THE SENTENCE AS A UNIT OF TRANSLATION

Belinda Maia

Maio, 1996
0 Introduction

This topic sounds very “linguistic” for discussion within the theme of “Tradução, Cultura, Sociedade”. It may even sound as if I have got stuck back in the days when translation theory was merely a branch of contrastive linguistics. Now, the academic pendulum has swung towards the consideration of the text as a unit for translation, and translation theory has been taken over with enthusiasm by literary theorists. We now have the umbrella term of ‘translation studies’, designed by its creators to cover a wide range of the linguistic and literary academic spectrum.

Translation studies experts will usually make an overall appreciation of the time, place and context of the text - an essential form of analysis which should be applied to all types of translation. They will take particular delight in discussing examples where the whole text has been reinterpreted. Susan Bassnett McGuire(1991) draws our attention to some particularly interesting examples of poetry and drama. João Duarte(1994) shows how Mickle’s translation of the Lusiadas was twice as long as the original and, how many English-speaking readers in the past felt the translator had improved on the original. Accusations of colonisation of texts are made against people like Mickle. Another technique used to make a translated version more ‘readable’ is omission. An example of this is a translation I have found of Lobo Antunes (1986) Os Cus de Judas in which sizeable chunks of heavy description and more subtle cultural-political references have been simply ignored. The examples of cultural and social prejudice in translation are legion.

However, when the translation follows the original fairly closely, as is the case of 99% of translated texts, many of the details chosen by translation analysts to demonstrate the difficulties of translation are found at the lexical level. The use of a word in the 16th century may be quite different to its use in the 20th. An English-speaking feminist writing a book on childcare makes a political point by referring to the ‘generic’ child as ‘she’. Yet in any Portuguese translation of this book, her political stance will be lost as the most normal translation of child is criança, which is feminine anyway. There is no doubt that our use of the lexicon can produce endless material with
which to show how languages vary quite considerably in the way they perceive the
world. And I mean endless, because the lexicon is always changing and adapting to the
needs of the reality it is supposed to describe. Differences in syntax are much more
subtle, and often left to theoretical linguists, partly because translation analysts are not
always experts in linguistics, and partly because the results of the analysis seem to be of
more immediate interest to language philosophers and artificial intelligence than
translators.

I am not alone in feeling that one should not move too far away from the realities
of language towards the so-called realities of the world. Gideon Toury was one of the
first to show the relevance of Even-Zohar's polysystem theory to translation and to
explore the cultural, social and historical implications. Yet I recently heard him express
the view that he now found himself moving back to linguistics to look for explanations. I
am personally fascinated by the interpretation of culture and society to be found in the
lexicon and syntax, but what I want to discuss here are the differences reflected at the
level of the sentence.

1 The sentence as a basic unit of translation

What is meant by a “sentence”? Roger Bell (1991: 29) says that “there is good
psychological and linguistic evidence to suggest that the unit (the translator actually
processes in the course of translating) tends to be the clause”. From the cognitive point
of view - of studying how a translator actually processes language - it is difficult to
contradict him. Words have to be coordinated into clauses before they acquire their full
meaning, and form basic units of communication. Communication in spoken language is
broken up into these units, even if processes of ellipsis and mutual understanding
between speaker and listener have reduced them to a single word. If we wish, in fact, we
can make all our language production explicit in terms of the syntactically defined clause
patterns to be found in grammars. However, from a psychological point of view, it is
significant that we rarely actually do so. Clauses are combined in relationships of co-
ordination and dependency, and reduced, using various forms of what Halliday (1976)
describes as *cohesion*, in order to save time and trouble. The result of all the mental activity which combines these clauses into a unit is what normal language terminology describes as a *sentence*. This is the interpretation I shall give to *sentence* here.

Communication through written language can only assume as much common background with the reader as the situation allows, and has no facial expression, body language and cueing to provide feedback on how communication is progressing. Therefore, written language has to organise, control and concentrate the information of several clauses, or communication units, into sentences. This process is highly sophisticated and everyone knows how one can return over and over again to a sentence to ‘polish it up’. Sentences of this kind are what occupies the minds of most translators in the normal process of translating, and it is time to make a systematic analysis of them.

2. **A proposal on how one could systematically study sentence structure and style**

Let me begin my proposal of what I mean by systematic study by describing the present situation of lexicography. No one doing serious lexicography nowadays will do it without the support of large electronic corpora, and dictionaries written before their advent need to be updated by reference to them if they are to remain respectable. Even the Oxford English Dictionary, based on decades of exhaustive manual collection of examples from an enormous number of texts, and considered by many as tantamount to a lexicographer’s Bible, is now the subject of a massive project, led by Patrick Hanks, to make an electronic corpus of all the texts used - and probably many others - to find out whether the examples chosen on a qualitative basis stand up to scrutiny on a quantitative basis. Not so long ago, John Sinclair, Hanks and their colleagues on the Cobuild dictionary project were being ridiculed for thinking that their pioneering work on electronic corpora could do better than the intuitions and education of the traditional lexicographer. Now their ideas are those of the lexicographical ‘establishment’ and, although the first Cobuild dictionary was based on a corpus of 7.5 million words - enormous in its day - corpora of 100 million are no longer unusual.

4
Once the use of ever larger corpora was recognised as essential for lexicography, people began to see the relevance of corpora for studying other aspects of language. Concordancing programmes, designed to bring up large numbers of individual lexemes for scrutiny are now having to add other sophisticated tools for text analysis. WORDSMITH, produced by a group from the University of Manchester and available through the Internet and Oxford University Press, also includes automatic listing of word frequency in texts, and statistical analysis to show which are the ‘keywords’ in a particular text. There is also a text aligner which allows for a text and its translation to be studied on the screen at the same time. Another project at the University of Birmingham is dedicated to refining this type of analysis using parallel texts in most European languages.

For some years there have been several projects dedicated to the tagging of electronic texts so that different syntactic structures can be concordanced easily and studied. In 1994 the Text Encoding Initiative Guidelines were drawn up and people are working hard to make them work with several languages. This is not easy, but it is a big step in the right direction. Once a sizeable number of texts have been tagged for syntax in several languages, we shall be able to see what patterns occur in different situations and genres, and how these patterns differ from one language to another. At the level of phrase structure elements, we shall be able to study certain linguistic phenomena at the level of what actually happens - rather than at that of what some well-informed grammarian or translator says on the subject. There are plenty of questions that I should like to see answered, but some of the most important areas are those of sentence length, sentence order and the relationship between sentences, the areas studied by experts in discourse analysis, pragmatics and stylistics. The material for research on translation should be ‘comparable’ and ‘parallel’ texts.

2.1 Sentence length

---

1 Parallel texts = originals + translations. Comparable texts = Texts in similar genres dealing with the same topic.
First let us look at sentence length. Elementary analysis of sentence length is even done by the grammar tool of a word-processor nowadays. I have carried out a more detailed form of analysis on a small corpus of comparable texts in English and Portuguese and the result can be seen in Figure 1.

![Figure 1 - Words per sentence in English and Portuguese corpora](image)

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, with English and Portuguese texts being actually quite similar when writing in some genres - as with interviews for magazines. The more serious or academic the Portuguese text, however, the more the sentences lengthen in relation to their English counterparts. This fact reflects a cultural difference which suggests that Portuguese culture gives value to the ability to formulate large units of information. The functional intention of sentences like the one in the following example appears to be that of giving several units of interconnected information, while qualifying them in a way which shows you are academically aware of all the implications of your statement:

**Example 1**

Mas com os estudos dos últimos anos, em que fomos servidos pelo progresso geral da Arqueologia e pelo avanço alcançado pela arqueologia dos habitats, em especial, nas técnicas de localização, métodos de escavação, análise e interpretação da informação, em conjugação com os dados fornecidos pelas fontes literárias e epigráficas, substancialmente ampliadas com o decorrer do tempo e de importância
The basic structure - Conj Adj Adj P(se) Cs S - is actually quite simple, as is the case with many long sentences. The embedding of dependent clauses, apposition, and other forms of qualification are techniques which the syntax of Portuguese - with its systems of concord - makes possible. English simply cannot construct such structures without risking ambiguity, because it lacks this syntactic feature. Critical views of academic language are held by speakers of both languages. One refers to the richness of Portuguese style versus the English inability to conceptualise in a complex manner - or Portuguese pomposity versus the synthetic and concise style of English - depending on one’s point of view. Cultural virtues and vices of this kind, however, have a linguistic basis.

If one can show that a certain text type in one language systematically uses shorter sentences than another, what can this teach us about our duties as a translator? It is not enough to simply say that one must chop up texts differently according to the language one is using - making shorter sentences in English and longer ones in Portuguese, for example. A quick look at parallel texts will show that translators rarely do this.

A sentence expresses some sort of wholeness of ideas. If a Portuguese writer has constructed a 115 word sentence, one will naturally need considerable knowledge and control of English to put all the information into one English sentence - and, of course, one could say the same of the Portuguese writer in relation to Portuguese. However, if one tries to break it up, one may lose the thematic thrust of the message, and also run the risk of making the translation even longer by having to repeat lexical and syntactic
structures which have been ellipted in the original. Similarly, joining up the shorter English sentences of the English academic tradition to make a more ‘respectable’ Portuguese text is not as straightforward as it might seem.

2.2 Thematisation and Sentence order

An area of theoretical importance that I believe needs to be studied very carefully is 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. English and Portuguese vary even more than I believed possible when I first started to analyse sentences.

Halliday’s proposals for the notions of Theme and Reme, linking them to notions of Given and New information in the sentence, is based on a strong psychological basis. He argues that whatever comes first in a sentence, whatever its syntactic or semantic function, is the Theme (ibid: 36-7) and says that

The Theme is a function in the CLAUSE AS MESSAGE. It is what the message is concerned with: the point of departure for what the speaker is going to say.

Not everyone would agree him, but I shall not go further into that debate here, and will assume that Theme refers to the initial component in a sentence. Halliday also argues that the Given information in the sentence also tends to come at the beginning of the sentence and New information at the end. He is not alone in making such suggestions. The belief that the natural order of a sentence is SVO is not an invention of modern grammar. Philosophical arguments defending this as the most logical way of expressing oneself go back a couple of centuries - at least among French and English speakers.

When I started work on the analysis of sentences, I was looking for a variety of factors. I examined the basic sentence structure in terms of Subject, Predicate, Objects,
Complements and Adjuncts\textsuperscript{2}. As few real-life sentences actually consist of the simple one word components found in linguistic examples - how simple our work would be if they were - this proved a challenging task! However, my attention was soon caught by the obvious difference in thematisation between the languages.

The analysis I made of my small experimental corpus would suggest that, despite all the lengths English language teachers go to 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 of the 765 sentences considered in the English part of the corpus\textsuperscript{3}, 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. English sentences also showed a 64% preference for the Subject as Theme (see Figure 2), or as being the first component of the sentence, whereas only 31,8% of the Portuguese sentences did. In English it is rare for the Verb to come first except with imperatives and, if we count the auxiliary verb, in Yes/No questions. 8,2% of the English sentences studied were imperatives or questions - some of which were in SP order - and yet 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.

Portuguese does not demand a pronoun with its verb except in cases when it is needed for disambiguation or emphasis\textsuperscript{4} so, if the topic has been introduced earlier the verb occurs naturally on its own, often in theme position. The verb + reflexive -\textit{se} is also used in Theme position without an expressed subject, often to convey the meaning given by \textit{it} + passive in English, as in \textit{Sabe-se} or \textit{It is known}. The Verb can and does precede the subject in other Portuguese sentences, particularly when a long and complex subject

\textsuperscript{2} The classification was that used in functionalist grammars, like Halliday (1985) and Downing and Locke (1992).

\textsuperscript{3} The English corpus consisted of 13,487 words divided into 765 sentences. The Portuguese corpus consisted of 14,044 words divided into 599 sentences.

\textsuperscript{4} See Maia (unpublished)
is involved\textsuperscript{5}. This is a useful device to avoid 'top-heavy' sentences, but it is not always easy to translate into English. This thematizing of the Verb suggests that Portuguese speakers often prefer to give more initial emphasis to the action, than to the doer of the action.

\begin{center}
\begin{tabular}{|l|c|c|}
\hline
 & English & Portuguese \\
\hline
Subject as Theme & 64\% & 31,8\% \\
Verb as Theme & 3,8\% & 19,2\% \\
Object as Theme & 1,3\% & 1,7\% \\
Adject 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\% \\
\hline
\end{tabular}
\end{center}

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.

If an adjunct of time is put in theme position by the author, it is for a reason. If I have to translate:

Na noite de S. João, os portuenses passam a noite na rua.

I can say:

On St. John’s Eve, the people of Porto roam the streets all night.

but I could argue that English prefers adjuncts to occur at the end of the sentence, with

The people of Porto roam the streets all night on St John’s Eve.

The difference is perhaps minimal, but there is a difference. The context of the Portuguese sentence probably refers to St. John’s Eve, and the sentence is to tell you

\textsuperscript{5} See Baker (1992 : Ch.5) and Johns (1991)
what people do then - not to tell you that the people of Porto roam the streets, particularly on St. John’s Eve!

Admittedly this is a simple example, but the more complex examples require more complex treatment. One cannot simply oblige the information supplying conventions of one language to obey those of another without changing the message. The thematising of Adjuncts in Portuguese may have a deeper psychological function of needing to establish the circumstances before discussing the action.

The proportion of sentences in which Conjuncts or Disjuncts were thematized in both languages was similar - 7,8% English - 8,6% Portuguese for Conjuncts, and 4% English and 4,5% Portuguese for Disjuncts. Then there was a group of sentences in each corpus which did not obey normal grammatical rules - 3,1% in English and 7,1% in Portuguese. Some of these sentences - which were classified as 'abnormal' tended to come from texts where they were 'reporting' some sort of speech situation, and where the context would supply the ellided elements. However, especially in written Portuguese, these structures were usually functioning as a continuation of the previous sentence, having been separated from it for reasons of emphasis, or in order to give the speaker/reader some 'breathing space' with a long sentence.

2.3 The relationship between sentences

The abnormal sentences recorded drew my attention to the relationship between sentences. A systematic study of parallel and comparable texts could also help us study the conventions existing at this level. More detailed tagging could reveal patterns of cohesion at the level of reference and ellipsis. Perhaps too, one could study how Portuguese concord affects information across sentences as in:

"Coimbra é por múltiplas razões a mais apaixonante das cidades portuguesas. Aureolada com epítetos vários desde longa data - justificados uns, injustificados outros - reúne de facto um conjunto de características, que a tornam única e inesquecível, para quantos nela passam algum tempo da sua vida. Situada nas margens do Mondego a
Examples like this call into question the idea of Theme and its relation to that of Topic. Some languages, like Chinese, for example, do not use thematisation as we do. They establish a Topic and may then construe a number of sentences all of which relate to it. The Portuguese system of concord allows a similar phenomenon to occur with sentences like the above.

3 Some implications of these differences for translation

Several questions arise from a study of this sort, and most of them lead back to the usual controversy of free v. literal, source language orientated v. target language orientated translation. Where do one’s loyalties lie? Should one be loyal to the order in which the original writer formed the message? After all, the order in which the information is given in a sentence is not accidental. It is conditioned by the rest of the text, and by the emphasis the writer wishes to give to particular points.

Some translation theorists say one should adapt the message to suit the target language's sensibilities, or even standardize structures to produce 'normal' target language forms. This does happen up to a point in any case, usually at a syntactic level. For example, a study of parallel texts will reveal that translators into English have to 'regularize' the Verb-Subject order of Portuguese sentences, and produce pronouns or introduce Subjects where none exist in the original. The data from the comparable texts would also suggest that a test of native/non-native competence in producing and translating English and Portuguese could be made by looking at the way adverbs are distributed in parallel texts.

Translating from English into Portuguese presents fewer apparent problems in this respect, as the rules for word order are far less rigid, and the regularity of the English original can usually be reproduced quite literally in Portuguese. However, will it sound natural to the Portuguese reader if one always respects the original?
I can hear some of you thinking - “Has she ever heard of free translation?” My answer is to draw your attention to the fact that a large proportion of translation is carried out in a very literal fashion. It is not easy to say why, but one reason is because the order of the message is important. I believe that a systematic analysis of source language texts could lead to firm clues about what constitutes a ‘natural’ text. Such clues would be an improvement on the impressionistic generalisations made by many when discussing things as difficult to define as style and genre.

Comparing a large number of original texts with translations would help us to understand the constraints of different languages and their cultures. However, I do not think we shall ever come to a happy agreement over the loyalties of the translator, as I believe much depends on the context in which the translation is carried out. The considerable differences in a target language text caused by re-formulating the information in a text designed to sell a product may be very important in a commercial context, where loyalty to the author is secondary to economic need. The translator may need to exercise more imagination here than with more literary texts, where loyalty to the information strategies of the author is more defensible.

From the point of view of anyone interested in the teaching of language and translation, I think this type of analysis can be particularly valuable. It helps me to understand why a grammatically correct sentence still reflects the information order of the speaker’s native language. The reason why is because when we talk about language and translation, we are talking not just about a random selection of semantic notions tied together with syntactic string, but of an organised way of thinking about and interpreting the world which mirrors our culture and the society we live in.

BIBLIOGRAPHY


BAKER, Mona (forthcoming) "Corpora in Translation Studies". In Target vol 7.


MAIA, Belinda (unpublished) ‘Word Order and Thematisation with the first Person singular in Portuguese and English’.