On the use of parallel texts in the comparison of languages

Diana Santos[*]
Grupo de Linguagem Natural do INESC

This paper provides a critical presentation of what has recently come to be called "corpus-based contrastive studies", or, more specifically, the study of parallel corpora (corpora of pairs of texts in two different languages, which are translationally related).

After describing several questions posed by this research paradigm in Section 1, I suggest, in Section 2, an argument for its adequacy and present, informally, in Section 3, a particular method for dealing with tense and aspect taking into account the issues previously mentioned: the Translation Network.

The languages discussed will be Portuguese and English.[1]

1. Problems involved in parallel corpora

There has been some enthusiasm regarding the appearance of parallel corpora in the computational linguistics community, as is probably best echoed by Isabelle et al. (1993:205): "existing translations contain more solutions to more translation problems than any other available resource".

The problems involving the use of such data are many and varied, though, as becomes evident once one notes that they imply that, in some way or other, the notion of translation be made clear. This is not a simple matter, as the field of translation studies demonstrates. In fact, one thing is to define some correct translations from expressions in one language into the other (hardly a trivial task, as the machine translation community knows only too well), which can be thought of as modelling translation competence; still another is to handle translation performance, i.e., to be able to conveniently process really occurring translations.

In addition to all general problems of performance models compared to competence models, which corpus linguists are well aware of, for parallel corpora three other kinds of phenomena are relevant as well: translation quality, translationese, and the logical diversity of translation correspondences.

1.1 Translation quality

Everyone is aware that translators sometimes fail, and that translations can be rated.

If one tries to go deep in what is supposed to be preserved, it becomes at once obvious that most translations, in fact, fail to convey every piece of information, and only those, present in the source text. On the other hand, it has proved extremely hard, if not impossible, to define a minimum that must get transmitted, even though some kinds of errors are easy to detect and check.

Some non-trivial examples may help to prove this point. How good are the following genuine translations?

The animals from miles around came to drink from the little pools, and the wild sheep and the deer, the pumas and raccoons, and the mice--all came to drink.
De quilómetros em volta, os animais vinham beber àquelas lagoas.
'From miles around, animals came to drink to those pools.'
She looked up at Kino when he came back; she saw him examine her ankles, cut and scratched from the stones and brush,
e olhou para Kino, que voltava e lhe via os tornozelos esfolados e feridos das pedras e das urzes.
'and she looked at Kino, who was coming back and could see her ankles...'
the doctor said, and he saw Kino's eyes flick involuntarily to the floor near the side post of the brush house.
disse o médico, ao mesmo tempo que seguia os olhos de Kino, irresistivelmente parados no chão ao pé do pilar da cabana.
'said the doctor, at the same time which he followed Kino's eys, irresistibly fixed at the ground near the post of the house'
Era um longo hábito de frade o que vestia, quase até às sandálias que, na lama, se distinguiam pouco dos pés magros e ossudos.
He wore a long friar's habit that reached almost to the muddy sandals on his thin, bony feet.
'It was a long habit of friar he had on, almost to the the sandals which, in the mud, were hard to distinguish from his thin and bony feet.'
Aconchegou-se no manto, e subiu as escadas.
He pulled the mantle closer about him and climbed the stairs.
'He made himself comfortable in his mantle, and climbed the stairs.'
Quando elas o deixaram exaustas, levantou-se para espreitar.
When they were exhausted and left, he got up to peep around.
'When they left him, exausted, he rose to have a look'

For instance, the first example is not necessarily a mistranslation. As pointed out by Lauri Carlson (p.c.), it is possible that "the Portuguese translator omits listing all the local animals coming to the waterhole because she estimates the effect of the list on her supposed audience would be opposite of its effect on readers of the original --- making the image more concrete and familiar for the former who have intimate acquaintance with the local fauna, but causing puzzlement or estrangement for foreign readers". In general, in all examples above it is difficult to point out some smaller translation correspondences, but none displays a blatant error, either.

One of the most complicating factors in the measurement of translation quality is the fact that translation is non-unique in a different way than any original utterance. I.e., in addition to matters of monolingual paraphrasibility, an important issue concerns the choice among different pieces of information contained in a source text, which will be taken up again in Section 1.3.

This is especially relevant when that information is conveyed by a grammatical marker, which generally acquires a set of related meanings by a process Dahl called "conventionalization of implicatures": "if some condition happens to be fulfilled frequently when a certain category is used, a stronger association may develop between the condition and the category in such a way that the condition comes to be understood as an integral part of the meaning of the category" (Dahl, 1985:11). I have given a first description of clusters of such meanings of Imperfeito in

Santos (1994) and presented some analyses of their translation in Santos (1995b).

I present here some other examples. When the English present perfect describes both an event and the information that its consequences are relevant, Portuguese must choose between the description of the event, or between the expression of the result state, as can be seen in the different actual choices for exactly similar sources. The first translation was provided in the published translation, the second one (between parentheses) I suggest as possible as well:

> *"This pearl **has become** my soul,"*
> --*Esta pérola **tornou**-se a minha alma.*
> 'this pearl became my soul'
> (--*Esta pérola é a minha alma, agora* 'this pearl is my soul, now')
> *now that you **have become** a rich man*
> --*Agora, que **és** um homem rico.*
> 'Now, that you are a rich man'
> (-- *Agora que te tornaste um homem rico* 'now that you became a rich man')

With events, the Portuguese Pret rito perfeito composto (PPC) expresses both an interval until now and a repetition of an event an indefinite number of times. In English, one must supply both the interval and the repetition explicitly, which means that often some of the information gets lost, as is illustrated in the next example. Note, however, that, while the choice was obligatory in the previous case, here one could have rendered the two pieces of meaning in English, by using either an explicit adverbial or the present perfect progressive form.

> *E eu disse [...] que **tens trabalhado** muito e até **tens estudado** com o Padre Manuel para que as palavras santas fiquem todas certas e nos seus lugares.*
> *And I said [...] that you**'ve worked** much and **have** even **studied** much with Padre Manuel so that the holy words would come out right and in their proper places.*
> 'And I said [...] that you've been working much and have even been studying much with Padre Manuel so that the holy words would all come out right and in their places.'

1.2 Translationese

Translation, it is often noted, can be seen as a process or as a result. The 'translation as result' view is concerned with the properties that a given text has, precisely because it was produced by translating another one.

Now, there are three kinds of properties one may be interested in, namely, (i) properties of all translations, what have been called "universals of translation" (see Baker, 1993), (ii) properties of translations specific to a particular language pair (source-target), which I call translationese, and (iii) properties of particular translations, depending on the author and/or the translator.

Obviously, it is the second kind which is of interest for contrasting two languages, even though it may be advisable to keep the other kinds in mind, as well.

In Santos (1995a), I define translationese as "the influence of properties of the source language in a translated text in a target language". One can also describe this phenomenon as language interference brought about by translation.[2]

At closer inspection, this influence can result in:

- the presence of properties of the source language in the target language text

- the absence of properties of the target language absent in the source language text

- the presence of properties of the target language exaggerated by the influence of conscious contrastive knowledge on the part of the translator.

As for this last case, even though it may reflect a property of texts created by translation, I prefer the term "anti-translationese" to describe it. In any case, it shares with translationese proper the deviation from the target language norms, naturalness and/or prototypicality.

The term "translationese" was to my knowledge first used by Gellerstam (1986) to describe vocabulary differences between original Swedish text and Swedish translated from English. His study contrasted the two kinds of text in terms of frequency of lexical items, and, from his results, he tried to identify why such differences occurred at all. (Incidently, and without being aware of it then, I was also studying translationese in Portuguese when I compared the classifications of Imperfeito in original and translated text in Santos (1994).)

Gellerstam, however, concentrated on target language texts only. In Santos (1995a), I suggested to broaden the notion of translationese by analysing parallel corpora, and inspect and use as evidence the source text and the actual translation relations as well[3], integrated in a general program of identifying the differences and similarities of two languages.

I present here some obvious real examples of translationese:

> *Apoiado à mesa, arrastou-se até à outra ponta, e daí, deixou-se cair até à enxerga.*
> *Leaning on the table, he dragged himself to the far side and from there let himself fall to the cot.*
>
> *Deixou-se ficar estendida, saboreando uma incomodidade que era exaurido repouso.*
> *She let herself remain stretched out, savoring a discomfort that was to her an exhausted repose.*
>
> *Long after Juan Tomás had gone Kino sat brooding on his sleeping mat.*
> *Muito tempo depois de João Tomás ter saído, ainda Kino estava sentado na esteira, meditando.*

1.3 The logical diversity of translation relations

Inspecting a number of actual human translations, I was able to arrive at the following pre-theoretical typology (the adjective "pre-theoretical" refers to the fact that no semantic description is provided underlying the particular analyses presented). The typology is intended to classify all translation pairs according to degrees of information preserving, or, rather, of semantic relationship between the two elements of the pair.

(Related typologies have been presented in the literature, regarding translation problems (Nida, 1959), linguistic properties (Keenan, 1978, Bar-On, 1993) and syntactic correspondence

(Wollin, 1981), Platzack, 1983). None concerned, however, the logical relationship between the two elements of the pair.)

Case A: Preservation of information. This category can actually be further subdivided into two different cases, though hard to distinguish in practice:

A1. Semantic equivalence. Through different linguistic means, the two elements of a translation pairs say "exactly the same thing", although in two different linguistic systems[4]. These cases are harder to find than one might expect. A possible non-trivial case is

E se não dava com a sepultura?
And if he couldn't find the grave?

A2. Contextual equivalence. I.e., even though the meaning of the two sentences differs, in the particular context they are used, they happen to convey the same information. Consider[5]

the stinging pain of the bite ***was going away***
a dor da picada ***diminuía***
'the pain was diminishing'

The English sentence describes a situation with a defined goal, namely the pain disappearance, and states that a process leading eventually to such a goal is in process. The Portuguese sentence, in turn, describes a gradual process occurring. That process may eventually lead to total disappearance, but that is not inherent to the meaning of the word *diminuir* ('diminish'). However, neither language in fact states that the pain went away. Both express only a gradual decrease. Thus, even though *diminuir* and *go away* are not equivalent semantically, they convey equivalent information in this particular context.

A more complex case is displayed in the following example, which involves, for its assessment, a context larger than the sentence:

He ***was quiet*** *now*
Agora a criança ***acalmara***
'now the child had calmed down'.

Clearly, for one to be quiet one does not need to have calmed down the moment before, and so the two sentences do not convey the same meaning in isolation. But, given their place in the narrative structure (roughly, following a description of a situation when that baby had been screaming with pain), they can be considered contextually equivalent.

I should note that the more different two languages are the more difficult it is to state that the non-contextualized linguistic meanings of the two elements of a translation pair are the same. On the other hand, given that natural language always occurs in context, one might even question the interest of the A1 category in itself.

Case B: Subsumption. By supsumption I mean that the translation is less specific than the source. Logically, the source element of the pair entails the target one.

I subdivide this category into two cases.

B1. Proper subsumption. A standard example seems to be:

where the pearl was ***buried***
onde a pérola estava ***guardada***
'where the pearl was kept'.

Omitting information generally not relevant for the target language seems, actually, to be a very common situation: In this case, it is well known that manner of position is fairly secondary in Portuguese. This particular example, however, is additionally interesting because it raises the question of whether there would be the possibility of information preservation in Portuguese at all.

At first sight, the word *enterrar* (the standard translation of *bury*) would be the right choice, and thus the preference of the translator in using a less specific verb would illustrate a strong preference of the target language, going against exact translation in that particular translation pair.

However, Lauri Carlson (p.c.) has pointed to me that *bury* in English is also used with the figurative meaning "hide". If this is not the case with *enterrar* (which I believe it is true, i.e., *enterrar* is not standardly used to express "esconder debaixo de alguma coisa"), then we could have the English sentence using *bury* to convey "hide under something", while *enterrar* would mean the more specific "hide underground". So, there would not be a (lexical) term in Portuguese with the meaning of figurative *bury*, and the translator would have been obliged to use a subsumption strategy.

Now, more interesting still, one could in this case argue that actually the pearl was buried in the sense of "hidden underground", because that much is clear from the previous text, and so the translation *enterrar* would be exact anyway. Still, it could never be objectively decided whether the particular use of *bury* in the sentence in question was meant to convey that much. This, therefore, illustrates well the subtleties involved in translation choice.

<u>B2. Partial translation.</u> Partial translation occurs when the translation reflects a choice among different pieces of information. Although it can be logically described in the same way, i.e., the translation is implied by the source element, one feature of it strikes me as being sufficiently different to deserve separate recognition: Partial translation involves, essentially, choice among arbitrary collections of information encoded in a very same item, compactly.

Contrarily to the lexical case above, classified as proper subsumption, where one could (and would) talk about one genus term and differential features, in the cases of partial translation there does not seem to exist a hierarchy among the pieces of information conveyed. So, the choice among them is all the more arbitrary.

Another way to put the difference between the two kinds of subsumption is to note that there is a unique dominating node in the subsumption hierarchy in the proper case, but not so in the other one.

An example may help to explain what I have in mind. Let us note the progressive form:

> *it felt for the source of the death that* ***was coming*** *to it*
> *queria encontrar a causa da morte que o* ***rondava***
> 'it wanted to find the cause of the death that surrounded him'

In the English sentence, two reasons for the use of the English progressive can be invoked: the approach of a goal (the arrival of death), and the temporal co-occurrence of *feel* and *come*.

The translation preserves the temporal co-occurrence connotation, but fails to convey the approaching to the goal/end, since it chooses a manner of movement, *rondar* (typically with connotations of evil).

What is relevant to distinguish this case from proper subsumption, I claim, is that there is no general criterion to decide which feature is more basic than the other. Approach of a goal and co-temporality are simply not comparable.

Case C: Information addition. More information than the one present in the source language is displayed in the translation. Addition of information, actually a very common case, may be required by the target language or triggered by personal choice, and in general involves a combination of both.

In a pure logical formulation, one could describe this case as the translation entailing the source text. This formulation sounds misleading, though, because it is against the actual direction of translation (it seems to suggest that the source was derived from the translation). There is nevertheless a case in which this formulation is adequate: when the translator's understanding or interpretation of the source text corresponds to a more specific situation than that actually expressed by the original text; situation which is nevertheless conveyed by the target text. I will call this <u>C1. Reconstruction</u>. Examples are

Haveis-me entendido, haveis gostado?
Have you understood me and have you liked ***what you heard****?*
'Did you understand, did you like?'

Ficou olhando as chispinhas delicadas que a candeia fazia
He ***sat*** *watching the delicate sparks that the lamp gave out*
'He stayed looking at the delicate sparks...'

In the first example, *like* strongly expects a direct object, while in the second the manner of location, *sat*, seems to be the only natural way of expressing the sentence in English. (However, it should be noted that while the first addition seems to agree with the original meaning, in that it is a question asked after a speech, in the second text there is absolutely no information that could be said to support the position introduced by the translator.)

The other case I call <u>C2. Free addition</u>, which is not logically related in any way to the original. There is not much one can say about these additions except for considerations ranging from adequacy from a literary point of view to unjustifiable damage. The following examples illustrate: in the first pair, the translator added an explicit comparison, in the second, graduality.

Kino hovered; he was helpless, he was in the way
Kino ficou para ali, sem esperança, ***como um espantalho***
'Kino was there, without hope, like a scarecrow'

Coyotito's screams turned to moans
os gritos de Coyotito ***iam-se*** *transformando em gemidos*
'Coyotito's screams gradually turned into moans'

One might consider this last example as reconstruction, since it adds manner exactly in the same way as the last example of C1. However, the target language in this case does not require manner specification in order to sound natural, and, therefore, the reconstruction was "free". I believe, therefore, that a fine classification of translation pairs should also take into account the two language systems involved.

In fact, one specially interesting case of information addition concerns translation pairs where the source element is vague regarding a given piece of information which the target language is forced to specify (such cases have been described in Nida (1959), Keenan (1978), and, since Kameyama et al. (1991), are commonly called "translation mismatches"). Naturally, the target language is forced to add information that is not present in the source language; however, this addition may entail a significant loss as well, whenever the other choice(s) were also conveyed by the source text), i.e., when the source language description was more encompassing. Clear cases in the English-Portuguese pair concern temporal reference:

He was trapped as his people ***were*** *always trapped*
Estava peado, como todos os da sua raça sempre ***tinham estado***
'He was trapped as his people had always been'

Portuguese has to choose either remote or co-occurrent time specification, since there is no way to convey, with a single verb-tense combination, both past and simultaneous validity as in English.

An analogous situation can be found in this example

Perhaps he alone did this and perhaps all of his people ***did*** *it.*
Talvez ninguém mais fizesse aquilo e talvez todos os seus o ***tivessem feito****.*
'Maybe nobody else did that and maybe all of his had done it'

In general, translation problems related to vagueness are pervasive. Other such cases between English and Portuguese involve grammatical gender for human nouns, and decision between inception or continuous state:

But Kino ***was*** *in motion.*
Mas Kino ***começou*** *a mexer-se*
'But Kino started to move'

In this case, the Portuguese translator had to choose between an inceptive or a stative description, while one can say that English is not specific about it.

Case D: Translation through intermediary representation. Another theoretical possiblity, suggested to me by Lauri Carlson (p.c.), is the case where both elements of the pair are related to a third one, intermediary, as it is, which originates from the understanding of the translator, and which may entail (or be entailed by) both elements, which do not therefore stand in a direct relationship between each other. Suggested examples are represented with the intermediary situation in intermediate position:

He lived on charity -> HE WAS POOR -> il n'avait pas un sou

The man was hanged -> HE WAS EXECUTED -> l'homme a été guillotiné

In both cases, two different specific instances are related to the same general information, so one could talk about moving to a sister node instead of going up in an information hierarchy (subsumption) or down (adding information).

A very common case is the translation of causes by results and vice versa (the intermediate representation, not expressed in either language, would contain both the cause and the result). Examples are

> *-- Quere-as? -- e as flores já se* ***alteavam*** *nas mãos do outro.*
> *--"Do you want them?"--and he* ***placed*** *the flowers in the other man's hands.*
> '"Do you want them?" And the flowers already rose in the other's hands'

> *and the pearl, knocked from his hand,* ***lay winking behind*** *a little stone in the pathway. It glinted in the soft moonlight.*
> *Mas a pérola, que lhe saltara da mão,* ***rolara*** *na terra* ***para trás de*** *uma pedra do caminho e cintilava sob a pálida lua.*
> 'But the pearl, who had jumped from him from the hand, had rolled into behind of a path stone and twinkled under the pale moon.'

This paves the way to the next (and final) case, that of translation mistakes. Because the understanding of the translator may be hindered, the intermediary situation may not be logically related to the source one, and thus semantically unrelated translations may occur.

Case E: Translation failure. Translation failure, defined here as the case of no semantic relationship between the source and the translation, is not a random process. In fact, one can attribute it mainly to two different reasons:

<u>E1. Misrepresentation.</u> As mentioned above, misrepresentation of the situation described in the source text, for lack of knowledge of the language, or of the situation it is depicting (a cultural problem), can result in wrong translation.

<u>E2. Interference.</u> Interference because of (misleading) similarities between formal systems of the two languages. The problem of false friends in the lexicon is well known, but I could observe the same problem involving grammatical markers fairly frequently, as well.

It is interesting to note, actually, that the existence of E2 cases provides evidence for the claim that human translators may often be structure driven, and not only interpretation driven; on the other hand, cases D and E1 can only be explained through (often rather elaborate) reasoning from the part of the translator.

As hinted above during my discussion of translation quality, translation failure is a difficult notion. How far, and how much must one disagree not to consider a given pair neither addition of information, partial interpretation, nor parallel (in the sense of sister node) rendering, but rather translation failure *tout court*? In general, it is perhaps better to speak of relative translation failure.

2. The argument for the use of parallel corpora in language contrast[6]

Considering what has been said in the previous sections, the proposal of quasi automatically eliciting semantic data from large parallel corpora does not look especially bright. In fact, two major difficulties arise:

1. When (if ever) can one take the texts in the two languages to express the same meaning?

(Irrespective of which definition of meaning one is working with.)

Section 1.3 suggested a fivefold classification which is illustrative of the problems of an "equivalence of meaning" assumption. One must be aware of the several semantic relationships actually occurring in translation pairs if one wants to use translations as data.

2. When (if ever) can one take the target language text to be an adequate sample of the target language (in the way it conveys meanings, for example)?

Here, again, there is no simple answer. As discussed in Section 1.2, a translated text is surely not a typical target language text, given that it was conceived in the source language system, with different expressive means, different discourse strategies, different cultural backgrounds, and a different stock of lexical items. (Actually, Baker (1993) has even proposed the study of translations as a text type in their own right.) One must also study original target language texts, as is proposed as a bilingual corpus principle in Johansson & Hofland (1993).

These two questions do no more than re-state one dialectics involved in the process of translation, namely, the desire to follow two principles, which are often incompatible: being faithful to the original (often called exactness or adequacy) and being faithful to the target language (also standardly referred to as naturalness).

One should thus address the fundamental question of whether contrastive studies should rely on real translations at all, given the problems highlighted above.

Now, on the one hand, according to van Buren (1990:85), "it is logically impossible to engage in contrastive analysis without postulating common categories of one sort or another since, more generally, it is logically impossible to compare any two entities without using the same frame of reference".

On the other hand, there is no guarantee that underlying semantic categories in any two languages cover the same range, nor is it necessary that they are involved in the same situations. I will present some examples presently that show that this is not the case.

Therefore, it seems that these two claims cannot be held simultaneously: rather, one seems to be forced either to abandon the requirement that each language be described in terms of its own categories, or to give up the hope of comparing two languages before a "language-independent language" be found that can be used as means of comparison.

There is, fortunately, one way out: that of <u>measuring one language via the other language's categories</u>. This is, in fact, what people have always (implicitly) done, often using English as the measuring rod (as I noted in Santos, 1992).

It is, moreover, what translation in practice does: it expresses the categories of one language through those of another (the target language).

The recognition of this fact thus solves our problem in a surprising manner: The use of real translations turns out to be in fact the best unbiased way (if not the only one) to get objective data for the contrast or comparison of two languages, provided, of course, that we are aware of the relativity of each language's concepts, and of the distortions possibly brought about by looking at one language with spectacles from another.

This view allows one, furthermore, to look at translation phenomena in a different perspective: Instead of viewing real translations as deviating from the ideal of "equal content in each side of the translation pair", now it is obvious and predictable that two translationally related entities do not necessarily mean the same.

An interesting case is when two items in a translation pair are instances of different categories of different languages, and meanings which are secondary or derived implications in one language are instead expressed through the core of the translated category in the other language.

For example, English marks "in progress" and encodes lexically the feature "has an inherent end". On the other hand, Portuguese has a category "temporary state". Now, it is natural (and actually frequent) that something marked as being in progress in English is translated into Portuguese by something temporary. After all, something which has an inherent end and is in progress is bound to finish. However, a Portuguese temporary state is not necessarily something which is in progress and ends.

Another example: Portuguese marks "according to plan", while English encodes present relevance grammatically. Often, the occurrence of what is according to plan is relevant to the present, and thus these grammatical features are translationally related.

This view of considering the process and result of translation as the way to do contrastive studies suggests precise guidelines for language comparison:

1. The analysis of each language must be done following the categories suggested by the language itself. Therefore, a different metalanguage should be used (which is trivially simple if one uses a natural metalanguage).

2. Languages should be contrasted without resorting to a third metalanguage ("interlingua", independent knowledge representation language, or the like), but rather by using each other’s metalanguage.

3. The contrast must always be directional and performed in both directions, i.e., L1 must be seen through L2, and L2 through L1.

4. Objective data for this contrast should be furnished by the analysis of translations in the two directions. This analysis must be fine-grained, though, because the considerations discussed above and summarized here concerning translations as data must be kept in mind:

a) translation is not unique:

- there is choice among the several features expressed by the source utterance

- there is information which is conveyed unintentionally (it is not functionally relevant)

- there will be arbitrary choices taken in translation, which are neither representative of the source nor of the target language

b) translation is not perfect:

- the translator may have a deficient knowledge of the source language

- the translator may have insuficient knowledge of the subject matter, culture or simply of the kind of situation depicted in the source text

- s/he may be influenced by formal similarity (interference)

- random phenomena may occur (such as typos or omissions in the texts)

c) the language of a translation is not a good representative of the target language: it will tend to select the categories which it shares with the source language (or at least uses to render them) and will have less emphasis on its own categories which are not shared by the source language (or, at most, are not useful in translation).

d) there will be cases in which the lack of parallelism is so huge that one cannot compare (formal) categories (for example, different sentence structure or radically different lexical items may prevent meaningful comparison of tenses).

3. Comparing the English and Portuguese tense and aspect systems

Considering the more restricted task of comparing the tense and aspect systems of the two languages, I suggest here a model that takes into consideration most issues described above, the Translation Network, consisting of two aspectual networks (one for each language) joined by a set of directional links.

For lack of space, I will not be able to describe an aspectual network in any detail. The reader is referred to Moens (1987) for the original proposal, and to Santos (1993) for some changes to it. To understand what follows it suffices to know that it is a model of aspect, which postulates categories (depicted as nodes) which can be expressed lexically, gramatically or even contextually. Aspectual operators are modelled as specifying input and output categories, and are depicted by labelled arcs in the network. Unlabelled arcs are meant to indicate that an aspect change may be induced by the context without an overt indication of it.

A translation network is meant as a device which allows a fine-grained description of actual translation pairs, in addition to being a model of the differences between the two systems. The complexity of the links accounts for the complexity of the language comparison/contrast, while the comparison of the particular paths in particular translation pairs provides a model of the complex processes involved in translation.

A translation network is, basically, a set of unlabelled directed arcs, which bridge the two (sets of) monolingual categories. These arcs, which I call "translational arcs", correspond to the linking of two different language systems. Some properties of this device are worth noting at once:

1. A translation network is, by definition, directional. Even though such a requirement follows from my previous discussion on translation, I would like to stress this point once more.

2. Secondly, the source nodes and the target nodes do not usually have the same names. Rather, it is often the case that they do not, since it is known that most categories (and thus node names) do not have a perfect correspondence in the target language.

It is a very satisfying property of this model, in my opinion, that it uses no common metalanguage. It allows for different semantic concepts and different semantic primitives in each language, as well as different semantic operations. This is, as argued in Santos (1992), important also from a practical point of view: a) Grammar developers for each individual language would certainly benefit from using a metalanguage closer to their intuitions. b) The increasing use of on-line dictionaries, thesauri, and corpora to semi-automatize language processing implies that the metalanguage used has to be somehow related to the original natural language. c) It will not conceal dissimilarities between "similar" constructions in different languages.

3. A third property of the translation network is that it can be read as a static description of the relationship between the two language systems, or be used for the description of a particular translation pair. Considering this latter use, actual translations are represented by one (or *n*) path in the source network, one (or *m*) path in the target network, and one (or *n*x*m*) set of pairings among nodes of the two paths. This set of pairings is intended to model a possibly parallel compositional derivation. Pairings occur when there are nodes linked in the two networks by translational arcs, occurring in a compatible sequence. Only one pairing is required (at the sentential level) for two sentences to be translations of each other, however. (These pairings can play a role in the evaluation and classification of translation pairs, as should be obvious.)

In individual translation pairs, translational arcs may obviously perform a level shift. For example, they may replace a whole complex expression by a lexical item, or vice versa.

4. There is only one feature of the translation network I propose which does not necessarily follow from the setup described, i.e., from the directional coupling of two aspectual networks. This feature is the allowance of unlabelled transitions in the source language, as "coercion brought about by translation". I use it, for example, to model cases involving vagueness: a vague sentence in the source language is translated (and interpreted) as belonging to a more specific category of that very same language, which is then standardly translated into the "corresponding" category of the target language.

This constitutes an interesting property, namely, that the coupling of two aspectual networks may produce some changes / additions to the source network. Actually, it implies a subtle perspective difference regarding translationese, which, contrarily to the way it is generally conceived, is not modelled in the translation network by influence of the source language over the target language, but the other way around. Namely, a different interpretation of the <u>source</u> language text brought about by translation is posited (in other words, interpretation through target language eyes). Now, this may still result in special properties of the target text, but their explanation may prove to be more adequate this way -- at least when natives of the target language perform the translation.

5. Finally, the translation network accounts neatly for the non-uniqueness of translation: given a particular sentence in one language there are often many different ways to render it in the other, neither of which can be claimed to be better on objective grounds. In the network, this is

modelled in two different ways: (a) by the possibility of several different target paths sharing the same number of translational arcs, albeit different; and (b) by the possibility of choosing, among several possible source paths for the same sentence, a particular one to translate.

Two specific questions concerning this model must still be addressed:

First of all, if one wants to measure the quality of a given translation pair, one should add a different parameter, namely, a measure of translationese, to what is provided by the translation network. In fact, a given translation may succeed in conveying every transition of the original, but at the cost of being extremely awkward as a target language sentence. One should thus rate that translation as poor. Two separate measures are thus necessary to evaluate a given translation, one based on the translation network concerning closeness to the original meaning, the other reflecting typicality in the target language system. These two measures should, moreover, be kept distinct, not only because it is unclear how to combine them but also because they could be used for different purposes. For example, by generalizing over cases of marked translationese one could be able to pinpoint actual language differences not as yet noticed, as well as classify translation problems for translation teaching.

Then, the hardest problem involved in the formalization of the translation network is the status of the translational arcs (a question which will then have import on the evaluation and description of the translations). The problem is that there is no sound evidence for the translational arcs, except perhaps translation practice. But, as noted by Neubert, *it is an illusion to think that practical results alone will decide what is communicatively equivalent or not* (Neubert, 1986:35). In fact, it is natural to believe that not all translational arcs are equal either. Some of them entail more loss or addition of information than others. Yet one other factor must then be introduced in the evaluation of a translation, weighting translational arcs: the semantic relationship among nodes in the two different language systems.

4. Conclusion

I claimed in this paper that parallel corpora are more useful than damaging to contrastive analysis. Their exploration must, however, be done with care, since they provide us with data whose rationale can be manifold, ranging from systematic language differences, systematic language influence, to non-systematic individual interpretations and individual stylistic options. Even if it is possible to separate and concentrate only on systematic phenomena, one must still have some model of each language system to be able to draw conclusions about their interrelationship, given that one cannot assume that translationally related categories mean the same. Here, I suggested a fairly sophisticated model to rigorously state contrastive observations in the tense and aspect domain. That it is feasible to deal with large quantities of data using such a model is something I must demonstrate later.

References

Baker, Mona. 1993. "Corpus Linguistics and Translation Studies: Implications and

Applications", in M. Baker, G. Francis & E. Tognini-Bonelli (eds.), *Text and technology. In honour of John Sinclair*, Benjamins, 1993, 233-250.

Bar-On, Dorit. 1993. "Indeterminacy of Translation: Theory and Practice", *Philosophy and Phenomenological Research*, Vol. LIII, No. 4, December 1993, 781-810.

van Buren, 1990. "Contrastive analysis", in Fisiak, Jasek (ed.), *Further insights into contrastive linguistics*, John Benjamins, 1990, 83-117.

Dahl, Östen. 1985. *Tense and Aspect Systems*, Basil Blackwell, 1985.

Gellerstam, Martin. 1986. "Translationese in Swedish novels translated from English", in Lars Wollin & Hans Lindquist (eds.), *Translation Studies in Scandinavia*, CWK Gleerup, Lund, 1986, 88-95.

Isabelle, Pierre, Marc Dymetman, George Foster, Jean-Marc Jutras, Elliot Macklovitch, Fran ois Perrault, Xiaobo Ren & Michel Simard. 1993. "Translation Analysis and Translation Automation", *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation, TMI'93* (Kyoto, July 14-16, 1993), 201-17.

Johansson, Stig & Knut Hofland. 1993. "Towards an English-Norwegian parallel corpus", in U. Fries, G. Tottie & P. Schneider, *Creating and using English Language corpora*, Rodopi, Amsterdam, 1993, 25-37.

Kameyama, Megumi, Ryo Ochitani & Stanley Peters. 1991. "Resolving Translation Mismatches With Information Flow", *Proceedings of the 29th Annual Meeting of the ACL* (Berkeley, 18-21 June 1991), 193-200.

Keenan, Edward L. 1978. "Some Logical Problems in Translation", in F. Guenthner & M. Guenthner-Reutter (eds.), *Meaning and Translation: Philosophical and Linguistic Approaches*, Duckworth, 1978, 157-89.

Moens, Marc. 1987. "Tense, Aspect and Temporal Reference", PhD thesis, University of Edinburgh, 1987.

Neubert, Albrecht. 1986. "On the interface between translation theory and translation practice", in Lars Wollin & Hans Lindquist (eds.), *Translation studies in Scandinavia*, CWK Gleerup, Lund, 1986, 25-36.

Nida, Eugene A. 1959. "Principles of translation as exemplified by Bible translating", in Brower, Reuben A. (ed.), *On Translation*, Harvard University Press, 1959, 11-31.

Platzack, Christer. 1983. "Six translations into Swedish of Lewis Carroll's Alice in Wonderland", in Gunnel Engwall & Regina af Geijerstam (eds.), *Från Språk till Språk*, Sjutton uppsatser om litterar oversattning. Lund, 1983.

Santos, Diana. 1992. "Semantics and (machine) translation", *Actas do VIII Encontro da Associação Portuguesa de Linguística* (Lisboa, 1-3 Outubro 1992), 451-464.

Santos, Diana. 1993. "Integrating tense, aspect and genericity", *Actas do IX Encontro da Associação Portuguesa de Linguística* (Coimbra, 29 Setembro - 1 Outubro 1993), 391-405.

Santos, Diana. 1994. "Imperfeito: a broad-coverage study", *Actas do X Encontro da Associação Portuguesa de Linguística* (Évora, 6-8 de Outubro de 1994), 525-539.

Santos, Diana. 1995a. "On grammatical translationese", *Short papers presented at the Tenth Scandinavian Conference on Computational Linguistics* (Helsinki, 29-30th May 1995), compiled by Kimmo Koskenniemi, 59-66.

Santos, Diana. 1995b. "L'Imperfeito portugais: Etude systématique de ses fonctions et de comment en rendre compte en traduisant vers l’anglais", *Actes du XXIV Colloque sur la linguistique des langues romanes* (Palermo, 18-24 Séptembre 1995), to appear.

Wollin, Lars. 1981. *Svensk latinoversattning*, I. Processen. Lund, 1981.

---

[*] I gratefully acknowledge a PhD scholarship granted by JNICT (Programa PRAXIS XXI). I am indebted to Lauri Carlson for his supervision, and I thank Jan Engh and Stig Johansson for their helpful comments to preliminary versions of this text. I would also like to thank Maria Fernanda Bacelar do Nascimento for her invitation to participate in this round table and for teaching me a lot about corpora processing and Portuguese.

[1] The corpus from which all examples are drawn is a collection of seven Portuguese short stories by Jorge de Sena (taken from *Novas e Antigas Andanças do Demónio*, 1984) and their translation into (American) English (*By the rivers of Babylon*, ed. by Daphne Patai, 1989), and a novel by John Steinbeck (*The Pearl*, 1975), translated into (European) Portuguese by Mário Dionísio (*A pérola*, 1977).

[2] I am aware that, as usual, there is no agreed upon definition of the term. For example, Baker (1993:149) writes: "in some cases, when an unusual distribution of features is clearly a result of the translator’s inexperience or lack of competence in the target language, this phenomenon is referred to as ’translationese’", while Gellerstam (1986:88) is careful to state that "translationese is not to be equated with translation errors: I use *translationese* in reference to what I mean to be systematic influence on target language (TL) from source language (SL)". Given this situation, I try to give here a precise definition of what the term is meant to denote in this paper. Even though Stig Johansson has called my attention to the fact that nouns in *-ese* have a depreciative connotation in English, viz. *journalese*, I still prefer the term "translationese" to "interference", which in my view (possibly due to Portuguese interference) carries worse connotations still.

[3] I am indebted to Stig Johansson for pointing this difference to me.

[4] For example, one could have a & b <=> a' & b', while a <-> a' and b<-> b'.

[5] All examples in this paper are uniquely meant to illustrate the phenomena specifically discussed in their connection, generally from the tense and aspect domain, presented in boldface. So, for example, I am NOT claiming the contextual equivalence of the subject NP's in the next example.

[6] I should note that, while the argument is my own, I am not claiming to be the first one to argue for its conclusion (or at least a weakened form of it), namely, that translated texts are a useful tool for contrastive analysis. On the contrary, this is perhaps the dominant view, and in fact it explicitly motivates the constitution of parallel corpora; cf. e.g. Johansson & Hofland (1993).