

TNE, Budapest '96

Sentence structure and thematization
in comparable and parallel texts

by
Belinda Maia
Universidade do Porto

0 Introduction

My interest in sentence structure has grown out of years of teaching language and translation. However, if I were asked what forced me to face up to the need to analyse the problem more scientifically, one significant factor would be a student who I can only describe as suffering from some kind of ‘sentence structure aphasia’. We are all familiar with the student who creates ‘unnatural’ structures by simply translating word-for-word - but this was more complicated.

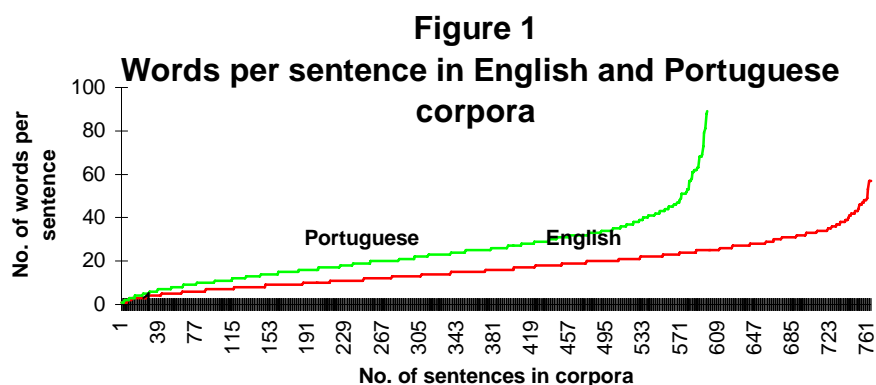
Attempts to teach her functional sentence analysis proved difficult, but I did find that such analysis helped others to analyse problems of text structure. It also led to speculation about what was ‘typical’ about the sentence structure of the two languages. Halliday’s theories on thematisation and information structure proved a useful basis for further work, and the adjustment of these theories to explain Portuguese proved illuminating.

Translation obviously involves far more than transposing the words and syntax of a text from one language to another, but we should not forget that this transposition is precisely what happens in much everyday translation. Although I agree that making students aware of the untranslatability of certain - usually lexical - facets of language is essential, it should also be possible to draw up some guidelines on what constitutes normal sentence structure in different languages and different genres.

1 Analysis of ‘comparable’ texts

A quantitative study I made last year Maia(1995), using a corpus of a variety of comparable texts showed that different languages vary quite considerably in the way their syntax encourages different types of sentence structure and thematization. This analysis showed up certain tendencies which could be seen as ‘typical’ of English and Portuguese, but it also demonstrated that some types of texts are more different than others, and that non-linguistic factors also affect the data on the different genres.

When I analysed a small corpus of comparable texts in English and Portuguese for sentence length, I found that there was a difference, as can be seen in Figure 1.



As expected, the average sentence length in the English texts was shorter than in the Portuguese texts - 17,6 to 23,5 words per sentence. However, the interesting point is that some genres are more different than others. English and Portuguese text conventions were actually quite similar when writing in some genres - as with interviews for magazines, but

with the more serious or academic the Portuguese text, the more the sentences lengthen in relation to their English counterparts. Although Portuguese syntax - with its inflections and system of concord - allows for greater complexity of structure than English, this fact reflects a cultural difference which suggests that Portuguese academics give value to the ability to formulate large units of information.

Another area of theoretical importance which showed distinct differences was that of sentence order. It seems natural to many people that the order of information is a function of the syntax of the sentence, and that this order exists for various psychological and cultural reasons. If this is true, then we have to account for the fact that different languages have quite different ways of organising information. Halliday's proposals for the notions of Theme and Rheme which, together with his notions of Given and New information in the sentence, is based on a strong psychological basis. My analysis of sentence structure takes this into account using the terms of Subject, Predicate, Objects, Complements and Adjuncts¹, and there is an obvious difference in thematisation between the languages.

Figure 2

	<i>English Corpus</i>	<i>Portuguese Corpus</i>
Subject as Theme	64%	31,8%
Predicate as Theme	3,8%	19,2%
Object as Theme	1,3%	1,7%
Complement as Theme	0,4%	0,5%
Adjunct as Theme	15,6%	26,6%
Conjunct as Theme	7,8%	8,6%
Disjunct as Theme	4%	4,5%
Abnormal structures	3,1%	7,1%

It would seem that, despite all the lengths English language teachers to which go in order to teach their foreign students different ways of thematizing sentences, the fact remains that natural English texts overwhelmingly favour the Subject-Predicate order. In the analysis, 88% of all sentences, whether or not the Subject was the theme of the sentence, followed this order, whereas only 51,1% of the Portuguese sentences used SP order. As can be seen in Figure 2, English sentences also showed a 64% preference for the Subject as Theme, whereas only 31,8% of the Portuguese sentences did².

These are the basic facts that I was able to establish with a small selection of texts. I am now working on collecting more comparable texts of certain genres in order to establish more specific characteristics.

2 Analysis of parallel texts

¹ The classification was that used in functionalist grammars, like Halliday (1985) and Downing and Locke (1992).

² In English thematisation of the Verb is rare, except with imperatives and, if we count the auxiliary verb with questions, and the Verb was only thematized in 3,8% of all the sentences. However, the Verb occurs in Theme position in 19,2% of all the Portuguese sentences studied. For years I have discouraged Portuguese students of English from giving the same generous distribution of adverbs about their English sentences that they use in Portuguese. Quantitatively it would seem I was right about their using them in initial position, with 26,6% of the Portuguese sentences having an Adjunct as theme, as against 15,6% in English.

Once one has laid the groundwork for characterising what is ‘normal’ sentence structure and thematisation in a certain genre in a certain language and shown how it differs from the ‘normality’ of another language, one can begin to consider how far translations measure up to the originals or to the expectations of the TL. When academics analyse originals and their translations, they usually focus the differences between them. However, it does not reflect the reality of everyday translation, or even much literary translation, the vast majority of which is surprisingly literal, as anyone who has searched for parallel texts which provide plenty of anomalies to discuss in a Translation theory class will be well aware. However, the parallel texts I chose for analysis turned out to be more alike than I had expected. Sentence length varied little and, although translators, with their tendency to greater explicitness, occasionally divided one sentence into two, cases of amalgamation of two short sentences were very rare.

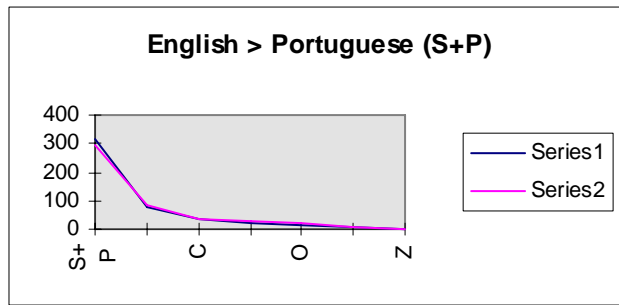
The aim of the work carried out was to see how far the translator is able, or obliged, to diverge from the original information structure by the syntactic or pragmatic conventions of the languages concerned. If theories of thematisation and information structure are to be taken seriously, one must accept that the order in which information is given is important, and that if a translator substantially alters it, s/he will be adapting it to the perceived needs of the TL - and often to her/his own more subjective interpretation.

The thematisation of the Predicate in Portuguese can be explained in various ways. One can simply attribute it to the syntactic behaviour of the verb which obviates the need for a pronoun by using inflection. However, as I have argued elsewhere, this is a simplistic attitude because it does not account for those occasions when a pronoun is used with an inflected verb. To suggest that the Predicate as Theme shows a focus on the action rather than the doer is a fairly strong semantic hypothesis, but quite feasible. Another explanation is to suggest that Portuguese texts should be analysed using the notion of Topic rather than Theme. In any case, the translator from English to Portuguese will frequently be obliged to make decisions as to whether to include the pronoun subject.

The use of Adjuncts, Conjuncts and Disjuncts as Theme vary considerably with text type - Conjuncts and Disjuncts, for example, being more common where the written word is imitating the spoken word. However, systematic thematisation of these elements also indicate a need by a speaker or writer to qualify statements. Hence their frequency in speech, where norms of politeness require such usage. If a genre requires their frequent use, the reasons are socio-semantic. If a large number of texts demonstrate that one language uses them more than another, as is the case with Portuguese in relation to English, one can hardly fail to conclude that this phenomenon reflects social usage. The decisions taken by the translator must take this into consideration as well as the more syntactic rules which prescribe certain unmarked positions for Adjuncts, particularly in English.

The differences in thematisation between the original texts and their translation were minimal, particularly if one considered Subject and Predicate thematisation together, as can be seen from Figure 3:

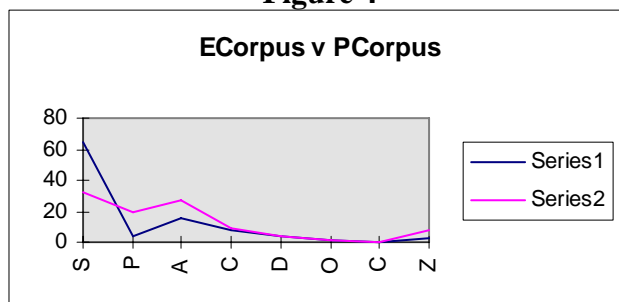
Figure 3



(Series 1 = English; Series 2 = Portuguese)

If Subject and Predicate thematisation were considered separately the difference is easy to see, as we can see in Figure 4:

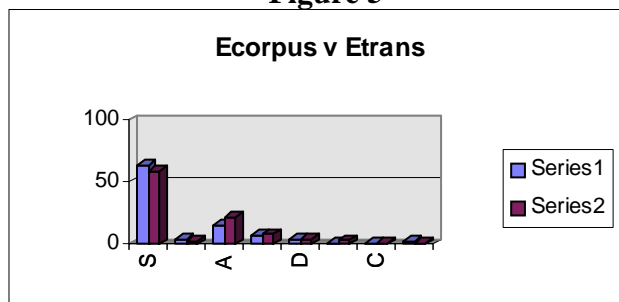
Figure 4



(Series 1 = English; Series 2 = Portuguese)

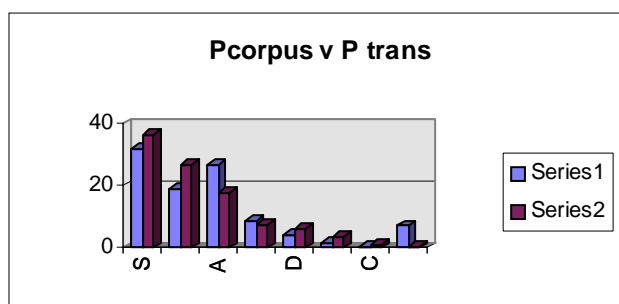
There was often more difference between one set of texts and another. However, when the translations were compared to the norms established for originals in the same language, differences could be noted, as can be seen in Figures 5 & 6:

Figure 5



(Series 1 = English original; Series 2 = English translation)

Figure 6



(Series 1 = Portuguese original; Series 2 = Portuguese translation)

Figure 5 shows that the English translation corpus shows an increase of Adjunct as theme and a decrease of Subject as Theme, which shows the influence of the Portuguese ‘norm’. However, the English syntax has pulled the texts into line with the regularisation of SP order.

Figure 6 shows a corresponding decrease in Adjunct as Theme and increase of Subject as Theme in the Portuguese translation corpus. There is also an increase in Predicate as Theme which reflects the SP overall dominance in English. There is also a decrease in anomalous sentences, which probably reflects the translator’s tendency to explicitness and correctness.

It would seem that, although the translations showed the pull of the norm of the TL language enough to differ a little from the norm of the SL text, especially in English, the pull of the norm of the original was also clear. Studying the variation, or lack of it, between original and translation might be a useful instrument in the evaluation of translatory competence in texts where acceptability in the TL is desirable.

However, perhaps the most interesting lessons to be learnt resulted from the differences between both original and translation and the norms established on the basis of a variety of texts. An analysis of the various psychological and cultural factors which encourage deviation from the structural norm in any specific text or genre could teach us something about underlying personal and cultural attitudes.

3 Analysis of individual texts and Genre analysis

When one thinks of individual texts and genres, one must also think of style, a notoriously difficult area for objective research. However, electronic corpora have now made it easier to perform analysis, and have enabled one to check whether certain words or phrases are in fact ‘typical’ of a particular genre or author. Farrington (1996), in her introduction to the Cusum Technique, cites such ‘well-known’ examples which have not withstood the test of corpora analysis. However, this technique, like others for establishing authorship, have shown that it is possible, using statistical techniques related to word and sentence length, to establish the writer’s equivalent of a ‘fingerprint’. It has also shown that this ‘fingerprint’ can be found on translations done by specific individuals.

One must therefore ask if any attempt at drawing general conclusions about texts is possible. However, we do recognise different texts as belonging to specific genres and so, providing we accept the limitations imposed by individual tendencies, it should be possible to explain why this is so fairly objectively. This can be done by collecting a representative corpus by a variety of authors in a recognised genre and analysing them. From the point of

view of translation, one should also collect a similar corpus of parallel texts, or original texts + their translations.

The form this analysis should take is obviously open to debate, but I would suggest that it could start by consideration of sentence and paragraph length, thematisation, information structure, and functional sentence analysis of the type already suggested. Once the more general factors have been established, one could then proceed to more detailed syntactic analysis of elements such as noun phrase structure, a particularly difficult area for translators of English and Portuguese.

4 Applications of this type of analysis

Texts chosen for students to translate should represent the content of the translation course one is teaching. However, even if people take a classification of texts - such as Mary Snell-Hornby's - as a basis³, a look at the texts which teachers assign their students for translation, shows that these selections tend to be idiosyncratic. Technical texts are chosen so that students will be forced to search for specialised dictionaries or consult specialists on the subject. General texts tend to focus on idiomatic usage or vocabulary which is difficult to translate for more cultural reasons. However, most are chosen to draw attention to lexical problems rather than structural ones.

Obviously, this lexical analysis is essential, but there should be a more systematic treatment of structure as well. This requires careful analysis of the text first, as general impressions of texts are often unreliable because if our attention is drawn to the unusual, we tend to ignore the surrounding normality⁴. If some form of preliminary examination of the text is carried out, with the help of the software for corpora analysis at our disposal, our choice can become more scientific.

Finally, since one of the objectives of conferences is, or should be, to promote cooperation between people with common interests, may I take the opportunity to express a wish to join forces with other people interested in assembling and analysing both comparable and parallel corpora in various languages. Doing it on one's own, or even in a small group, is time-consuming. The various ways in which these corpora could be used would benefit all areas of language study.

REFERENCES

³ Personally, since I teach translation into L2, I find I now select texts which I know non-native speakers are actually asked - and often paid - to do. These texts are usually of the informative type, but this does not mean that they are easy. In fact, they are often more difficult to do than the type of literary or journalistic text favoured by many teachers for the cultural brain-teasers posed by idiomatic language.

⁴ One of the texts which I chose for analysis was from Carl Sagan's *Cosmos*. On reading it through I noticed that he had used marked thematisation of Objects and Complements effectively. However, when the whole text had been analysed, I found that the incidence of this type of structure was no different than in other texts.

- DOWNING, Angela & LOCKE, Philip. 1992. *A University Course in English Grammar*. New York : Prentice-Hall.
- FARRINGDON, Jill M. 1996. *Analysing for Authorship - a Guide to the Cusum Technique*. Cardiff : University of Wales.
- HALLIDAY, M.A.K.1985. *An Introduction to Functional Grammar*. London : Edward Arnold.
- MAIA, Belinda. 1996. ' Parallel text and the study of sentence formation' in THELEN, Marcel and LEWANDOWSKA-TOMASZCZYK, Barbara. *Translation and Meaning - Part 4*. Maastricht : Universitaire Pers Maastricht.
- SNELL-HORNBY, Mary. 1988. *Translation Studies - An Integrated Approach*. Amsterdam/Philadelphia : John Benjamins.