant literally, but treated as an idiom when it is endowed with a figurative meaning, as in *He opened my eyes to this new reality*. But it is not always *that* simple.

Second, as pointed out a few paragraphs above, it is sometimes difficult to say whether a given phraseological unit in the TT is similar or different to its corresponding ST segment, as it may be similar in some respects and different in others. And even within one single aspect, sameness / difference may be a matter of degree. For practical purposes, we decided to rule out the possibility of two phraseological units being *identical* and restrict ourselves to the twofold distinction *similar / different*. In order for two phraseological units to be similar, they are required to show relevant similarity in both overall meaning and metaphorical base. If the overall meaning is different, then the two units are obviously different, although this does not occur frequently; and if the metaphorical base is perceived as different in any relevant respect, then the two units are regarded as different. But even with these criteria in mind it is sometimes difficult to make up one's mind.

4. Findings

Once the searching and selection process is completed, and each pair of ST and TT segments is classified according to the translation technique employed, we are faced with the results shown in tables 2 and 3. Table 2 shows how our 483 occurrences, or parallel concordance lines, are distributed across a) the different fiction texts in which they occur, and b) the translation techniques employed, in raw frequencies. Table 3 shows the same results in percentage form.

TABLE 2

Distribution of occurrences across fiction texts and translation techniques (raw frequencies)

	Omission	Direct copy	PU → No PU	PU → Collocation	PU → Similar PU	PU → Different PU	Collocation → PU	No PU → PU	Total
Bartleby			1		5	3		2	11
Billy Budd			3	1	2			6	12
Bottle Imp			5		8	5		2	20
Bruce-Partington				2	6	1		1	10
Dazzler			14	4	12	8	4	3	45
Dead			4		8	1	2	5	20
Dunwich Horror			3		3			6	12

	Omission	Direct copy	PU → No PU	PU → Collocation	PU → Similar PU	PU → Different PU	Collocation → PU	No PU → PU	Total
Ghostly Rental			2		4	1	1	1	9
Gold Bug			1		2		1	1	5
Grizzly King	2		8	8	16	15		4	53
Hadleyburg			1	1	10	2		1	15
Jekyll	2		9	1	14	9		3	38
Keeper			6	2	12	6		4	30
Mysteries Paris	1		10		8	7	2		28
Norwood Builder			3		6	1		1	11
Pit and Pendulum			1		4		1	1	7
Rocannon			3	1	12	2	2	19	39
Secret Garden					6	2	1	7	16
Suicide Club			3	1	16	8	1	1	30
Tobermory			1		5	1		3	10
Typhoon			5	3	13	6		7	34
Vancouver			1	1	4	5		2	13
Wizard Oz			1		6	1		7	15
Total	**5**	**0**	**85**	**24**	**181**	**86**	**15**	**87**	**483**

TABLE 3

Distribution of occurrences across fiction texts and translation techniques (percentages)

	Omission	Direct copy	PU → No PU	PU → Collocation	PU → Similar PU	PU → Different PU	Collocation → PU	No PU → PU	Total
Bartleby	0.00	0.00	9.09	0.00	45.45	27.27	0.00	18.18	100.00
Billy Budd	0.00	0.00	25.00	0.00	8.33	16.67	0.00	50.00	100.00
Bottle Imp	0.00	0.00	25.00	0.00	40.00	25.00	0.00	10.00	100.00
Bruce-Partington	0.00	0.00	0.00	20.00	60.00	10.00	0.00	10.00	100.00
Dazzler	0.00	0.00	31.11	8.89	26.67	17.78	8.89	6.67	100.00
Dead	0.00	0.00	20.00	0.00	40.00	5.00	10.00	25.00	100.00
Dunwich Horror	0.00	0.00	25.00	0.00	25.00	0.00	0.00	50.00	100.00
Ghostly Rental	0.00	0.00	22.22	0.00	44.44	11.11	11.11	11.11	100.00
Gold Bug	0.00	0.00	20.00	0.00	40.00	0.00	20.00	20.00	100.00
Grizzly King	3.77	0.00	15.09	15.09	30.19	28.30	0.00	7.55	100.00
Hadleyburg	0.00	0.00	6.67	6.67	66.67	13.33	0.00	6.67	100.00
Jekyll	5.26	0.00	23.68	2.63	36.84	23.68	0.00	7.89	100.00
Keeper	0.00	0.00	20.00	6.67	40.00	20.00	0.00	13.33	100.00
Mysteries Paris	3.57	0.00	35.71	0.00	28.57	25.00	7.14	0.00	100.00
Norwood Builder	0.00	0.00	27.27	0.00	54.55	9.09	0.00	9.09	100.00
Pit and Pendulum	0.00	0.00	14.29	0.00	57.14	0.00	14.29	14.29	100.00
Rocannon	0.00	0.00	7.69	2.56	30.77	5.13	5.13	48.72	100.00
Secret Garden	0.00	0.00	0.00	0.00	37.50	12.50	6.25	43.75	100.00
Suicide Club	0.00	0.00	10.00	3.33	53.33	26.67	3.33	3.33	100.00
Tobermory	0.00	0.00	10.00	0.00	50.00	10.00	0.00	30.00	100.00
Typhoon	0.00	0.00	14.71	8.82	38.24	17.65	0.00	20.59	100.00
Vancouver	0.00	0.00	7.69	7.69	30.77	38.46	0.00	15.38	100.00
Wizard Oz	0.00	0.00	6.67	0.00	40.00	6.67	0.00	46.67	100.00
Total	**1.04**	**0.00**	**17.60**	**4.97**	**37.47**	**17.81**	**3.11**	**18.01**	**100.00**

What these tables show is that there is a strong tendency in the corpus to translate a ST phraseological unit by means of a target-language phraseological unit,

whether it be (relatively) similar to or different from the one encountered in the source. These two techniques together account for over half (55.28%) of the concordance lines under scrutiny. Then, on either side of these two central techniques are those translation solutions which imply either *phraseological loss* (to the left) or *phraseological gain* (to the right). Leaving aside omission, which is almost negligible, and direct copy, not represented in these results, the techniques PU → No PU and PU → Collocation can be taken to represent non-phraseological solutions to ST phraseological units, and together they account for 22.57% of the cases. Parallel to this, the techniques Collocation → PU and No PU → PU stand for non-phraseological segments translated as phraseological units, and together they account for 21.12% of the cases. If, as hypothesised above, phraseological usage in translated text is regarded as an indicator of normalisation, the results just shown do not lend support to the normalisation hypothesis in the COVALT corpus. The 21.12% of cases in which there is phraseological gain fail to match the 22.57% involving loss – though by a narrow margin. As far as figures are concerned, that is how matters stand.

5. Discussion

However, this initial conclusion based on figures needs qualifying – it needs to be put in perspective. Phraseological units are often – though by no means always – difficult to translate. Even if a translator is familiar with a phraseme they come across in a text, they need to check whether its target-language surface equivalent can function as an equivalent in context, whether its register, cultural or connotative values are (roughly) the same, etc. Mismatches across phraseological systems are frequent and false friendship is not a rare phenomenon. It is often the case that the target language does *not* have a surface equivalent at all, in which case the translator can only look for a different phraseological unit which fits the context in some respect, or paraphrase the meaning of the ST unit by means of non-phraseological material. When all these difficulties are taken into account (and other, more subjective ones could be added, concerning the translator's motivation, the amount of time and documentation tools at their disposal, etc.), it seems fair to say that an even higher degree of loss would not have been surprising. And, as we have just seen, the margin is narrow. Therefore, on the basis of the data analysed, it could be tentatively argued that contemporary translators of fiction into Catalan in the Valencian region regard phraseological usage in translated texts as something positive and therefore labour to preserve and even create an acceptable degree of phraseological activity. That may not be the case in all genres, but it sounds especially plausible for literary texts. Regardless of other possible textual and stylistic contributions they can make, phraseological units tend to add a layer of expressivity (e.g., González-Rey 1998, Ruiz Gurillo 1998) which is usually highly valued in literary discourse. And it may not be the case either in all sociocultural contexts, but it seems to be true of the texts comprised by the English-Catalan subcorpus of COVALT. As our data shows, there is almost as much phraseological activity in the target texts as in the source texts, and that can hardly be said to come about naturally, i.e., effortlessly.

Let us examine some examples to illustrate this. It is true that the corpus offers many instances of phraseological loss (PU → No PU) which could have been easily avoided, such as the following:

Table 4

Example of phraseological loss (I)

Node	ST	ST segment	TT segment	Technique
Hand	Bruce-Partington	Give me a *hand*, Watson, and I'll do the same for you	Ajude'm, Watson, i faré el mateix per vosté	PU → No PU

The ST phraseological unit has readily available equivalents in the TL (**tirar una mà**, **donar un colp de mà**), but they are not used and the flatter, non-phraseological **Ajude'm** (*Help me*) is preferred instead. However, there are many other instances of phraseological loss for which it would be difficult at any time to find a phraseological solution. That is the case in the English idiom *in (one's) mind's eye / to see (something) with/in/before (one's) mind's eye*, as in the following example:

Table 5

Example of phraseological loss (II)

Node	ST	ST segment	TT segment	Technique
Eye	Jekyll	For once more he saw before his mind's *eye*, as clear as transparency, the strange clauses of the will	I dubtava perquè, una vegada més, havia recordat, tan clares com la mateixa transparència, les estranyes clàusules del testament	PU → No PU

He saw before his mind's eye is translated here as **havia recordat** (*had remembered*). In fact, this idiom occurs five times in the corpus and it is always rendered non-phraseologically. The most plausible reason for this is that the Catalan system does not possess an equivalent idiom; in other respects it may be said to be quite transparent and therefore easy to understand and translate. This also happens in biblical allusions which have been idiomatised in English but not in Catalan.

On the other hand, there are instances of PU → Different PU and, above all, of No PU → PU which strike one as remarkable, as they seem to point in the direction suggested above, i.e., a certain degree of effort on the translator's part to maintain an acceptable level of phraseological activity in the TT. In PU → Different PU, the translator is forced to search for a phraseological unit in their system which matches the ST unit in overall meaning, even if the metaphorical base is different. But it is in No PU → PU where active effort is more visible, as this technique consists in rendering phraseologically what was not phraseological in the original. Let us look at a couple of examples.

Table 6

Examples of phraseological gain

Node	ST	ST segment	TT segment	Technique
Mà(ns)	Secret Garden	Under a tree in this commanding yet neglected spot was an old ramshackle wooden seat. On this seat sat the two priests still in serious speech together	Just en aquest lloc imposant i deixat de la *Mà* de Déu hi havia un arbre i davall d'aquest, un banc de fusta desgavellat	No PU → PU

Mà(ns)	Secret Garden	In Greenford itself twenty people could swear to me for all that time	A Greenford mateix, vint persones podrien posar la *Mà* en el foc per mi durant tot aquest temps	No PU → PU

In the first, *neglected spot* is translated as **deixat… de la mà de Déu** (*forsaken by God's hand*). In the second, *twenty people could swear to me* is rendered as **vint persones podrien posar la mà en el foc per mi** (*twenty people could put their hand into the fire for me*). Admittedly, not all instances of No PU → PU are so colourfully expressive as these. Some of them come about as a result of the (almost logical) activation of a TL phraseme prompted by a given ST segment, even if the latter is not phraseological in nature. In those cases, it might be argued, the degree of phraseological gain is not too high. But then the converse is also true: many cases of loss are not outstanding either, as the ST phraseme is rather of the run-of-the-mill kind. At bottom, what *is* remarkable is that overall loss should be so small, indeed smaller than might have been predicted.

6. Conclusion

To sum up, the following moves have been made in the present article.

First, a conceptual or interpretive move: it has been assumed that phraseological usage can be seen as an indicator of normalisation in translated texts, as phraseological units are target-language standardised forms belonging to its lexical repertoire. As a consequence of their being frequently used, they have become part and parcel of the target language conventions. Since normalisation is all about *textual conventionality* (to put it in Vanderauwera's terms), the assumption made here does not seem hasty or unwarranted;

Second, a descriptive move: drawing on data yielded by the English-Catalan subcorpus of the COVALT corpus, it has been established that Catalan translated texts are *less phraseological* than their corresponding English source texts, but they are so only by a narrow margin. 483 bilingual concordance lines (containing phraseological units in English or Catalan or both) were analysed and labelled for the translation technique used. Some techniques implied the presence of a phraseological unit both in the source and in the target, but others resulted in either phraseological gain or loss. Instances resulting in gain failed to make up for instances of loss, but by a narrower margin than might have been predicted considering the translation difficulties often posed by phraseology. The narrow margin just referred to seems to point towards some effort on the part of translators to retain or recreate a noticeable degree of phraseological activity in translated texts.

The third move should be explanatory and take the form of a correlation between description of corpus findings and some other variable or factor *external* to the corpus, be it sociocultural or cognitive. This is exactly the kind of explanation advocated by Toury (1995) or Chesterman (2000) and sorely missing in most corpus-based studies. The main reason for this absence probably lies in the fact that a whole new set of methods would be needed for the task, mainly in the form of questionnaires, interviews and think-aloud protocols. As a result, corpus-based studies typically end on a new plaintive note: *now we know what happens (with regard to a given feature)*

in a cross-section of translator behaviour (representative of a specific cultural and temporal environment), but we do not know why it happens.

In fact, a few explanatory hypotheses have been put forward, though – symptomatically enough – not in empirical but speculative work. I would like to mention two of these hypotheses, which might be brought to bear on the findings presented here. The first is Halverson's (2003) argument that translation universals may have a cognitive basis, which, in a nutshell, she formulates as follows:

> The basic idea is straightforward: in a translation task, a semantic network is activated by lexical and grammatical structures in the ST. Within this activated network, which also includes nodes for TL words and grammatical structures, highly salient structures will exert a gravitational pull, resulting in an overrepresentation in translation of the specific TL lexical and grammatical structures that correspond to those salient nodes and configurations in the schematic network (Halverson 2003: 218).

The *highly salient structures* mentioned by Halverson are two: the category prototype and the highest level schema, i.e., structures which are more prototypical, less peripheral, and structures with a high level of generality, as opposed to more particular ones. This *gravitational pull* exerted by prototypical and general structures is claimed to provide a cognitive basis for such translation universals as simplification, normalisation and generalisation. The fact that, according to current literature in the field, a similar phenomenon has been observed in second language acquisition seems to lend support to Halverson's hypothesis.

The second explanatory hypothesis mentioned in the previous paragraph has been put forward by Pym (2007) and concerns translator attitude. Pym argues that the common denominator to such features as simplification, explicitation, normalisation, *and* interference (the latter posited by Toury but not even mentioned by scholars working with a methodology based on comparable corpora) is *risk aversion*. Translators avoid taking risks either by resorting to target-language accepted features and structures (which leads to simplification, explicitation, and normalisation, seen as overlapping and making target texts easier to read, in line with Toury's law of growing standardisation) or by falling back on the authority of the source text (which leads to interference). According to the author, this happens because there are no rewards for translators taking a riskier course of action. On the contrary, they are expected to be "basically nurturers, helpers, assistants, self-sacrificing mediators who tend to work in situations where receivers need added cognitive assistance (e.g. easier texts)" (Pym 2007: 16). Pym formulates his explanatory law as follows: "Translators will tend to avoid risk by standardizing language and/or channelling interference, if and then there are no rewards for them to do otherwise" (Pym 2007: 20).

Other explanatory hypotheses might possibly be put forward. Be that as it may, they will remain on the level of speculation unless they are put to the test of empirical data of the kind that can only be obtained by means of questionnaires, interviews and protocols, i.e., through methods focused on the process. Corpus-based research is an invaluable tool for description; for explanation, other complementary methods need to be brought into the picture. The latter fall outside the scope of the present article, but the confluence of product-oriented and process-oriented methods looks promising for the near future of our discipline.

NOTE

1. Research funds for this article have been provided by two research projects: HUM2006-11524/FILO, funded by the Spanish Ministry of Education and Science, and P1·1B2006-13, funded by the *Caixa Castelló – Bancaixa* Foundation, within the framework of an agreement with the Universitat Jaume I.

REFERENCES

Baker, Mona (1992): *In Other Words*. London: Routledge.

Baker, Mona (1993): Corpus Linguistics and Translation Studies – Implications and Applications. *In:* Mona Baker, Gill Francis, and Elena Tognini-Bonelli, eds. *Text and Technology. In Honour of John Sinclair.* Amsterdam: John Benjamins, 233-250.

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

Baker, Mona (1996): Corpus-based Translation Studies: The Challenges that Lie Ahead. *In:* Harold Somers, ed. *Terminology, LSP and Translation: Studies in Language Engineering, in Honour of Juan C. Sager.* Amsterdam: John Benjamins, 175-186.

Bernardini, Silvia (2005): Reviving old ideas: parallel and comparable analysis in translation studies – with an example from translation stylistics. *In:* Karin Aijmer and Cecilia Alvstad, eds. *New Tendencies in Translation Studies.* Göteborg: University of Göteborg, 5-18.

Bernardini, Silvia (2007): Collocations in Translated Language. Combining parallel, comparable and reference corpora. *In:* Matthew Davies, Paul Rayson, Susan Hunston, and Pernilla Danielsson, ed. *Proceedings of the Corpus Linguistics Conference.* (Corpus Linguistics Conference, Birmingham, 27-30 July 2007). Visited July 2nd, 2009, <http://ucrel.lancs.ac.uk/publications/CL2007/paper/15_Paper.pdf>.

Chesterman, Andrew (2000): A causal model for translation studies. *In:* Maeve Olohan, ed. *Intercultural Faultlines. Research Models in Translation Studies 1: Textual and Cognitive Aspects.* Manchester: St. Jerome, 15-27.

Chesterman, Andrew (2004): Hypotheses about translation universals. *In:* Gyde Hansen, Kirsten Malmkjær and Daniel Gile, eds. *Claims, Changes and Challenges in Translation Studies.* Amsterdam: John Benjamins, 1-13.

Corpas, Gloria (1997): *Manual de fraseología española.* Madrid: Gredos.

Corpas, Gloria (2003): *Diez años de investigación en fraseología: análisis sintáctico-semánticos, contrastivos y traductológicos.* Madrid: Iberoamericana.

Delabastita, Dirk (1996): Introduction. *The Translator.* 2(2):127-139.

Eskola, Sari (2004): Untypical frequencies in translated language: A corpus-based study on a literary corpus of translated and non-translated Finnish. *In:* Anna Mauranen and Pekka Kujamäki, eds. *Translation Universals: Do they Exist?* Amsterdam: John Benjamins, 83-99.

González-Rey, Maribel (1998): Estudio de la idiomaticidad en las unidades fraseológicas. *In:* Gerd Wotjak, ed. *Estudios de fraseología y fraseografía del español actual.* Madrid: Iberoamericana, 57-73.

Guzman, Josep R. and Serrano, Àlvar (2006): Alineamiento de frases y traducción: AlfraCOVALT y el procesamiento de corpus. *Sendebar.* 17, 169-186.

Guzman, Josep R. (2007): El uso de COVALT y AlfraCOVALT en el aprendizaje traductor. *In:* *Actas del XXIV Congreso Internacional de AESLA. Aprendizaje de lenguas, uso del lenguaje y modelación cognitiva: perspectivas aplicadas entre disciplinas.* Madrid: UNED, 1743-1749.

Halverson, Sandra (2003): The cognitive basis of translation universals. *Target.* 15(2):197-241.

Kenny, Dorothy (2001): *Lexis and Creativity in Translation. A Corpus-based Approach.* Manchester: St. Jerome.

Laviosa, Sara (2002): *Corpus-based Translation Studies: Theory, Findings, Applications.* Amsterdam: Rodopi.

Mauranen, Anna (2000): Strange Strings in Translated Language: A Study on Corpora. *In:* Maeve Olohan, ed. *Intercultural Faultlines. Research Models in Translation Studies 1: Textual and Cognitive Aspects.* Manchester: St. Jerome, 119-141.

Mauranen, Anna (2004): Corpora, universals and interference. *In:* Anna Mauranen and Pekka Kujamäki, eds. *Translation Universals: Do they Exist?* Amsterdam: John Benjamins, 65-82.

Mauranen, Anna (2005): Contrasting languages and varieties with translational corpora. *Languages in Contrast.* 5(1):73-92.

Molina Martínez, Lucía (2006): *El otoño del pingüino. Análisis descriptivo de la traducción de los culturemas.* Castelló: Servei de Publicacions de la Universitat Jaume I.

Neubert, Albrecht and Shreve, Gregory M. (1992): *Translation as Text.* Kent: The Kent State University Press.

Olohan, Maeve (2004): *Introducing Corpora in Translation Studies.* London: Routledge.

Parks, Tim (1998): *Translating Style. The English Modernists and their Italian Translations.* London: Cassell.

Pym, Anthony (2007): On Toury's laws of how translators translate. Visited on July 2nd, 2009, <http://www.tinet.org/~apym/on-line/translation/2007_toury_laws.pdf>.

Ruiz Gurillo, Leonor (1998): Una clasificación no discreta de las unidades fraseológicas del español. *In:* Gerd Wotjak, ed. *Estudios de fraseología y fraseografía del español actual.* Madrid: Vervuert, 13-37.

Teich, Elke (2003): *Cross-Linguistic Variation in System and Text.* Berlin: Mouton de Gruyter.

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

Van Lawick, Heike (2006): *Metàfora, fraseologia i traducció. Aplicació als somatismes en una obra de Bertolt Brecht.* Aachen: Shaker Verlag.

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

Vinay, Jean-Paul and Darbelnet, Jean (1958): *Stylistique comparée du français et de l'anglais. Méthode de traduction.* Paris: Didier.