

Chapter 3: Translation

One reason for including a chapter on translation is the simple (though often overlooked) fact that, as Lab (1990:52) put it, having a linguistic description of two languages is not the same as having a linguistic description of the translation among them.

In Section 3.1 I start by discussing machine translation. One reason for this is that it is easier to define and characterize than human translation. Then, in Section 3.2, I describe some linguistic approaches to (human) translation. Afterwards, my discussion of the concept of translation mismatch, in Section 3.3, paves the way for the introduction of a typology of kinds of translation instances in Section 3.4. This section and Sections 3.5 and 3.6, concerning translationese and translation quality respectively, introduce the most relevant issues for the description of translation performance. Finally, the next two sections discuss the use of real translations in contrasting languages, from a theoretical (Section 3.7) and practical point of view (Section 3.8), while Section 3.9 summarizes the main points made about translation and the comparison of languages in this and the previous chapter.

3.1 The contribution of machine translation

Research in machine translation (MT) is relevant for a definition of translation in general, and for the particular question of language comparison for two distinct reasons: It has prompted some to reject the 'meaning preservation across languages' assumption, while, on the other hand, its implementation actually requires explicit encoding of the relation(s) between the two languages.

As to the first issue, some MT researchers have highlighted, to a smaller or larger extent, matters of language differences. For example, Tsujii (1986) claimed that there are meaning distinctions that could only be noticed contrastively; Tsujii (1988) suggested that languages differ not only in syntactic and semantic properties, but in rhetorical structure as well; and Kameyama et al. (1991) noted that there are different distinctions required by different language systems. This last claim attracted a growing interest of the MT community to the so-called translation mismatches, a subject which will be addressed in some detail in Section 3.3 below.

But the issue of the comparison of the two (or more) languages involved is rarely explicit in MT work. Rather, the nature of the translation relation itself has been a hotter topic; see Santos (1992b) for a more extensive survey of what MT researchers have had to say on meaning and translation. Basically, among those who have been discussing translation theoretical issues, one can distinguish between

- a) the approaches based on the principle of meaning preservation in translation (which assume that languages convey the same meaning); and
- b) the approaches focussing on a weaker definition of the translation relation.

On a different level altogether, another contribution of the MT community is often overlooked: MT developers, by actually developing translation machines with limited or no resources as to

semantics, have demonstrated that much of the content of an utterance is related to its structure and lexicon. In that way, relatively simple formal translation relations hold between the source and target languages, so that, even without explicitly handling meaning, they manage to preserve it fairly often. Inspired by the discussion in Schubert (1988), one may even conclude that most successes of MT were due to the sharing of implicit information across languages. Schubert also makes the controversial claim that it is not "really necessary for machine translation to make semantic relations explicit" (Schubert, 1988:600). Rather, they can remain implicit, provided one uses a human language as intermediate representation. Dyvik (1994:17) also mentions "*implicative relations* holding among aspects of the translational relation", as a reason why translation among closely related languages is easier: "as long as we can trust such implicative relations we know that we do not have to worry about recreating more than the properties that imply all the others -- the rest will follow". This should be an antidote to excessively radical conclusions that might follow from the rest of this chapter.

Returning to the question of language differences, let me start with those researchers who took the standpoint that languages convey exactly the same meaning. For example, Landsbergen states that in the ROSETTA system "a sentence s' of L' is considered a possible translation of s of L if s and s' are derived from corresponding basic expressions by applying corresponding syntactic rules" (Landsbergen, 1987:353), i.e., he requires a strong equivalence between languages. Similarly, Beaven and Whitelock's (1988) MT system uses the notion of bilingual sign, where the source and target semantics are constrained to be shared.¹

These assumptions are not without theoretical consequences, though. For example, Landsbergen (1987) actually defined meaning in one language depending on the way it was conveyed in the other language, by requiring an isomorphy between compositional meaning construction in the two languages. Thus, he modelled the translation relation as mere identity of meanings, but at the cost of giving up the concept of meaning in one language.

On the other hand, Beaven & Whitelock (1988) and Dyvik (1990) formalized the translation relation by unification of meanings, the first using first order predicate calculus, and the second situation schemata, to represent the semantics. In order for two descriptions to unify, however, the same metalanguage (or semantic representation language) has to be used -- and thus, in the example presented by Beaver & Whitelock, we see that the semantics of the Spanish distinction between *perna* and *pata* is cast in terms of the "interlingual" concepts LEG-OF and HUMAN. Analogously, Dyvik (1994) uses the concept "sense" EAT' to represent both the English and Norwegian "related" verbs *spise* and *eat*. To avoid such commitments, I have argued in Santos (1992b) for the use of a different metalanguage, a subject which I shall develop at length in Chapter 5 below.

Other researchers have actually answered the question as to what the translation relation is in a different way: Kaplan et al. (1989) suggested an equational definition of the notion of translation, namely that the target structure must be subsumed by a minimal solution to the transfer description. If

¹ This sharing is additionally constrained by the requirements: 1) that there is not a more specific sign in either language that also unifies with that semantic description, and 2) that, after the full derivation, unification still holds.

one assumes that such transfer description can involve meaning changes as well as structure changes (only the latter were exemplified in the paper), they in fact place no constraints on what can be related by translation. An interesting property of their proposal, mentioned in Kaplan et al. (1989:276), is that the transfer equation allows for more than one specification in the target language concerning the cases where the source language is vague, i.e., a unique solution of the transfer equation is not required.

The field of translation studies accepts that "Currently it is acknowledged that translation is an operation, which implies a dynamic process and static results; the term translation itself is ambiguous on that matter" (Garnier, 1994:79, my translation). Several other researchers have accordingly defined translation in a more dynamic way: Hobbs & Kameyama (1991:141) claimed that "MT requires that the appropriate inferences be drawn and that the text be interpreted to some depth", and they suggested abduction as a computational method of actually performing translation.

Finally, Kameyama et al. (1991) defined translation as information flow, which can invoke subsumption, assumption, and merger of concepts in a lattice, where, in addition to content and context (discourse specific and utterance specific), also contrastive linguistic information is present (of both the lexical and grammatical kinds).

My conclusion is that it is clear that, even without taking into account the problems of human creativity and of the translator's choice, the task of analysing and formalizing translation (of sentences) requires more complex models than those who were involved in the first attempts at automating translation had realized.

3.2 Linguistic theories of translation

I will now turn to linguistically motivated definitions of translation, without, of course, attempting to cover the field exhaustively.² Rather, I will look at translation mostly from the point of view of bridging two language systems, i.e., from a linguistic point of view. In fact, I believe that translation is a linguistic phenomenon, and should therefore be primarily viewed in that way: After all, translation is the form of natural language communication which is more constrained or circumscribed by language and less by the external world (it is the production of a text from a primary source which is a written text, and often the translator has not even access to the extralinguistic context present at the writing of the original).³ So, all claims that translation is primarily cultural, for example, are only true to the extent that natural language itself is seen as a primarily cultural phenomenon.

In his seminal book on translation, Catford (1967) describes the problems of translation very clearly, contending that:

1. Translatability is a question of degree: "SL [source language] texts and items are *more or less* translatable rather than absolutely *translatable* or *untranslatable*" (Catford, 1967:93).

² Good discussions on translation and the field of translation studies can be found e.g. in Brower (1959), Bassnet-McGuire (1980) and Gutt (1991).

³ This point will be taken up again in Chapter 4 below.

2. There is a clear distinction between the two concepts of transference and translation: Transference is a (not very common) process whereby "parts of the TL [target language] text, do have *values set up in the SL*: in other words, have *SL meanings* [... while] in normal translation [...] the TL text has a *TL meaning*" (Catford, 1967:43).

3. There are more often than not translation shifts, i.e., the information conveyed by one level⁴ of the linguistic system in one language is rendered by another level in the other language. Therefore, a rank-bound translation, i.e., a translation where a limited number of linguistic levels are used (such as word-bound, or phrase-bound), is necessarily defective in some cases, as opposed to total translation.

4. "for translation equivalence to occur, both SL and TL text must be relatable to the functionally relevant features of the situation" (Catford, 1967:94). When there is no way to build a functionally relevant feature in the context of the TL text, untranslatability arises.

The distinction between functionally-relevant distinctions and those forced by the language system (linguistically-relevant) is of major importance: while some contrasts are intended, others are forced by the SL language. Only functionally relevant distinctions should (be attempted to) be preserved in translation. For example, in *Je suis arrivée - dit Marie*, gender marking is obligatory in French but is not intended, nor necessary, for the understanding of the text. Therefore, it can be left out in the translation into any language where gender marking is not obligatory. However, were gender functionally necessary, and not possible to specify in the target language, the text would be untranslatable.

I should stress that functional relevance is a property of linguistic items in context, and this is possibly why it is so difficult to define functional relevance in a precise way, as Catford acknowledges.

5. Untranslatability can be due to linguistic or cultural factors. Linguistic untranslatability is related to formal (i.e., linguistic) properties, and is best exemplified by puns where ambiguity or polysemy play a main role. "Cultural untranslatability arises when a situational feature, functionally relevant for the SL text, is completely absent from the culture of which the TL is a part" (Catford, 1967:99). The examples here are not so clear-cut, and the translations lead to what Catford calls "collocational shock": (*they lay in the*) *bathroom* for Finnish *sauna*, or (*hotel*) *bath-robe* for Japanese *yukata*.

Another interesting theory of translation is put forward by Gutt (1991), who, however, does not cite Catford. Gutt argues against the idiomatic and the dynamic equivalence approaches to translation, represented by Beekman and Callow (1974)⁵ and Nida and Taber (1969)⁶ respectively, on

⁴ Catford distinguishes between shifts between grammar and lexis, called level shifts, and shifts inside grammar proper, namely structure shifts, class shifts, unit shifts, etc., which he calls category shifts. For my purpose, this distinction is irrelevant, and I use the word "level" corresponding to both his level and category.

⁵ "whose primary concern is that the relationship between receptor and message should be substantially the same as that which existed between the original receptor and the message" (Gutt, 1991:66).

⁶ which emphasizes "'naturalness' of language use and ease of comprehension rather than receptor response" (Gutt, 1991:66).

the grounds that they both produce inadequate analyses in some cases and do not furnish principles to decide how to actually translate. Gutt suggests an explanatory theory of translation, as opposed to a descriptive or prescriptive one, and bases himself on a theory of language behaviour, Sperber and Wilson's (1986) relevance theory. One of his claims is precisely that "relevance theory alone is adequate -- there seems to be no need for a distinct general translation theory" (Gutt, 1991:viii).

He thus models translation as "interlingual interpretative use", a relevance theoretic concept that is defined roughly as "two utterances interpretively resemble each other iff they produce the same effects on the listeners uttered in the same context" (Gutt, 1991:44), or, if two different contexts are at stake, if they share the implicatures and explicatures. In addition, in cases where the original style matters, they still preserve (as much as possible) the communicative clues of the original, or resemblance in linguistic properties; cf. "the point of preserving stylistic properties lies not in their intrinsic values, but rather in the fact that they provide clues that guide the audience to the interpretation intended by the communicator" (Gutt, 1991:127).

Gutt's emphasis is, however, on the claim that a translated text is a communicative object, and thus should be created with the intended audience in mind, and he is very definite in stating that "the aim of conveying the same message does not provide a tenable basis for a theory of translation" (Gutt, 1991:99).

Despite the overall interest of his book, my impression is that many of his arguments are not so much concerned with languages but with cultures: most of the time, what he is discussing is cross-cultural (un)translatability, which falls outside of what I would call translation proper. I believe that most of the cultural mismatches he mentions could happen using only one language, namely, the problems of lack of relevance of a text for a particular culture, lack of previous knowledge for a second culture, and of excessive wrong implications in the the case of a third, which lead Gutt to conclude that one not only has to choose "how to say it" but actually "what to say". In any case, the problems do not have to do with the specific pair of languages involved in any linguistically relevant way - and this is why, I believe, he could manage with relevance theory alone.

3.3 On translation mismatches

To the man in the street, a "translation mismatch" is a place where a (particular) translation went wrong, where the target text does not convey what the source text conveyed. However, the term has acquired a technical meaning with Kameyama et al.'s (1991) influential paper, a meaning which I believe to be inadequate in several ways. I will discuss it at length, because a correct classification of translation pairs is crucial for this dissertation.

The original definition was: "TRANSLATION MISMATCHES [...] are found where the grammar of one language does not make a distinction required by the grammar of the other language" (Kameyama et al., 1991:194). Consequently, translation mismatches are generally held to involve the following two situations (cf. Kameyama et al. (1991), Eberle et al. (1992), Barnett et al. (1994)):

a. A particular distinction is made through linguistic means in the target language and not in the source language. So, the target language is forced to decide.

b. A particular distinction is made through linguistic means in the source language and not in the target. So, the target language eliminates the distinction.

Let me start by pin-pointing a vagueness of the term with respect to the type/token distinction: it seems to refer both to actual cases of translation problems (tokens), which derive from general differences in language systems, and to these general differences themselves (types). The term is thus related, on the one hand, to the assessment of particular translation pairs, and, on the other hand, to a specific kind of language difference, which can be held to constitute a problem for translation in general, prior to obtaining any translation pairs whatsoever.

The first definition in the machine translation literature seems to invoke the two questions: While the statement that "divergences are instances of lexical mismatches [...] resolved within a sentence by cooccurring lexemes" (Kameyama et al., 1991:194f) can only refer to actual translation pairs, the explicit introduction of the concept, quoted above, seems to point in the other direction, by mentioning grammar (systems).

Of course, one could claim that translation mismatches are precisely those cases of translation problems which are derived from this kind of language differences. But this definition is, on closer examination, rather unsatisfactory from a theoretical point of view. In fact, I will show that it singles out a set of cases that is not principled by any of the two possible approaches:

If one is interested in a typology of actually occurring translation pairs, there are other cases of (translation) mismatches - and very much so. The so-called translation mismatches are at most the limiting cases of an array of different possibilities, as will become clear in Section 3.4.2 below.

On the other hand, if one is interested in actual insurmountable *a priori* problems for translation, not all (and in fact only a small minority) of these so-called translation mismatches result in real problems, since the mismatches are defined irrespective of their functional relevance (à la Catford).⁷ In fact, a crucial property of the concept of translation mismatches is that they are defined in comparing languages, and not texts, and therefore no reference to the context is possible.

What have been called translation mismatches (in the type interpretation) would be better described by the name "Language Mismatches Relevant for Machine Translation", since neither do they necessarily bring about translation problems nor are they necessarily found (or defined) in actual translations. Their existence is an issue in the comparison of languages, but not primarily in translation (except in so far as all contrastive knowledge is relevant to translation). At this point, one should note that it is relevant for machine translation, though, where it must be addressed generally.

I should actually stress the fact that the whole concept of translation mismatch originated in the context of machine translation. Some explanation for this lies in the fact, noted by Dyvik (1994), that the building of an MT system requires encoding translation competence, not performance. The ultimate goal of MT researchers is to render the translation choice automatic, through specifying in advance the possible translation choices (which will then be chosen, on the fly, given the context). While it is true that there is no analogue to free translation here, or, at least, that the translation pairs

⁷ This point is also made by Bar-On (1993), explicitly quoted in Section 3.4.1 below.

resulting from machine translation will be a subset of those found in human translation, it is equally true that the requirement of full specification of the links between the two language systems causes problems which are not pertinent when assessing, studying, or analysing real human translations.

Now that a separation between the question of incommensurable grammatical systems and that of the analysis of actual translations has been established, let us look more closely into the latter.

If one looks at actual translations in context, one may be able to observe, in fact, some cases where the language difference resulted in some translation loss. But much more conspicuous will be the existence, in parallel with cases a. and b. above, of many cases where

c. (a particular text in) one language makes a distinction that is not explicit in the other, even though it could have made it explicit.

Or, in other words, languages differ not so much as to what they can say, but in what they do (incidentally, it appears that I claim the contrary of Jakobson's famous quote: "Languages differ essentially in what they must convey and not in what they may convey" (Jakobson, 1959:236)).⁸ And it was this observation that made me look for a more appropriate typology for the description of real translation pairs.

But, first, it is necessary to discuss what exactly is meant by "translation pair".

The term "translation mismatch", discussed above, was introduced in contrast to the term "translation divergence" in Kameyama et al. (1991), "divergence" denoting the cases where translation equivalence is achieved through level shifts below the sentence rank, while a "mismatch" was considered to occur when translation equivalence did not obtain at the sentence rank (or at no rank whatsoever).

This definition requires a specific answer to the question: in translation, at what level can one meaningfully talk about equivalence (or about meaning), and, consequently, use the expression "translation pair"? Only at the level of the sentence and above, or at any linguistic level? Or, as e.g. Givón (1978:272) contends, at the level of discourse (i.e., over and above sentences)? The answer to this question conditions the attitude towards the dichotomy mismatch/divergence. If one can only talk about equivalence at the sentence level, then the notion of divergence is simply not defined, since *strictu sensu* there is no translation between parts of the sentence. On the other hand, if one treats all linguistic levels as able to stand in the relation of 'translationally related' (as is done in standard translation practice), then there seems to be little reason to separate the cases above and below the sentence by using different names.

Rather, it would make more sense to distinguish the cases where (no matter which level) there is preservation of (enough) content (or information value) from the cases where the target language text fails to convey significant parts of the source text. In Santos (1994a), I called the latter

⁸ Strictly speaking, my statement is but a weaker version of Jakobson's, with *do* substituted for *must*, not the contrary. But I read Jakobson's statement as implying that languages (for example, in translation) will be able to say very much the same thing, or even that they will say very much the same thing, whereas in my statement I want to express precisely the opposite conclusion, i.e., even though they may be able to say the same thing, they won't. Even in a constrained context like translation, different languages will not in general convey the same information.

"semantic" mismatches, and the former could be called merely "syntactic", dropping the notion of divergence altogether.⁹ Syntactic mismatches are thus strictly speaking all instances of translation which do not result in failure to convey the source message, since every syntactic device is (in a way) language specific. It is, however, not obvious in all cases how to separate syntactic from semantic mismatches, especially when dealing with a translation level above the sentence, and so I may use the cover term "translation mismatches" applied to any non-trivial translation relation.

The detailed study of actual texts convinced me, however, that a twofold distinction (between syntactic and semantic mismatches) was not sufficient to account for real translations, and that the variety of cases found required finer discrimination.

3.4 A typology of translation pairs

Before presenting my own classification in Section 3.4.2, I will survey some typologies of translation mismatches which have appeared in the literature, in Section 3.4.1.

3.4.1 Previous classifications

Keenan's (1978) article, although concerned with the theoretical question of (un)translatability and not with the performance of actual translators, has deeply influenced my way of thinking, as described in the previous chapter. In his paper, Keenan argues that languages differ in what respects they are vague and in what respects they are precise, and this is why exact translation is often impossible. He gives the following definition of vagueness: "if a speaker can remain indifferent between alternatives *a* and *b* and still meaningfully assert some sentence, then the sentence is vague according as *a* or *b* obtains" (Keenan, 1978:173). An interesting consequence of the difference between ambiguity vs. vagueness, as pointed out by Keenan, is that in cases where the target language has no corresponding ambiguous or vague device, in the former case a translation can be exact, while in the latter it will always be inexact.¹⁰

Keenan presents a considerable number of language phenomena that illustrate "semantic properties which are overtly represented in the syntactic structure of the sentences in question" (Keenan, 1978:167) in one language but not in others, namely i) semantic gaps, ii) pragmatic presuppositions, iii) structure destroying operations, and iv) positive and negative co-reference. It can thus be said that Keenan's typology is arranged according to semantic phenomena.

Bar-On's (1993) analysis of mismatches, directly influenced by Keenan's, is relevant because it distinguishes between competence and performance, i.e., she separates possible from actual translation problems: "Our examples of translatability failures have mostly traded on the fact that in some contexts idiosyncratic features were relevant to the understanding of an SL item [...]. We have

⁹ To be fair, I think that syntactic mismatches were precisely what Dorr (1990) wanted to express when she introduced the term "divergence"; however, the lack of a sufficiently clear definition, and Kameyama et al.'s (1991) contrast with "mismatch", makes it advisable to drop the term.

¹⁰ In other words, if a device is ambiguous, it may be disambiguated in context, which makes it possible that a non-ambiguous TL device is an exact translation of it. On the other hand, if a device is vague, no choice in the TL can be an exact translation. This last assertion will be discussed in many places in the present dissertation.

not ruled out the possibility that in other contexts those features would not be relevant, so that exact translation would become possible" (Bar-On, 1993:792). This is certainly something one should bear in mind when talking about such issues as translatability and translation quality. To use Catford's terms again, systematic differences in language systems are not (for the most part) functionally relevant (see Section 3.2 above) in a particular text.

Bar-On's typology, in terms of linguistic features, is however of considerably less interest. She considers

a) lexical referential mismatches: "if a designator stands for idiosyncratic aspects (or elements) of the extra-linguistic environment of the SL speakers -- whether natural or socio-cultural -- which are completely missing from the environment of the TL speakers" (Bar-On, 1993:787).

b) lexical linguistic mismatches: which "occur when the relevant elements [...] are [...] compressed by a single designator but not in the other" (Bar-On, 1993:787), and are basically MT's concept of lexical gaps (see Santos, 1990).

c) grammatical mismatches: exemplified by the Malagasy sentence glossed as "the table which was basket-put by Rabe was damaged", underspecified with respect to location of basket in relation to table, and English *I have just broken up with my lover*, unspecified for sex of both *I* and *my lover*.

d) pragmatic mismatches: if the mismatch "occurs at the level of pragmatics" (Bar-On, 1993:791)

In spite of furnishing interesting examples, it is not evident what Bar-On's use of linguistic levels (lexical-grammatical-pragmatic) attempts to accomplish. For one thing, the classes defined are not necessarily disjoint;¹¹ in addition, she does not even provide a definition of grammatical or pragmatic mismatches. For example, why is the French *tu/vous* distinction considered pragmatic and not grammatical, when it is both grammatically encoded and as obligatory as gender encoding in Hebrew, which she presents as a grammatical mismatch?¹²

Nida's (1959:22ff) classification is more detailed, and more interesting as well. While still comparing languages as systems, Nida differentiates translation problems, as faced by translators, according to their causes. Here is my rendering of it:

a) the source language lacks information which is obligatory in the target language

b) the source language obscures information which is obligatory in the target language

c) the source language is ambiguous¹³ regarding information which is obligatory in the target language

d) the source language conveys implicitly information which is obligatory in the target language

e) the source language conveys explicitly information which requires a heavy apparatus when it is rendered in the target language

¹¹ This was pointed to me by Lauri Carlson (p.c.).

¹² True, natural gender has non-overlapping domains, while this is not so in address conventions: At least in some cases, the speaker has the choice. In most of the cases, however, the social relationship between speakers, or their sex, is out of their control, and is clearly extralinguistic.

¹³ In fact, I think that it should be termed "vague" instead, but I am using Nida's terms.

f) the source language conveys explicitly information which is optional in the target language

Looking critically at Nida's distinctions, "lacks" and "obscures" are not really different things, the latter being just a degree of lack of information, and thus a) and b) boil down to the same. On the other hand, I split Nida's E category, namely, "cases where explicit information should not be rendered explicitly", into points e) and f) since, as Nida himself states, they are clearly different: They are "two exceptions to this general rule [explicit information in the source language should be communicated in the receptor language]" (Nida, 1959:24). It is due to this general rule (contested, by the way, by Gutt) that Nida's description of problems for translation is heavily geared to cases of having to add information in the translation which is not explicit in the source, rather than cases of having to dismiss explicit information.

An interesting point made by Nida is that implicitness can be very culturally dependent, in fact, more so than is actually in the text, and, therefore, the translator must sometimes add explicit information to prevent unwanted "implicit" inferences to creep in. This brings about a new category which is not covered by d), or, alternatively, would cause d) to be rephrased as "making explicit something that is only implicit in the source language, either because of obligatoriness, or because of the need to prevent unwanted interpretations".

It is nevertheless important to note that Nida's classification is not one of resulting translations, but of difficulties for the translator. This is most conspicuous in cases e) and f), whose proposed solution requires a global (macro-level) perspective: using a natural frequency of the corresponding target language devices is not a property of individual translation pairs, but a global translation strategy.

This leads me back to my statement in Chapter 1: Translations can be (orthogonally) evaluated relative to their closeness to the original text, and relative to their naturalness (closeness to the target language style); cf. Toury. The first issue can be tackled by looking at individual translation pairs, while the second, measuring what is typical in a system, can only be dealt with by comparing individual cases to a prototype computed from a large set of instances.

In the following section, I will present an alternative typology, whose main purpose is to accurately describe real translation pairs. Contrary to the typologies described in this section, it will not be organized according to cause.

3.4.2 A classification according to semantic relation

As I am concerned with all translation pairs (and not only those that constitute, for one reason or other, translation problems), I intend to describe all possible kinds of translation pairs, not only the problematic ones. A similar approach, concerning syntax, has actually been pursued by Dyvik, Wollin and Platzack. Dyvik (1993) classifies translations according to a fourfold scale: word for word correspondence; content word correspondence only; predicate-argument-structure correspondence only; and other; while Platzack (1983) uses eight categories based on Wollin's (1981) even more detailed scale: structural identity; transposition; functional adjustment; addition; deletion; divergence; convergence; and mixing.

Inspecting a number of actual human translations, I was able to arrive at the following pre-theoretical typology (the adjective "pre-theoretical" relates 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 preservation, or, rather, the semantic 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 they may be hard to distinguish in practice:

A1. Semantic equivalence. Through different linguistic means, the two elements of a translation pair say "exactly the same thing", although in two different linguistic systems.¹⁴ These cases are harder to find than one might expect. One 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:¹⁵

*the stinging pain of the bite **was going away**
a dor da picada **diminuí**
'the pain of the sting diminished'*

The English sentence describes a situation with a defined goal, namely the disappearance of pain, and states that a process leading eventually to such a goal is in process. The Portuguese sentence, in turn, describes a gradual process which is occurring. That process may eventually lead to total disappearance, but that is not inherent in 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

¹⁴ For example, one could have a & b \Leftrightarrow a' & b', while a \nrightarrow a' and b \nrightarrow b'.

¹⁵ These examples 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.

relevance of the A1 category in itself.

Case B: Subsumption. By subsumption, 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 which is 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 information preservation is possible 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, as Lauri Carlson (p.c.) has pointed out to me, *bury* is also used with the figurative meaning "hide" (cf. COBUILD, p.187, II). If this is not the case with *enterrar* (which I believe 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 argue in this case that actually the pearl was buried in the sense of "hidden underground", because this 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.

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

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

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

*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 source of the death that prowled it'

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).

It is relevant to distinguish this case from proper subsumption, in my view, because 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 that present in the source language is displayed in the translation. Actually, addition is very common. It may be required by the target language or triggered by personal choice. In general, it 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 the one actually expressed by the original text; a situation which is nevertheless conveyed by the target text. I will call this C1. Reconstruction. 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 that the lamp made'

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 C2. Free addition, 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. Note the following examples: 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 to be 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 is "free". This seems to indicate that a fine classification of translation pairs must also take into account the two language systems involved.

In fact, one especially 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 (one of the two cases of 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 all of his people had always been'

Portuguese has to choose either remote or cooccurrent 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, and will be abundantly described in the present thesis. Other such cases between English and Portuguese involve 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 possibility, suggested 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 -> Não tinha um tostão ('He did not have a penny')

The man was hanged -> HE WAS EXECUTED -> O homem foi decapitado ('The man was beheaded')

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). They involve both addition and loss of information,

i.e., they combine cases B. and C. Note, however, that while the first corresponds to positing a situation which entails the two others, and could be termed D1. Entailments of a common situation, in the second case both are specific instances of one common situation, i.e., this could be analysed as D2. Sharing entailments.

A very common case of D1 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.'

In purely formal terms, any translation pair could be described as having an underlying intermediary representation (built by adding the two statements) from which both sides could be entailed. Obviously, something else is required to justify invoking a "common situation" in the first place.

This paves the way for 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 is defined here as the case where none of the semantic relationships A to C between the source and the translation described holds, and the (formally possible) version of D has not enough internal coherence. Obviously, and as will be additionally shown below during my discussion of translation quality, translation failure is a difficult notion. How great must a discrepancy be in order not to count as addition of information, partial interpretation, or parallel (in the sense of sister node) rendering, but rather as translation failure *tout court*? In general, it is perhaps better to speak of relative translation failure.

Translation failure, moreover, is not a random process.¹⁶ It can be attributed mainly to two different reasons:

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

E2. Interference. Translation failure may occur 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.

¹⁶ Forgetting for the moment the uninteresting question of typing errors.

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

Summing up, the above typology presents a classification based on several principles. Basically, translation pairs were organized according to a logical relation enriched with linguistic considerations. For the logician, the relevant categories ('=>' stands for implication) are then: A: $a \Leftrightarrow b$; B: $a \Rightarrow b$; C: $a \Leftarrow b$; D1: $c \Rightarrow a$ and $c \Rightarrow b$; D2: $a \Rightarrow c$ and $b \Rightarrow c$, and E: none of the above. Further distinctions drawn are relevant if one takes the linguistic systems into account: A1 versus A2 distinguishes between the meaning of linguistic categories in context and in vacuum;¹⁷ B1 versus B2 deals with further specification of the linguistic category *a*, specifically, from the assumption that *a* could be identified roughly as $c+b$, whether the nature of that "+" was essential or arbitrary; C1 versus C2 concerns essentially the same distinction, now having *b* identified as $a+d$; while E1 contrasts with E2 by depicting the cause for E obtained through two different linguistic levels (content versus form).

Before discussing some consequences of this typology for translation-based contrastive studies, let me address two other issues that must be mentioned before a conclusion can be drawn: translation quality and translationese. Although both make sense in the context of machine translation as well, they will mainly be discussed in relation to human translation, the object of this study.

3.5 Translationese

Translation, as already mentioned, 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.¹⁸

¹⁷ This is a distinction that cuts across all cases, but which was only identified separately for case A. In the other cases, the analyses already presupposed translations in context.

¹⁸ 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 take 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 dissertation. Even though Stig Johansson has called my attention to

On 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 typicality.

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 explain why such differences occurred at all. In this dissertation, I will be interested in the notion of translationese applied to grammatical features. (Incidentally, 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 (1994b) -- see also Chapter 10.)

Gellerstam, however, concentrated on target language texts only. In Santos (1995a), I suggested a broadening of the notion of translationese by analysing parallel corpora, and inspecting and using as evidence the source text and the actual translation relations as well,¹⁹ integrated in a general program of identifying the differences and similarities of two languages.

I present here some real examples where translationese is, in my opinion, obvious:

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.

'Leaning on the table, he dragged himself to the other corner and fell 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.

'She kept lying, savoring a discomfort that was 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.

In Part III, several other examples will be discussed, while some procedures to systematically detect grammatical translationese, building upon those suggested in Santos (1995b) will be presented in Section 3.8 below.

3.6 Translation quality

Everyone is aware that translators sometimes fail, and that translations can be rated (if not, why

the fact that nouns in *-ese* have a depreciative connotation in English, e.g. *journalese*, I still prefer the term "translationese" to "interference", which in my view (possibly due to Portuguese interference) carries worse connotations still.

¹⁹ I am indebted to Stig Johansson for pointing out this difference between Gellerstam's approach and mine.

would "translation prizes" exist for the best translation of the year in most European countries?). Translators themselves are willing to admit that "Even with such cautious checking and re-checking, errors will persist" (O'Brien, 1959:89).

Again, translation quality can be seen as referring both to content preservation, and to naturalness of the resulting text. But even separating these two criteria, it is hard to come up with a reasonably objective strategy, as books devoted to the task of "translation quality assessment", like e.g. House (1981), show.

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, scatched and hurt from the stones and brushes'

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 distinguíam 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, exhausted, 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 the examples above it is difficult to point out

translation correspondences between elements of the translation pairs (one might call this "non-compositionality"), but none displays a blatant error, either.

If one tries to go deep into what is supposed to be preserved, it becomes obvious at once that most translations, in fact, fail to convey every piece of information, and only those, present in the source text. I hope to have shown this clearly in Section 3.4.2. 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.

One of the most complicated 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 paraphrasability, an important issue concerns the choice among different pieces of information contained in a source text, as was again pointed out in Section 3.4.2.

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 has 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). Conventionalized implicatures, as should be expected, vary from language to language, thus making the translation of grammatical features especially difficult. As regards Portuguese tense and aspect markers, I have given a first description of clusters of such meanings of Imperfeito in Santos (1994b) and presented some analyses of their translation in Santos (1995b), which will be built upon in Chapters 10 and 14, respectively.

Summing up, in spite of the optimistic statements of some computational linguists, e.g. that "existing translations contain more solutions to more translation problems than any other available resource" (Isabelle et al., 1993:205), real translations display a fairly complex array of problems as well.

3.7 Discussion

Before addressing the issue of the use of translations for the study of language contrast, which is, after all, the *raison d'être* of the present and the preceding chapters, I want to draw some conclusions from the three previous sections, as well as evaluate what I have done so far.

I started with a critical presentation of the interrelationship(s) between meaning and translation, from the perspective of both the machine translation and the translation studies schools. After discussing in some detail a particular concept from the MT community, the translation mismatch, showing that it fell short of describing most mismatches in real translation, I argued that translation is not meaning preserving in general. I did this by presenting a semantic typology built upon the study of actual translation pairs which featured several different relationships between the meanings of the elements of the pair.

This, I feel, is the practical, or empirically based, complement of Santos (1992b), where I made the same point (lack of meaning preservation in translation) without resorting to real data.

(Incidentally, there, I cited philosophical approaches to meaning rather than claims from translation theorists, and discussed other possible relations that could be proposed to formalize translation.) I believe, nevertheless, that the argument presented here is considerably stronger, because it is based not on speculation but on real, documented, translation practice.

Then, I proceeded to discuss two other questions: the elusive issue of translation quality, and the phenomenon, more amenable to description, called translationese. My intention was to provide an informed overview of what may be involved in any particular translation pair.

Now, it is undeniable that the classification I provide is considerably more intricate than the one obtained by a simple generalization of the "translation mismatches" perspective (adding case c. - - even though the two languages can, they do not express the same content (cf. page 30) -- and d. -- they can and do express the same content). The reason lies, I believe, not in the failure of the researchers to address the relevant questions, but in the realization that the (computational) analysis of real translations is a task different in kind from, and possibly more complex than, machine translation.

I.e., although the notion of "translation mismatch" is relevant to the design of an MT system, it can nevertheless be of little value for the analysis of real (human) translations. Conversely, measuring translation quality or translationese is outside the scope of an MT system, which strives to render only good (though possibly translationese-prone) translations. Also, a little consideration of the issues described in C2, D or E of my typology in Section 3.4.2 shows that they cannot be involved in MT.

This, it is important to note, disagrees radically with some (far too optimistic, in my opinion) claims in the MT literature. For example, Isabelle asserts that "there are reasons why the outlook for bi-text production is much brighter than it is for MT systems. First, in contrast with the active linguistic capacity required for the production of translations, the reconstruction of translation correspondences in existing translations requires only a passive linguistic capability" (Isabelle, 1992:80). He is obviously not taking into consideration human imagination, creativity and other hard-to-formalize phenomena which are at play in human translation, allowing it to diverge/deviate in a much more radical way than machine translations.

This misconception of simplicity is well phrased again by Isabelle, when he suggests that it is the system developer's choice, after sentence alignment, to "further specify correspondences between lower-rank units: sentences, phrases, words and morphemes" (Isabelle, 1992:80). This, I contend, is not a question of choice, but of descriptive adequacy. The following examples demonstrate, anticipating the description of alignment in Chapter 9, that it is the texts to be aligned that dictate which resolution is in fact possible, i.e., a specific bi-text may allow some sentences to be aligned in a word-to-word fashion, while other sentences can only be sentence-to-sentence aligned. I have therefore underlined minimal correspondences in some translation pairs:

And, since early Mass was over and business was slow, they followed the procession, these endless searchers after perfect knowledge of their fellow men, to see what the fat lazy doctor would do about an indigent baby with a scorpion bite.

e, como a missa da manhã estava acabada e aquilo já pouco rendia, juntaram-se à procissão como incansáveis pesquisadores do perfeito conhecimento do próximo, para verem o que aquele gordo e preguiçoso médico faria por uma criança pobre picada por um lacrau.

'and since the morning mass was finished and that already little payed, joined the procession as untirable searchers after the perfect knowledge of the fellow man, to see what that fat and lazy doctor would do for a poor child bit by a scorpion.'

And the baby was weary and petulant, and he cried softly until Juana gave him her breast, and then he gurgled and clucked against her.

A criança estava fatigada e agitada, e só deixou de chorar quando ela lhe mostrou o seio, a que se lançou com avidez.

'The child was tired and agitated, and only stopped crying when she showed him the breast, to which he moved with greed.'

(Digressing for a moment, I note that Isabelle displays a fairly simplistic view of (machine) translation when he states that "one cannot translate a unit (say a sentence) without at the same time translating its components (phrases, words)" (Isabelle, 1992:80). One can translate a unit without at the same time translating its components, not only in human translation, as shown above, but in machine translation as well. Santos (1990) describes an implementation that demonstrates this claim.)

But let me leave aside for the moment the question of machine versus human translation, and concentrate simply on what other conclusions can be drawn from the typology of Section 3.4.2.

It became evident -- see Chapter 7 for extensive illustration -- that, even when there were no formal marks requiring specification in one system but not in the other (the classical definition of translation mismatch), more often than not a situation other than equivalence (case A) occurred.

One way to go about it would be to attribute it to the unsystematicity of the lexicon. In other words, most translation mismatches would be due to lexical mismatches, caused by idiosyncratic lexemes in either language. Not denying that there are idiosyncratic lexical items, I believe that lexical organization is by and large related to (or, in harmony with) grammar. And hence most lexical mismatches in the realm of tense and aspect should be derivable from a lexical organization which is dependent on the same principles of grammar. They should not be random.

Another alternative that has been proposed is to postulate quantitative rules as formalizing tendencies: Instead of simple prohibition or obligatoriness, one should introduce plausibility indices in grammar description. I believe now that this position stems from a confusion between the empirical data and the hypotheses to explain it. While a different distribution should be noted as an empirical fact, a systematic description should not involve quantitative factors in order to have explanatory power. In my opinion, a systematic description should rank categories and distinctions among categories, and will predict that the lower an item is hierarchically, the less frequent (more marked) it can be expected to be.

I will, therefore, attempt in this thesis to describe systematically the differences between English and Portuguese -- which can be objectively assessed by different distributions -- by postulating different categories and/or different ranking order among corresponding categories, not by introducing numerical "explanations".

3.8 On contrasting languages through translation

In the previous sections, I have surveyed the difficulties in coming to grips with the notion of translation, and the complexity involved in an adequate description of its result in semantic terms. Here, I will discuss how a text and its translation can be used in the comparison of two language systems.

3.8.1 The argument for translation-based contrastive studies²⁰

At first sight, there is not very much positive one can say in favour of the use of translations in language comparison. When one seriously addresses the question of eliciting semantic data from parallel corpora, 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 3.4.2 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 3.5, 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 (1994).

These two questions do no more than re-state the already mentioned 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 commonly 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, "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" (van Buren, 1990:85)

On the other hand, as I argued in Chapter 2 above, there is no guarantee that underlying semantic categories in any two languages cover the same range, nor is it necessary that they are

²⁰ 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 compilation of parallel corpora; cf. e.g. Johansson & Hofland (1994).

involved in the same situations.

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 measuring one language via the other language's categories. This is, in fact, what people have always (implicitly) done, often using English as the measuring rod (as noted in Santos (1992b) and above in Chapter 2).

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 in fact turns out to be the best unbiased way (if not the only one) to get objective data for the contrasting 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 guide-lines 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 some considerations 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

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 insufficient 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).

3.8.2 Practical contrastive studies based on parallel corpora

I have argued in the previous section that corpus-based studies on real translations are the best method to elicit systematic language differences.

But it is one thing to argue on theoretical grounds for the adequacy of a general approach. It is another matter, involving quite different concepts altogether,²¹ to demonstrate that it is practically possible, by suggesting an algorithm, or procedure, to do it.

Having shown that many complex phenomena are at stake, I will not present a full-fledged algorithm. I will, rather, concentrate on the identification of plausible cases of interest, assuming that, as everywhere else in scientific (empirical) investigation, data and hypotheses are intertwined in the process of constructing a scientific theory; cf. Hempel (1966).

The investigation process should thus be adequately describable as two activities in tandem: the

²¹ Cf. Hempel (1966:14): "scientific hypotheses and theories are usually couched in terms that do not occur at all in the description of the empirical findings on which they rest, and which they serve to explain. For example, theories about the atomic and subatomic structure of matter contain terms such as 'atom', 'electron', 'proton', 'neutron', 'psi-function', etc.; yet they are based on laboratory findings about the spectra of various gases, tracks in cloud and bubble chambers, quantitative aspects of chemical reactions, and so forth -- all of which can be described without the use of those "theoretical terms"."

guess of a general tendency, whose confirmation will come from actual translations displaying it; and the discovery of a particular difference in a translation pair, which will lead to an attempt to identify it as an instance of a general tendency. In any case, only a detailed study (as will be illustrated in Part III) can allow non-trivial conclusions; the analysis of individual translation pairs is as important as that of general distributional trends.

Before proceeding, let me note a specific restriction on the studies I am discussing here: Faithfulness (or high degree of semantic relationship) and source language interference can be observed (or violated) at the level of the individual translations themselves, in addition to being relevant at the level of the whole text. However, this dissertation will only address the individual translation level, and will therefore make the simplifying assumption that one whole text can be considered a sum of its parts.

Note, nevertheless, that even at the individual translation level, there is considerable difference between studying exactness or naturalness. For the former, the individual pair (supplemented with enough context, of course) is enough, while for the latter, as already mentioned, it is necessary to have a notion of typical, or natural, target language text, that can only be arrived at by studying large amounts of original texts in the target language. In addition, it would be desirable to obtain information on the markedness of the source text as well, because a translation should strive to preserve as much as possible the degrees of markedness in the original.²² A marked target language expression could thus be attributed to a conscious decision on the part of the translator to mirror a marked original form, or identified as a case of more or less clear translationese.

For ease of presentation, I will start by what I call a naive identification model, and I will point out its weaknesses one by one, suggesting corresponding improvements, in Section 3.8.1. These will be spelled out in a second identification model (Section 3.8.2), in which a list of plausibly interesting cases will be expressed. I should, however, note that this model is not a description of the method followed in the empirical studies in this dissertation; rather, it is a hopefully convenient abstraction to be useful for further work. A small example framed in this model will thus be given in Section 3.8.3.

3.8.2.1 A naive model

The naive model with which I start was actually the one I came up with when beginning (empirical) work in this domain. The direct influence of the literature on translation mismatches should be easy to spot.

Comparing grammatical systems, I assume that the factors at play are a finite number, usually small, of discrete observable units. I will represent them by capital letters. Let A, B, etc. denote formal features in a particular language and the dash the translation relation:

i. The existence of a one-to-many translation pairing may reveal objective distinctions which are only implicit (or partially specified) in the source language. For example, given the existence of

²² Incidentally, this is one of the most conspicuous translation losses in general. Translators tend to stay in a neutral register and miss metalinguistic features like dialects, archaic language, etc., which are known to be very difficult to translate. This has been noted i.a. in Gellerstam (1986) and Baker (1993), and it was strikingly true in my corpus (cf. in particular Chapter 12).

pairs A-B and A-C, one should investigate whether:

1. B,C represent meaning differences not overtly marked in the source language, but understandable from the context; (context-resolved ambiguity or vagueness)
2. B,C imply different meanings, which were simply left unspecified in the source text. Each choice is possible, but results in a different interpretation; (non-resolved ambiguity or vagueness)
3. B,C are more or less interchangeable (i.e., corresponding to distinctions which are not semantic in nature). (synonymy)

ii. On the other hand, a many-to-one translation pairing may reveal that a form of the target language has two (or more) distinct uses. For example, given the existence of the pairs A-D and B-D, one should investigate whether:

4. A,B correlate with some other mark in the target language beside D;
5. the target language loses information, i.e., there is no way of specifying just those meanings;
6. A,B are more or less interchangeable.

iii. Finally, the existence of relatively direct mappings may pin-point areas where the two systems use the same kind of devices to mark the same kind of distinctions (and therefore syntactically driven translation can be successful).

Now, after the extensive discussion in this and the previous chapter, some basic flaws of this reasoning must be obvious:

1. It is not formal features in themselves that should be compared, because one cannot identify a formal marker with only one meaning / function / use.
2. One should not treat all translation pairs equally, because they display different semantic relations and different levels of success (i.e., quality).
3. The particular context in which a particular translation occurs must be taken into consideration: (i) the context may enforce a particular meaning; (ii) the context may not be informative enough to discriminate among more than one possible meaning; (iii) the context may force the interpretation of a particular device to be different from the usual; (iv) the context may grant more or less weight to a particular feature, which therefore may be dropped or emphasized, by translation.

In other words, one cannot identify target text with target language. What one has access to is texts, not directly the two language systems, as presupposed in the above model.

A revision of the method taking into account these criticisms would then take A,B, etc. to be form-meaning pairs in context, and discriminate the dash (-) according to type of translation. For example, it would obviously be advisable not to use the pairs featuring translation failure as evidence for any systematic relationship between source and target language items.²³ Likewise, the translation element of the translated pair should be evaluated according to possible translationese.

²³ Even though the analysis of systematic errors could be important evidence to distinctions and differences between the two languages, which had not been mastered (or realized) by the translator.

In turn, these two issues (translationese and kind of semantic relation) imply that the direction of translation matters (and, therefore, in the model, the dash will be replaced by an arrow). This is, incidentally, one of the considerations that receives stronger confirmation by looking at actual instances of translation.

One final problem with the model above concerns the roles of the features themselves in relation to the system. It presupposes that, whenever more than one feature is at stake in one language, the features are in equipollent opposition, i.e., in case i., that B,C,D are sisters as values of one category. Consequently, matters such as default values or hierarchical relations between markers (privative oppositions, degrees of markedness, etc.) are not contemplated. Furthermore, B, C and D can even belong to different categories altogether.

This remark is especially relevant for obligatory categories (accidence categories in Dahl's (1985) terms) -- like tense in English or Portuguese --, because matters like default use are commonly involved in their description and explanation. Now, if A is a default value in one language's system which marks some features, say z , y and x , it is not in fact some positive properties of the utterance that are at play, but their absence. If the other language marks other features, like v , w and q , and also has a default case B, one may have to reckon first with the whole range of positively correlated cases in order to be able to make something out of the correlation of the two default markings (i.e., one may have to analyse the cases $z \rightarrow v$, $z \rightarrow w$, $z \rightarrow q$, $z \rightarrow B$, $y \rightarrow v$, $y \rightarrow w$, ..., $A \rightarrow v$, $A \rightarrow w$, $A \rightarrow q$ before looking at $A \rightarrow B$).²⁴

A particular case I would like to point out is the possibility of having one and the same category (across languages) with a markedness reversal, i.e., even if values of some category in one language A and B were associated to A' and B' in another language (a highly ideal case), still the marked element in one language could be the unmarked one in the other. I thus conclude that, although very common, discussions on e.g. which element of the pair imperfective/perfective is marked, based on data from many languages, simply do not make sense.

3.8.2.2 A more realistic model

I maintain the assumption that the factors at play are a finite number of discrete units, but they will now correspond to a pair of form and meaning in context, and I will also represent them by capital letters. An arrow stands for the translation relation in terms of competence, i.e., $A \rightarrow A'$ stands for "A can be translated by A'", while a double arrow, as in $A \Rightarrow A'$, stands for the actual occurrence in a translation pair of A translated by A'.

A model for corpus-based contrastive studies should first identify clearly what its input (the data on which it draws) and its expected output are. As can be seen at once, such studies draw on several kinds of data:

On the one hand, information on translation proper:

²⁴ An interesting consequence of this remark seems to be a prediction that the more marked the forms the more easily they lend themselves to a default translation pairing. By contrast, in order to explicate the pairing of default forms in two languages one must resort to highly complicated rules.

1. Competence-oriented models of translation, that say $A \rightarrow B, C$ (A is translated by B or C, or A can be translated by B and by C).
2. Actually occurring translation relations $A \Rightarrow B$ and their distribution.

On the other hand, information on each language:

3. Analyses of the monolingual systems to which A and B belong.
4. The distribution of markers in source and target languages: $\text{freq}_{\text{orig}}(A)$ and $\text{freq}_{\text{orig}}(B)$, as well as in translated text: $\text{freq}_{\text{trans}}(A)$ and $\text{freq}_{\text{trans}}(B)$.

The object of the study, then, is the differences and similarities between the two languages, from the semantic relationship between the actual markers A and B.

Let me note that this is a rather abstract model. In fact, the differences and similarities must also refer to the analyses of the monolingual systems: there must obviously be a non-null correlation between the output and the input of the study. Also, there must be some overlap between $A \rightarrow B$ and $A \Rightarrow B$ in order for both to be relevant kinds of input.

I proceed thus by selecting kinds of situations for further scrutiny.

1. The first kind corresponds to the standard confirmation of hypotheses in a corpus, and is of the form "competence model implies performance description".

Such an experimental rule attempts to verify a particular competence model (translation relation and/or description of language system) in the light of the data. If the performance description is met, one can say that it provided positive evidence for the competence model. Alternatively, if a particular situation concerning actual data can be described according to a performance description present in one of the posited experimental rules, it can lead to a hypothesis, so far unexpressed, concerning the language system and/or the translation relation.

Examples are:

a) $A \rightarrow B, C, D$ implies $\text{freq}_{\text{trans}}(B, C, D) \neq \text{freq}_{\text{orig}}(B, C, D)$, in particular $\text{freq}(B) / (\text{freq}(B) + \text{freq}(C) + \text{freq}(D))$ is higher in translated than in original text. In words, the above schematic description is intended to mean that, when there are a number of grammatical markers in the target language with (roughly) the same meaning or use as A in the source language, translationese will occur. In parallel with the standard translation of lexical items noted by Gellerstam,²⁵ I predict that one of the possible target translations will be preferred, having thus a significantly higher frequency than in original text (in the target language).

b) $A + \text{obligatory } B \rightarrow A' + \text{optional } B'$ implies $\text{freq}_{\text{trans}}(A' + B') \gg \text{freq}_{\text{orig}}(A' + B')$ I.e., if there is an optional marker in the target language ("standardly") corresponding to an obligatory one in the source language, it is to be expected that its frequency will increase. This is another kind of translationese.

c) $C = \text{compact } (A, B) \rightarrow A' + B'$ implies $C \Rightarrow A', C \Rightarrow B', C \Rightarrow A' + B'$ and $\text{freq}(A' + B') /$

²⁵ Gellerstam noted that "Standard -- or "press-the-button" -- translations seem to have certain characteristics in common. The "right" way of translating the English word is taught at school, you find it in wordlists and dictionaries. You ask anybody in Sweden what *arrive* is in English and you will get the answer *anlända*. The trouble is that *anlända* is a fairly rare word in Swedish. [...] These are typical translationese words with a wide stylistic gap between the English and the Swedish word" (Gellerstam, 1986:92).

(freq(A')+freq(B')+freq(A'+B')) is larger in translated than in original text. When there is a device which packages more than one meaning in a compact expression, while the target language must express them separately, more often than not only one of the features is conveyed in the translation. Still, it should be the case that translationese occurs, in that there will be more conjoined cases (A' + B', i.e., cases where both A' and B' are expressed) than in original text.

d) vague (A,B) -> A', B' implies either (i) freq_{orig}(A') ≈ freq_{trans}(A') and freq_{orig}(B') ≈ freq_{trans}(B') and neither A' nor B' are marked; or (ii) freq(A') >> freq(B') in translated versus original text and B' is marked. When a source item is vague between two (or more) cases, and the target language cannot preserve such vagueness, we face a translation mismatch as described originally by Kameyama et al. (see above). Strictly speaking, this is a case where the source language is unable to influence the translation, and so, no consistent pattern in the observed translation relation can be expected. In fact, it may be that A and B are not even really categories of the source language, and are just introduced for translation's sake. On the other hand, they may be derived from the context or from the translator's intuitive interpretation of the text. But there should be no source language oriented rules to decide between A and B, since the language is not sensitive to such a distinction. We are thus left with target language oriented strategies. If in the target language one of A' and B' is marked, the translator's choice would probably be the less marked, and translationese could show in the relative frequencies of A' and B'. In any case, I believe that purported vagueness in the source language is one of the most interesting cases to study in detail, and I shall employ this concept fairly often when describing translation between English and Portuguese in Chapter 7.

d') A -> A', C'=vague (A', B') implies freq(A')/(freq(A')+freq(C')) higher in translated text This is the dual rule of d) above, involving vagueness in the target language. Even though it would be more akin to the target language not to specify a particular feature, the influence of the source language is predicted to produce text in which such a feature is more often specified than usual.

2. The second kind of interesting situation involves the relationship between competence and performance descriptions. These situations do not allow one to formulate a rule, but surely require further study. Examples are:

e) A -> B and freq(A) ≈ freq(B) ≈ freq(A=>B) Expressed in words, correspondence between competence model and distribution seems to provide positive evidence for the equivalence of meaning and use of A and B.

f) A => B and there is no known relation A -> B. This may indicate that A is foreign to the target language, irrespective of whether these translations are due to translation errors or correspond to special cases.

g) A -> B; B -> A and freq(A => B) >> freq(B => A). Expressed in words, this corresponds to the case when it is generally held that A and B point to equivalent concepts in the two languages, but the translation from A to B is much more frequent than from B to A. This case may be due to several distinct possibilities:

i. freq(A) >> freq(B) The apparent asymmetry is due to the fact that the language of A expresses much more frequently the concept denoted by A and B;

ii. $B \rightarrow A, C, D \dots$ and $\text{freq}(A \Rightarrow B) \approx \text{freq}(B \Rightarrow A) + \text{freq}(B \Rightarrow C) + \text{freq}(B \Rightarrow D) + \dots$ The asymmetry is due to the existence of other translation choices. I.e., A is paradigmatically related to other devices C and D (at least from the language of B's point of view).

iii. $B \Rightarrow \text{vague}(A, C)$ The situation stems from the existence of an apparently more appropriate translation of B into the language of A which does not make A explicit (i.e., involves an expression that is vague relative to A).

iv. $B \Rightarrow \text{compact}(A, C)$ The situation stems from the existence of an apparently more appropriate translation of B into the language of A which does not make A explicit in isolation; rather, it packs it together with some other information.

v. $B \Rightarrow_{\text{add}} C$ The explanation for the asymmetry can be rendered in terms of semantic relation type: a significant number of translations of B involved explicit addition (" \Rightarrow_{add} " stands for "translated by adding", encompassing the semantic typology cases of information addition and translation through intermediary representation).

vi. $B \Rightarrow_{\text{dim}} C$ Likewise, a significant number of translations of B involved explicit loss (" \Rightarrow_{dim} " stands for "translated by removing some information", encompassing the semantic typology cases of subsumption and translation through intermediary representation).

vii. At least one of A or B is a default value, a case which has been discussed above.

3. Finally, a third case of interesting situations indexes translations according to semantic relation, something which has already been illustrated by 2.v. and 2.vi. above. Examples are:

h) $A \rightarrow B$ and $A \Rightarrow_{\text{fail}} B$ If, although the translation competence literature claims a particular translation relation, in practice it often results in translation failure, one must either dismiss the rule or use it as an error probe.

i) $A \rightarrow B$ and $A \Rightarrow_{\text{add}} C$ If the translation relation is more often than not that of information addition, one may conclude that B alone is rare in the target language, possibly because it occurs in compact predicates ($C = \text{compact}(B, x)$).

Summing up, this schematic presentation details what (semantically annotated) quantitative data can contribute to contrastive analysis. The analyses involved in each case are far from trivial, though, as will be demonstrated by the discussion of one particular example.

3.8.2.3 An example

I will discuss here some cases involving the progressive, since this device is not especially studied in this dissertation, despite being a major issue in English aspect.

The (relevant) input is the following data, displayed in Tables 3.1 and 3.2 (where "freq" stands for absolute frequency):

Table 3.1: Contrastive data

Competence	Performance
progP \rightarrow progE	$\text{freq}(\text{progP} \Rightarrow \text{progE}) = 12$
Imperfeito \rightarrow progE	$\text{freq}(I \Rightarrow \text{progE}) = 85$

Imperfeito -> simple past	freq(I=>past) = 623
Imperfeito -> gerund	freq(I=>ger) = 26

Table 3.2: Monolingual distribution of the progressive

freq(progE _{trans}) = 98
freq(progE _{orig}) = 58

The observation that there is a much larger number of progressives in English translated from Portuguese, as compared to the number in original English text, is straightforward. Most occurrences of English progressive, furthermore, (actually 85 out of 98) were translations of Imperfeito (only 12 translating Portuguese progressives). The question is whether this can be considered a case of translationese, and, furthermore, whether there is something one can learn from this distributional difference in what concerns the systems of the two languages. I will try to spell out in detail all assumptions and theoretical claims about the systems in the two languages required to draw any conclusions.

The first assumption required is (i) that there are no other grammatical features of Portuguese that consistently translate into English progressives, and could therefore explain the marked increase of this device. In this case, the assumption is uncontroversial.

(ii) The translation relation Imperfeito -> progressive should then be rendered more explicitly as Imperfeito -> past + progressive.

One possibility that readily comes to mind is to analyse this situation as translationese of kind (b), namely, A + obligatory B -> A' + optional B' leads to increase in B'. In order to do so, one needs to analyse Imperfeito as a binary combination of two obligatory features, and choose one of the two components of the past progressive as optional. This leads to assuming the next two hypotheses:

(iii) that Imperfeito can be analysed as "past + obligatory conveying of imperfectivity",

(iv) and that English progressive in the past is always optional, yielding past + obligatory imperfectivity -> past + optional progressive.

Now, hypothesis (iii) must of course be restricted to the cases where Imperfeito gets translated by the English progressive (otherwise it would be trivially false), but, given that we are dealing with form-meaning pairs and not simply with forms in our description of translationese, let us assume that that is all right. Hypothesis (iv), namely that, in the past tense, the progressive is never obligatory in English, is harder to confirm, but at least in some cases the progressive is optional; cf. *He wore/was wearing glasses when I saw him*.

Finally, in order to verify the case as an instance of the (b) type of translationese, one has to analyse the other translations of Imperfeito (or, better, those sharing the particular meaning at stake) in order to check whether (v) there are no other consistent translations of that form-meaning pair. But this is where the analysis breaks down: In addition to the obvious case of Imperfeito -> past (whose existence was actually required by the hypothesis), another consistent translation, that of Imperfeito -> gerund, could be observed. Gerund, in addition, can be analysed as leaving out past and stating

"progressiveness".

The hypothesis of translationese can be maintained, though, by substituting (c) for (b): Imperfeito = compact (past, imperfectiveness) -> past + prog. This should imply an increase of "past + prog" compared to "past" and/or "prog" alone. The performance description that must be met agrees with an increase of past progressive (past + prog), but requires now, additionally, that both simple past and "simple gerund" are found as translations of Imperfeito, and that both decrease in frequency compared to original English text.

Now, in order to see whether the particular sentences (whose number was) taken as support for this conclusion actually supported it, one would then have to analyse (at least) all translations of Imperfeito into respectively simple past, past progressive, and gerund. For those instances which could be considered to correspond to the same meaning of Imperfeito in Portuguese, relative frequencies should be computed; then relative frequencies in English original text should be obtained (for simple past, past progressive, and gerund for the meanings at stake), and finally compared to those of the English translations to see whether there was a statistically significant difference.

The steps described in the previous paragraph constitute in my view the sort of procedure that, because they are time consuming and relatively boring, should be automated as much as possible, allowing researchers to concentrate only on the parts requiring their judgement.

3.9 Conclusion

This chapter has several goals: Firstly, it presented a set of relevant problems in the formalization of translation as well as in the way it has been discussed. Secondly, building on the detailed study of actual translations that constitute the empirical work underlying this thesis, it suggested a typology of semantic relations between the two elements of a translation pair. Thirdly and perhaps most important, it argued for translation-based contrastive studies from a theoretical point of view. Finally, although it did not propose a full-fledged methodology for this kind of studies, it discussed a set of methodological issues and illustrated several apparently appropriate points of departure.

I summarize here the main ideas I have argued for in Chapters 2 and 3, concerning the contrast and comparison of the language systems.

- Languages constitute unique systems that should be studied in their own terms.
- According to Jakobson, languages differ not so much in what they can say, but in what they must say. But even though languages differ not so much in what they can say, from the study of real texts I conclude that they differ much more in what they do say.
- Translationally related grammatical devices in no way presuppose equivalence of meaning, be it meaning in context or systematic meaning.
- Translation (a real activity outside linguistics proper) is the best way to contrast and compare two languages.
- In order to compare two language systems with the help of translations, one must be able

to assess the semantic relationship between the two elements of those translations.

- Some properties of the target language text, rather than being typical of the target language, are induced by the translation process itself. Thus, looking into original target language texts as well is an important source of contrastive data.

After introducing tense and aspect properly in the next chapter, I present a computational model in Chapter 5 which is designed to describe parallel corpora as far as tense and aspect are concerned. This model is meant to take care of the claims suggested here as well as to be a practical aid in the contrastive task (among other possible applications).